From Population to Individual:

A Data-Driven Approach for Enhancing Patient-Centered Decision-Making in Hand and Wrist Care

Lisa Hoogendam

ISBN: 978-94-6522-258-5

Lay-out & cover design: Arina van Londen | www.ridderprint.nl

Printing: Ridderprint | www.ridderprint.nl

Van populatie naar individu:

Een datagedreven aanpak voor het verbeteren van patiëntgerichte besluitvorming binnen hand- en polszorg

From Population to Individual:

A data-driven approach for enhancing patient-centered decision-making in hand and wrist care

Proefschrift

ter verkrijging van de graad van doctor aan de

Erasmus Universiteit Rotterdam

op gezag van de

rector magnificus

Prof.dr.ir. A.J. Schuit

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

19 november 2025 om 13:00 uur

door

Lisa Hoogendam

geboren op 22 maart 1996 te Rotterdam

Erasmus University Rotterdam

Promotiecommissie:

Promotor:	Prof.dr. R.W. Selles
Overige leden:	Prof.dr. S.M.A. Bierma-Zeinstra
	Dr. T.J. Hoogeboom
	Prof.dr.ir. P.R. Rijnbeek
Copromotoren:	Dr. J.S. Souer
	Dr. R.M. Wouters

TABLE OF CONTENTS

1

GENERAL INTRODUCTION

FROM POPULATION TO INDIVIDUAL:

A Data-Driven Approach for Enhancing Patient-Centered Decision-Making in Hand and Wrist Care

Hand and wrist conditions affect a relatively large part of the population over the course of their lives. The yearly prevalence of musculoskeletal upper extremity conditions has been reported to range up to 41% (1), and in elderly, the monthly prevalence of hand pain is 17% (2). Hand and wrist conditions can affect patients in their daily functioning and their ability to work, and limit social participation.

Hand and wrist conditions are also associated with considerable societal costs and can directly impact quality of life (3-5). For example, in 1997-2009 the annual healthcare costs of hand and wrist injuries are estimated at €670 million in The Netherlands, which ranks them among the most expensive injuries (6), and these costs have likely only increased over time. In the United States, 5.9 million ambulatory care visits annually are due to musculoskeletal hand and wrist conditions (7).

Many hand and wrist conditions (e.g., osteoarthritis, nerve compression syndromes, and tenosynovitis) can negatively impact patients in their daily lives but are not considered medical emergencies. Thus, there is sufficient opportunity and time for patients to be informed about the treatment options (including refraining from treatment) and consider the possible risks and benefits of these options (8). Within a patient-centered care setting, it is the current best practice that patients decide on the most suitable treatment option with their treating physician while considering their preferences and values. This decision-making process is commonly referred to as shared decision-making (9).

Shared decision-making

While shared decision-making is considered the best practice and preferred by patients, in daily clinical practice, this is not always achieved (10-12). For example, in an outpatient vascular surgery setting in The Netherlands, 31% of the patients prefer themselves to have a more active role in shared decision-making, and 58% prefer a different decision-making approach than what they had experienced (11). Also, elderly patients with hand and wrist conditions indicated that they prefer a patient-directed or collaborative approach to decision-making, compared to a clinician-directed approach (13). For physicians, barriers towards shared decision-making include lack of time and when they see themselves as decision-makers (12, 14, 15). Some physicians fear to be seen as indecisive by their patients when they engage in shared decision making (14, 15). Additionally, specific patient characteristics (e.g., multiple comorbidities and lower education levels) are considered barriers to the shared decision-making process, possibly because treatment outcomes may be more uncertain for patients with certain characteristics or because physicians expect difficulties in adequately conveying treatment benefits and risks (16).

To overcome the barriers to the shared decision-making process, adequate, relevant, and unbiased information provision to patients about the treatment options, risks, and benefits is vital (16). Decision aids (possibly based on outcome information) may direct this information provision, but they typically provide general and not patient-specific information (17). However, for many conditions, information about all relevant outcome domains (e.g., pain, function, return to work, complications) for all treatment options is not directly available. In current daily clinics, decisions are usually based on clinician experience and guidelines and therefore not consistently personalized and at risk of being noisy due to between-clinician variance (18). While the increasingly routinely collected patient and outcome data is highly valuable to overcome this, it is still insufficiently used for providing unbiased, data-driven decision support. **Therefore, the main aim of this thesis is to provide patients with hand and wrist conditions and their clinicians with clinically relevant, personalized, and data-driven outcome information, to facilitate shared decision-making towards the most suitable treatment option.**

Shared decision-making is often mentioned in the context of value-based healthcare. This healthcare delivery model is focused on providing maximal value (i.e., positive health outcomes) for each dollar or euro spent (19). An important aspect of value-based healthcare is that value is defined from a patient's perspective (19). The shared decision-making approach, where patients' values, preferences, and goals play a central role in the decision-making, complements this (20).

Measuring outcomes routinely and learning from every patient systematically is commonly considered a first step towards value-based healthcare, including shared decision-making (19, 21, 22). The data collected through routine outcome measurement can be used for multiple purposes, including using the data to directly or indirectly improve patient care and facilitate research (23, 24). Patient care can be directly improved by directly providing individual patient data to clinicians. For example, a dashboard can show clinicians how this individual patient rated his/her symptom severity, which can serve as a starting point for the consultation. Additionally, routinely collected data from specific patient groups can be used to visualize outcomes over time for one or more treatment options and a particular diagnosis. Such applications can potentially make the shared decision-making process more data-driven than expert option-based (25).

Data collected through routine outcome measurement can also indirectly improve healthcare (23, 24). For example, the data can be used for research on treatment effectiveness and can inform guideline development or analyses regarding societal treatment costs. The data could facilitate benchmarking and healthcare costing experiments, such as bundled payments and pay-for-performance (21). Advantages of using routine outcome data for such research is that the patient samples are often larger and more representative of daily clinics than those used in randomized controlled trials, providing higher ecological validity

(26, 27). Subsequent implementation of these research outcomes can, therefore, improve daily clinical care.

Thumb base osteoarthritis

In this thesis, a condition of specific interest within the general group of hand and wrist conditions is thumb base osteoarthritis (OA). Thumb base OA can be a painful condition limiting activities of daily living that affects a relatively large part of the population, mostly females (28, 29). In the general population, the radiographic prevalence is estimated at 33% for males at age 80 and 39% for females at age 80 (30), and the symptomatic prevalence at a mean age of 59 is estimated at 2% for males and 7% for females (31).

There is a wide range of treatment options for thumb base OA, both nonsurgical and surgical. Nonsurgical treatment can comprise analgesics, orthotics, hand therapy, or corticosteroid injections (32). Surgical treatment options include isolated trapeziectomy, trapeziectomy with tendon interposition and/or ligament reconstruction (LRTI), arthrodesis, joint replacement, and denervation (33). In the Netherlands, patients with symptomatic thumb base OA are first treated nonsurgically, according to the Dutch guideline (34). When nonsurgical treatment fails to relieve symptoms sufficiently, typically surgical treatment is proposed, such as trapeziectomy (with LRTI) (35, 36). However, despite many systematic reviews and meta-analyses (37-39), the optimal treatment strategy for thumb base OA in terms of (cost)effectiveness is unknown. Many previous studies included a small number of patients, had a short follow-up period, or described outcomes that may not be clinically relevant or directly meaningful to patients (e.g., clinician-reported treatment success). **Consequently, a subgoal in this thesis was to describe clinically relevant outcomes, compare treatment outcomes, and assess the cost-effectiveness of possible treatment strategies for thumb base OA.**

The considerable variation in outcomes of (non)surgical) treatment options for thumb base OA between patients is an additional challenge for clinicians (40, 41). For clinicians, it can be relevant to know which characteristics are related to good and poor treatment outcomes. These prognostic factors can help understand the variation in treatment outcomes and can help clinicians inform their patients whether they can expect somewhat better or worse treatment outcomes than the average patient, which might lead to a different treatment recommendation. A particular interest is the role of mindset factors (defined as symptoms of depression, anxiety, pain catastrophizing, and illness perceptions) and outcome expectations. In orthopedic surgery, mindset factors have previously been shown to be related to treatment outcomes (42). However, it is unclear whether this association is also present for treatment outcomes of hand and wrist conditions. Moreover, the role of expectations is even less clear, while expectations might be modifiable by clinicians (43). **Therefore, another subgoal in this thesis was to study how mindset factors and**

expectations relate to treatment outcomes for patients receiving treatment for thumb base OA.

Prediction modeling

While prognostic factors provide some global insight into variations in treatment outcomes, it can be difficult to translate this knowledge into specific expected treatment outcomes for individual patients. Ideally, a clinician would have access to reliable, patient-specific estimates of the expected treatment outcomes for each available treatment option for each individual patient. To this end, prognostic prediction models can be used since they predict treatment outcomes for individual patients based on predictor variables (44). Such predictor variables can consist of but are not limited to, patient characteristics (e.g., age, sex, and comorbidities) and disease characteristics (e.g., patient-reported symptom severity and radiographic stage of osteoarthritis) and do not need to be causally related to the outcome of interest. However, before prediction models can safely be used in clinical practice, they need to be validated to ensure that the model will also make reliable predictions for future patients (45). Particularly in cases where the likely treatment outcomes are uncertain and where many treatment options are available, such as in thumb base OA, implementing validated prediction models in clinical practice could have a great potential to facilitate shared decision-making by informing clinicians and patients. **Therefore, our subgoal in this thesis was to develop and validate prognostic prediction models for clinically relevant outcomes of multiple treatments for hand and wrist conditions and to assess whether using these models would improve clinical decision-making.**

Patient-Reported Outcome Measures

In this thesis, data from the Hand-Wrist Study Group cohort were used. This cohort results from fifteen years of routine outcome measurement at Xpert Clinics, a dedicated center for elective hand and wrist care (46). As of 2024, the cohort contains data of roughly 185.000 patients with >500 diagnosis-treatment combinations. To incorporate the patients' perspective that plays a pivotal role in the value-based healthcare philosophy, patient-reported outcomes and patient experiences are integral components. Therefore, patients with a hand or wrist disorder are asked to complete Patient-Reported Outcome Measurements (PROMs) at fixed time points prior to and during their treatment, as well as Patient-Reported Experience Measurements after consultation and treatment. Also, Clinician Rated Outcome Measures (e.g., range of motion and grip strength), sociodemographic characteristics, and measures of treatment outcome expectations and mindset (e.g., pain catastrophizing, anxiety) are collected.

PROMs are questionnaires regarding outcomes such as pain, hand function, or symptom severity. PROMs can capture the patients' perspective on his or her illness, therefore providing valuable information for clinicians (19). A challenge when using PROMs is

that they are often on a numeric scale, such as 0-100, and therefore lack intrinsic meaning for patients and clinicians. This makes it difficult to determine whether a statistically significant improvement following treatment is also relevant for patients and clinicians (47-49). To overcome this issue regarding the interpretation of PROMs, Clinically Important Outcome Values (CIOVs) have been introduced (50, 51). The Minimally Important Change and the Patient Acceptable Symptom State are the most well-known examples of CIOVs. The most frequently used metric is the Minimally Important Change, which refers to the minimal change in an outcome where patients consider themselves improved (52-54). Conversely, the Patient Acceptable Symptom State (PASS) is a CIOV that does not reflect a threshold based on a change score, but instead an outcome value at a chosen time point where a patient is satisfied with his or her current level of symptoms (55).

While CIOVs are highly relevant for interpreting PROM values, there are methodological limitations to the values currently presented in hand and wrist literature, which limits their applicability. Multiple calculation methods are used (56-58), even though several frequently used methods can provide biased estimates (59, 60). Moreover, there are indications that CIOVs vary for the same PROM depending on the diagnosis and treatment (61), while most CIOVs are based on either a diverse set of hand and wrist conditions or only one specific diagnosis-treatment combination. This means that the currently available CIOVs often cannot be directly applied to the broad spectrum of hand and wrist diagnoses and treatments. There is a need for a golden standard regarding calculation methods and an overview of unbiased CIOV estimates that can readily be used in patients with specific diagnoses and/or treatments. **Therefore, our final subgoal in this thesis was to determine the Minimally Important Change and Patient Acceptable Symptom State values for hand and wrist conditions and investigate how these values should be calculated.**

AIMS AND OUTLINE OF THIS THESIS

The aim of this thesis is to provide patients with hand and wrist conditions and their clinicians with personalized, data-driven information on clinically relevant treatment outcomes, to facilitate shared decision-making towards the most suitable treatment option. I aimed to achieve this by describing average treatment outcomes, assessing which factors explain variation in treatment outcomes, and predicting individual treatment outcomes. However, it should first be determined what patients consider a meaningful outcome. Therefore, this thesis starts by defining clinically meaningful outcomes for patients with hand and wrist conditions. This thesis is structured accordingly, as seen below.

PART I: DETERMINING CLINICALLY IMPORTANT OUTCOME VALUES

The first part of this thesis aimed to determine Clinically Important Outcome Values for PROMs that are frequently used to measure outcomes of hand and wrist treatments. In **Chapter 2**, we determined the MIC for 36 combinations of hand or wrist conditions and associated treatments. In **Chapter 3**, we evaluated the accuracy and precision of multiple methods, including a new Item Response Theory-based method, to calculate the PASS. In **Chapter 4**, we used the new Item Response Theory-based method to determine the PASS for all multi-item PROMs for 35 diagnosis-treatment combinations for hand and wrist conditions.

PART II: CLINICALLY IMPORTANT OUTCOMES OF TREATMENT (STRATEGIES) FOR THUMB BASE OSTEOARTHRITIS

The second part of this thesis aimed to determine clinically relevant outcomes of both nonsurgical and surgical treatment options for thumb base osteoarthritis. In **Chapter 5**, we evaluated patient-reported outcomes and conversion to surgery five years after nonsurgical treatment for thumb base OA. In **Chapter 6**, we evaluated the prevalence of complications and the association between complications and treatment outcomes following trapeziectomy with a Weilby sling in patients with thumb base OA. In **Chapter 7**, we compared treatment outcomes of four different surgical techniques to treat thumb base OA. In **Chapter 8**, we combined outcomes with costs to evaluate the cost-effectiveness of four potential treatment strategies for thumb base OA over a 10-year period, taking both treatment-related and societal costs into account.

PART III: PROGNOSTIC FACTORS FOR CLINICALLY IMPORTANT TREATMENT OUTCOMES (FOR THUMB BASE OSTEOARTHRITIS)

The third part of this thesis aimed to study which factors explain variation in treatment outcomes. In **Chapters 9** and **10,** we studied how expectations and mindset factors are associated with pain and hand function before and three months after the start of nonsurgical treatment for thumb base OA. In **Chapter 11**, we aimed to explain satisfaction with treatment results following nonsurgical thumb base OA treatment. In **Chapter 12**, we aimed to identify factors associated with satisfaction with the treatment results and willingness to undergo the treatment again for six common hand and wrist disorder treatments. Since **Chapters 10**, **11**, and **12** showed that expectations were strongly associated with treatment outcomes, we tried to understand which factors drive pre-treatment expectations for

patients with hand and wrist conditions in **Chapter 13**. In **Chapter 14,** we studied the outcomes of patients involved in a personal injury claim. In other fields, personal injury claim involvement has been associated with worse treatment outcomes after surgery, but it is unknown whether this association is present for patients treated (non)surgically for hand and wrist conditions.

PART IV: DEVELOPMENT, VALIDATION, AND ADDED VALUE OF PREDICTION MODELS IN HAND SURGERY

The final part of this thesis focused on developing and validating individual prediction models to inform patients about their individual probability of a clinically important treatment outcome, aiming for the models to be implemented in clinical practice. In **Chapters 15 and 16**, we developed and internally validated prediction models for the probability of a clinically relevant improvement in symptoms following carpal tunnel release surgery and surgical treatment for thumb base OA, respectively. In **Chapter 17**, we evaluated the added clinical value of the carpal tunnel release prediction model by comparing predictions made by hand surgeons with predictions made by the prediction model. In **Chapter 18**, we performed a systematic review of the literature regarding prediction models for outcomes of upper extremity surgery, where we assessed the clinical applicability of published prediction models. Finally, we evaluated whether a single prediction model can be used for multiple hand and wrist conditions and compared the performance of this model to the typical condition-treatment-specific prediction models in **Chapter 19**.

This thesis will be concluded with a general discussion, where the findings and future perspectives will be discussed (**Chapter 20**).

REFERENCES

1. Huisstede BM, Bierma-Zeinstra SM, Koes BW, Verhaar JA. Incidence and prevalence of upper-extremity musculoskeletal disorders. A systematic appraisal of the literature. BMC Musculoskelet Disord. 2006;7:7.
2. Dahaghin S, Bierma-Zeinstra SM, Reijman M, Pols HA, Hazes JM, Koes BW. Prevalence and determinants of one month hand pain and hand related disability in the elderly (Rotterdam study). Ann Rheum Dis. 2005;64(1):99-104.
3. Kovacs L, Grob M, Zimmermann A, Eder M, Herschbach P, Henrich G, et al. Quality of life after severe hand injury. J Plast Reconstr Aesthet Surg. 2011;64(11):1495-502.
4. Srikesavan C, Dean B, Toye F, Bolt A, Thurley N. Experiences of people with hand osteoarthritis and perceptions of clinicians and carers on quality of life and treatment services: A qualitative synthesis. Hand Ther. 2023;28(4):121-32.
5. Gabrielli AS, Lesiak AC, Fowler JR. The Direct and Indirect Costs to Society of Carpal Tunnel Release. Hand (N Y). 2020;15(2):NP1-NP5.
6. de Putter CE, Selles RW, Polinder S, Hartholt KA, Looman CW, Panneman MJ, et al. Epidemiology and health-care utilisation of wrist fractures in older adults in The Netherlands, 1997-2009. Injury. 2013;44(4):421-6.
7. Song A, Kim P, Ayers G, Jain N. Characteristics of Non-Spine Musculoskeletal Ambulatory Care Visits in the United States, 2009-2016. PM R. 2021;13(5):443-52.
8. Stiggelbout AM, Pieterse AH, De Haes JC. Shared decision making: Concepts, evidence, and practice. Patient Educ Couns. 2015;98(10):1172-9.
9. Spatz ES, Krumholz HM, Moulton BW. Prime Time for Shared Decision Making. JAMA. 2017;317(13):1309-10.
10. Haesebaert J, Adekpedjou R, Croteau J, Robitaille H, Legare F. Shared decision-making experienced by Canadians facing health care decisions: a Web-based survey. CMAJ Open. 2019;7(2):E210-E6.
11. Santema TB, Stoffer EA, Kunneman M, Koelemay MJ, Ubbink DT. What are the decision-making preferences of patients in vascular surgery? A mixed-methods study. BMJ Open. 2017;7(2):e013272.
12. Hahlweg P, Harter M, Nestoriuc Y, Scholl I. How are decisions made in cancer care? A qualitative study using participant observation of current practice. BMJ Open. 2017;7(9):e016360.
13. Dardas AZ, Stockburger C, Boone S, An T, Calfee RP. Preferences for Shared Decision Making in Older Adult Patients With Orthopedic Hand Conditions. J Hand Surg Am. 2016;41(10):978-87.
14. Schoenfeld EM, Goff SL, Elia TR, Khordipour ER, Poronsky KE, Nault KA, et al. A Qualitative Analysis of Attending Physicians' Use of Shared Decision-Making: Implications for Resident Education. J Grad Med Educ. 2018;10(1):43-50.
15. Schoenfeld EM, Goff SL, Elia TR, Khordipour ER, Poronsky KE, Nault KA, et al. Physician-identified barriers to and facilitators of shared decision-making in the Emergency Department: an exploratory analysis. Emerg Med J. 2019;36(6):346-54.
16. Waddell A, Lennox A, Spassova G, Bragge P. Barriers and facilitators to shared decision-making in hospitals from policy to practice: a systematic review. Implement Sci. 2021;16(1):74.
17. Stacey D, Legare F, Col NF, Bennett CL, Barry MJ, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2014(1):CD001431.

18. Kahneman D, Rosenfield AM, Gandhi L, Maser T. Reducing Noise in Decision Making. Harvard Bus Rev. 2016;94(12):18-.

19. Porter ME. What is value in health care? N Engl J Med. 2010;363(26):2477-81.

20. Chen M, Sepucha K, Bozic KJ, Jayakumar P. Value-based Healthcare: Integrating Shared Decision-making into Clinical Practice. Clin Orthop Relat Res. 2023;481(3):448-50.

21. Bernstein DN, Calfee RP, Hammert WC, Rozental TD, Witkowski ML, Porter ME. Value-Based Health Care in Hand Surgery: Where Are We & Where Do We Go From Here? J Hand Surg Am. 2022;47(10):999-1004.

22. Silveira Bianchim M, Crane E, Jones A, Neukirchinger B, Roberts G, McLaughlin L, Noyes J. The implementation, use and impact of patient reported outcome measures in value-based healthcare programmes: A scoping review. PLoS One. 2023;18(12):e0290976.

23. Crossnohere NL, Anderson N, Baumhauer J, Calvert M, Esparza R, Gulbransen S, et al. A framework for implementing patient-reported outcomes in clinical care: the PROTEUS-practice guide. Nat Med. 2024.

24. PROTEUS Consortium. Practice guide [Available from: https://theproteusconsortium.org/proteus-practice/proteus-practice-guide/.

25. Feitz R, van Kooij YE, Ter Stege MHP, van der Oest MJW, Souer JS, Wouters RM, et al. Closing the loop: a 10-year experience with routine outcome measurements to improve treatment in hand surgery. EFORT Open Rev. 2021;6(6):439-50.

26. Rabeneck L, Viscoli CM, Horwitz RI. Problems in the conduct and analysis of randomized clinical trials. Are we getting the right answers to the wrong questions? Arch Intern Med. 1992;152(3):507-12.

27. Concato J, Horwitz RI. Beyond randomised versus observational studies. Lancet. 2004;363(9422):1660-1.

28. Anakwe RE, Middleton SD. Osteoarthritis at the base of the thumb. BMJ. 2011;343:d7122.

29. Marshall M, Peat G, Nicholls E, van der Windt D, Myers H, Dziedzic K. Subsets of symptomatic hand osteoarthritis in community-dwelling older adults in the United Kingdom: prevalence, inter-relationships, risk factor profiles and clinical characteristics at baseline and 3-years. Osteoarthritis Cartilage. 2013;21(11):1674-84.

30. van der Oest MJW, Duraku LS, Andrinopoulou ER, Wouters RM, Bierma-Zeinstra SMA, Selles RW, Zuidam JM. The prevalence of radiographic thumb base osteoarthritis: a meta-analysis. Osteoarthritis Cartilage. 2021;29(6):785-92.

31. Haugen IK, Englund M, Aliabadi P, Niu J, Clancy M, Kvien TK, Felson DT. Prevalence, incidence and progression of hand osteoarthritis in the general population: the Framingham Osteoarthritis Study. Ann Rheum Dis. 2011;70(9):1581-6.

32. Wilkens SC, Meghpara MM, Ring D, Coert JH, Jupiter JB, Chen NC. Trapeziometacarpal Arthrosis. JBJS Rev. 2019;7(1):e8.

33. Athlani L, De Almeida YK, Martins A, Seaourt AC, Dap F. Thumb basal joint arthritis in 2023. Orthop Traumatol Surg Res. 2024;110(1S):103772.

34. van Uchelen J, Beumer A, Brink SM, Hoogvliet P, Moojen TMS, A.J.; , Vermeulen GM, et al. Primaire artrose van de duimbasis 2014 [Available from: https://richtlijnendatabase.nl/richtlijn/primaire_artrose_van_de_duimbasis/primaire_artrose_duimbasis_-_korte_beschrijving.html.

35. Deutch Z, Niedermeier SR, Awan HM. Surgeon Preference, Influence, and Treatment of Thumb Carpometacarpal Arthritis. Hand (N Y). 2018;13(4):403-11.

36. Parker S, Riley N, Dean B, Oxford Upper Limb C. Management of osteoarthritis at the base of the thumb. Bone Joint J. 2020;102-B(5):600-5.

37. Challoumas D, Murray E, Ng N, Putti A, Millar N. A Meta-analysis of Surgical Interventions for Base of Thumb Arthritis. J Wrist Surg. 2022;11(6):550-60.

38. Knightly N, Sullivan P. Surgery for Trapeziometacarpal Joint Osteoarthritis: A Meta-Analysis on Efficacy and Safety. J Hand Surg Asian Pac Vol. 2021;26(2):245-64.

39. Yuan F, Aliu O, Chung KC, Mahmoudi E. Evidence-Based Practice in the Surgical Treatment of Thumb Carpometacarpal Joint Arthritis. J Hand Surg Am. 2017;42(2):104-12 e1.

40. Tsehaie J, Spekreijse KR, Wouters RM, Slijper HP, Feitz R, Hovius SER, Selles RW. Outcome of a Hand Orthosis and Hand Therapy for Carpometacarpal Osteoarthritis in Daily Practice: A Prospective Cohort Study. J Hand Surg Am. 2018;43(11):1000-9 e1.

41. Wilcke M, Roginski M, Astrom M, Arner M. A registry based analysis of the patient reported outcome after surgery for trapeziometacarpal joint osteoarthritis. BMC Musculoskelet Disord. 2020;21(1):63.

42. Sheikhzadeh A, Wertli MM, Weiner SS, Rasmussen-Barr E, Weiser S. Do psychological factors affect outcomes in musculoskeletal shoulder disorders? A systematic review. BMC Musculoskelet Disord. 2021;22(1):560.

43. Waljee J, McGlinn EP, Sears ED, Chung KC. Patient expectations and patient-reported outcomes in surgery: a systematic review. Surgery. 2014;155(5):799-808.

44. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009;338:b375.

45. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605.

46. Selles RW, Wouters RM, Poelstra R, van der Oest MJW, Porsius JT, Hovius SER, et al. Routine Health Outcome Measurement: Development, Design, and Implementation of the Hand and Wrist Cohort. Plast Reconstr Surg. 2020;146(2):343-54.

47. Leopold SS, Porcher R. Editorial: The Minimum Clinically Important Difference-The Least We Can Do. Clin Orthop Relat Res. 2017;475(4):929-32.

48. Leopold SS, Porcher R. Editorial: Threshold P Values in Orthopaedic Research-We Know the Problem. What is the Solution? Clin Orthop Relat Res. 2018;476(9):1689-91.

49. Marks M, Rodrigues JN. Correct reporting and interpretation of clinical data. J Hand Surg Eur Vol. 2017;42(9):977-9.

50. Martin R, Disantis A. Editorial Commentary: Patient-Reported Outcome Measures With Established Clinically Important Outcome Values Should Be Used for Hip Arthroscopy Patients: Sifting Through the White Noise. Arthroscopy. 2021;37(10):3122-4.

51. Cepeda NA, Polascik BA, Ling DI. A Primer on Clinically Important Outcome Values: Going Beyond Relying on P Values Alone. J Bone Joint Surg Am. 2020;102(3):262-8.

52. Terwee CB, Peipert JD, Chapman R, Lai JS, Terluin B, Cella D, et al. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res. 2021;30(10):2729-54.

53. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407-15.

54. Copay AG, Subach BR, Glassman SD, Polly DW, Jr., Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J. 2007;7(5):541-6.

55. Tubach F, Dougados M, Falissard B, Baron G, Logeart I, Ravaud P. Feeling good rather than feeling better matters more to patients. Arthritis Rheum. 2006;55(4):526-30.

56. Stjernberg-Salmela S, Karjalainen T, Juurakko J, Toivonen P, Waris E, Taimela S, et al. Minimal important difference and patient acceptable symptom state for the Numerical Rating Scale (NRS) for pain and the Patient-Rated Wrist/Hand Evaluation (PRWHE) for patients with osteoarthritis at the base of thumb. BMC Med Res Methodol. 2022;22(1):127.

57. Wang Y, Devji T, Qasim A, Hao Q, Wong V, Bhatt M, et al. A systematic survey identified methodological issues in studies estimating anchor-based minimal important differences in patient-reported outcomes. J Clin Epidemiol. 2022;142:144-51.

58. Macri EM, Young JJ, Ingelsrud LH, Khan KM, Terluin B, Juhl CB, et al. Meaningful thresholds for patient-reported outcomes following interventions for anterior cruciate ligament tear or traumatic meniscus injury: a systematic review for the OPTIKNEE consensus. Br J Sports Med. 2022;56(24):1432-44.

59. Terluin B, Griffiths P, van der Wouden JC, Ingelsrud LH, Terwee CB. Unlike ROC analysis, a new IRT method identified clinical thresholds unbiased by disease prevalence. J Clin Epidemiol. 2020;124:118-25.

60. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010;63(5):524-34.

61. Schrier VJ, Gelfman R, Amadio PC. Minimal clinically important difference is lower for carpal tunnel syndrome patients undergoing injection versus surgery. J Hand Surg Eur Vol. 2020;45(1):90-2.

Part I

DETERMINING CLINICALLY IMPORTANT OUTCOME VALUES

2

WHAT ARE THE MINIMALLY IMPORTANT CHANGES OF FOUR COMMONLY USED PATIENT-REPORTED OUTCOME MEASURES FOR 36 HAND AND WRIST CONDITIONS?

Hoogendam, L.*, Koopman, J. E.*, van Kooij, Y. E., Feitz, R., Hundepool, C. A., Zhou, C., Slijper, H. P., Selles, R. W., Wouters, R. M., The Hand-Wrist Study Group (2022). *Clinical orthopaedics and related research*, *480*(6), 1152–1166.

* Both authors contributed equally

ABSTRACT

Background

Patient-reported outcome measures (PROMs) are frequently used to assess treatment outcomes for hand and wrist conditions. To adequately interpret these outcomes, it is important to determine whether a statistically significant change is also clinically relevant. For this purpose, the minimally important change (MIC) was developed, representing the minimal within-person change in outcome that patients perceive as a beneficial treatment effect. Prior studies demonstrated substantial differences in MICs between condition-treatment combinations, suggesting that MICs are context-specific and cannot be reliably generalized. Hence, a study providing MICs for a wide diversity of condition-treatment combinations for hand and wrist conditions will contribute to more accurate treatment evaluations.

Questions/purposes

(1) What are the MICs of the most frequently used PROMs for common condition-treatment combinations of hand and wrist conditions? (2) Do MICs vary based on the invasiveness of the treatment (nonsurgical treatment or surgical treatment)?

Methods

This study is based on data from a longitudinally maintained database of patients with hand and wrist conditions treated in one of 26 outpatient clinics in the Netherlands between November 2013 and November 2020. Patients were invited to complete several validated PROMs before treatment and at final follow-up. All patients were invited to complete the VAS for pain and hand function. Depending on the condition, patients were also invited to complete the Michigan Hand Outcomes Questionnaire (finger and thumb conditions), the Patient-rated Wrist/Hand Evaluation (wrist conditions), or the Boston Carpal Tunnel Questionnaire (nerve conditions). Additionally, patients completed the validated Satisfaction with Treatment Result Questionnaire at final follow-up. Final follow-up timepoints were 3 months for nonsurgical and minor surgical treatment (including, trigger finger release), and 12 months for major surgical treatment (such as, trapeziectomy).

Our database included 55,651 patients, of whom we excluded 1528 patients who only required diagnostic management, 25,099 patients who did not complete the Satisfaction with Treatment Result Questionnaire, 3509 patients with missing data in the PROM of interest at baseline or follow-up, and 1766 patients who were part of condition-treatment combinations with less than 100 patients. The final sample represented 43% (23,749) of all patients and consisted of 36 condition-treatment combinations. In this final sample, 26% (6179) of patients were managed nonsurgically, and 74% (17,570) were managed surgically. Patients had a mean ± SD age of 55 ± 14 years, and 66% (15,593) of patients were women. To estimate the MIC, we used two anchor-based methods (the anchor mean change and the

MIC predict method), which were triangulated afterwards to obtain a single MIC. Applying this method, we calculated the MIC for 36 condition-treatment combinations, comprising 22 different conditions, and calculated the MIC for combined nonsurgical and surgical treatment groups. To examine whether the MIC differs between nonsurgical and surgical treatments, we performed a Wilcoxon signed rank test to compare the MICs of all PROM scores between nonsurgical and surgical treatment.

Results

We found a large variation in triangulated MICs between the condition-treatment combinations. For example, for nonsurgical treatment of hand OA, the MICs of VAS pain during load clustered around 10 (interquartile range 8 to 11), for wrist osteotomy/carpectomy it was around 25 (IQR 24 to 27), and for nerve decompression it was 21. Additionally, the MICs of the MHQ total score ranged from 4 (nonsurgical treatment of CMC1 OA) to 15 (trapeziectomy with LRTI and bone tunnel), for the PRWHE total score it ranged from 2 (nonsurgical treatment of STT OA) to 29 (release of first extensor compartment), and for the BCTQ SSS it ranged from 0.44 (nonsurgical treatment of carpal tunnel syndrome) to 0.87 (carpal tunnel release). An overview of all MIC values is available in a freely available online application at: https://analyse.equipezorgbedrijven.nl/shiny/mic-per-treatment/. In the combined treatment groups, the triangulated MIC values were lower for nonsurgical treatments than for surgical treatment ($p < 0.001$). The MICs for nonsurgical treatment can be approximated to be one-ninth (IQR 0.08 to 0.13) of the scale (approximately 11 on a 100-point instrument), while surgical treatment had MICs that were approximately one-fifth (IQR 0.14 to 0.24) of the scale (approximately 19 on a 100-point instrument).

Conclusion

MICs vary between condition-treatment combinations and differ depending on the invasiveness of the intervention. Patients receiving a more invasive treatment have higher treatment expectations, may experience more discomfort from their treatment, or may feel that the investment of undergoing a more invasive treatment should yield greater improvement, leading to a different perception of what constitutes a beneficial treatment effect.

Clinical Relevance

Our findings indicate that the MIC is context-specific and may be misguiding if applied inappropriately. Implementation of these condition-specific and treatment-specific MICs in clinical research allows for a better study design to achieve more accurate treatment evaluations. Consequently, this could aid clinicians in better informing patients about the expected treatment results and facilitate shared decision-making in clinical practice. Future studies may focus on adaptive techniques to achieve individualized MICs, which may ultimately aid clinicians in selecting the optimal treatment for individual patients.

INTRODUCTION

In recent years, the use of patient-reported outcome measures (PROMs) has become standard practice in clinical research and daily clinics for interpreting treatment results from the patient's perspective [28]. Unfortunately, clinicians are often faced with evidence that is challenging to interpret clinically as conclusions about treatment effects are often made from a statistical point of view [15, 16, 23, 27]. Statistically significant changes do not provide information about the magnitude of a treatment effect and therefore may not be meaningful to patients or clinicians. To interpret whether a statistically significant treatment effect is also clinically relevant, the minimally important change (MIC) concept is essential [23, 30, 40]. The MIC refers to the smallest change from baseline to post-treatment that patients perceive as important. Although often used interchangeably with the minimum clinically important difference (MCID), the MCID indicates the importance of a difference in outcomes between treatment groups (such as, fasciectomy versus collagenase) and thus represents a distinct entity [23, 30, 40].

In hand surgery, the MIC has been determined for multiple PROMs, including the Michigan Hand outcomes Questionnaire (MHQ) [13, 18, 20-22, 35], the Patient-rated Wrist/Hand Evaluation (PRWHE) [31, 36], and Boston Carpal Tunnel Questionnaire (BCTQ) [21, 25, 32]. However, we noticed in previous literature that MIC values for each PROM differ substantially not only between hand conditions, but also between different treatments for the same hand condition. For example, the MIC for the MHQ total score ranges from 9 for patients undergoing trigger finger release [13] to 18 for patients after proximal interphalangeal joint arthroplasty [22]. In addition, patients with carpal tunnel syndrome treated surgically yielded a higher MIC than patients receiving steroid injections [32].

Rationale

These findings suggest that MICs are context-specific and cannot always be reliably generalized to other condition-treatment combinations. For example, as treatments differ in invasiveness, rehabilitation periods, and experienced discomfort by patients, it is plausible that patients undergoing surgical treatment require a larger improvement before being satisfied compared with undergoing nonsurgical treatment. Hence, a study providing specific MICs for various condition-treatment combinations for hand and wrist conditions will contribute to more accurate treatment evaluations in clinical research. As a result, this could aid clinicians in better informing patients about the expected treatment results, which may facilitate shared decision-making in clinical practice.

Therefore, we asked: (1) What are the MICs of the most frequently used PROMs for common condition-treatment combinations of hand and wrist conditions? (2) Do MICs vary based on the invasiveness of the treatment (nonsurgical treatment or surgical treatment)?

PATIENTS AND METHODS

Study Design

This study is based on data from a longitudinally maintained database of patients with hand and wrist conditions (the Hand-Wrist Study Cohort), reported according to the Strengthening the Reporting of Observational Studies in Epidemiology statement [41]. The cohort and data collection [34] and their use in daily clinical care [11] have been described in more detail.

Setting

Between November 2013 and November 2020, we collected data at Xpert Clinics Hand and Wrist Care and Xpert Clinics Hand Therapy. Xpert Clinics currently comprises 26 locations, 23 European Board-certified hand surgeons, and more than 150 hand therapists. As part of routine outcome measurements, we invited all patients to complete PROMs before and at fixed timepoints after treatment based on the measurement track [34, 43].

Participants

For each outcome, we included all condition-treatment combinations consisting of at least 100 patients with data for the PROMs of interest at baseline and at the final follow-up examination. For nonsurgical treatments (such as hand therapy for thumb base osteoarthritis or a steroid injection for trigger finger) and minor surgical treatments (such as trigger finger release or carpal tunnel release), the final follow-up was at 3 months post-treatment. For major surgical treatments (for example, trapeziectomy or proximal row carpectomy), the final follow-up was 12 months post-treatment (Supplementary Table 1; supplemental materials are available with the online version of *CORR*®).

To prevent ceiling effects, we excluded patients with a baseline score of 90 or more points for the MHQ total or subscale scores or VAS pain, at most 10 points for the PRWHE total score or VAS hand function, 5 points or less for the PRWHE pain and hand function score, or no more than 1.4 points for the BCTQ subscales because this may result in an underestimation of the MIC [18]. The exact numbers of included patients for each PROM subscale and each condition-treatment combination can be found in our freely available online application at: https://analyse.equipezorgbedrijven.nl/shiny/mic-per-treatment/.

During the study period, 55,651 patients were treated for a hand or wrist condition. We excluded 1528 patients who only required diagnostic management (such as, diagnostic wrist arthroscopy), 25,099 patients who did not complete the Satisfaction with Treatment Result Questionnaire, 3509 patients with missing data in the PROM of interest at baseline or follow-up, and 1766 patients who were part of condition-treatment combinations with less than 100 patients. Our final sample represented 43% (23,749 of 55,651) of all patients and

consisted of 36 condition-treatment combinations (Fig. 1). Of these, 26% (6179 of 23,749) received nonsurgical treatment and 74% (17,570 of 23,749) received surgical treatment (Table 1). The mean ± SD age of the nonsurgically managed patients was 57 ± 15 years, and 73% (15,593) were women. The surgically managed patients had a mean ± SD age of 55 ± 14 years and 63% (11,114) were women.

Variables and Measurements

As part of routine follow-up, we invited all patients to complete the VAS for pain and hand function, regardless of the condition-treatment combination. Additionally, patients were invited to complete a more disease-specific PROM (the MHQ, the PRWHE, or the BCTQ). This was dependent on the condition-treatment combination and associated measurement track [34]. In brief, wrist conditions were assessed with the PRWHE, finger and thumb conditions with the MHQ, and nerve conditions with the BCTQ.

We used the VAS to examine pain (scale of 0-100; higher scores indicate more pain) and function (scale of 0-100; higher scores indicate poorer function). The VAS for pain was measured for three situations: pain at rest, pain during physical load, and average pain during the past week. The VAS has high test-retest reliability, good ability to detect change, and acceptable concurrent validity [12].

The MHQ consists of six domains (overall hand function, work performance, activities of daily living, pain, aesthetics, and satisfaction with hand function), each with a score ranging from 0 to 100 [4]. Higher scores indicate better performance, except for the subscale of pain. For interpretability, we reverted the pain subscale such that higher scores indicate less pain. We analyzed scores of the affected hand. The MHQ has been shown to have a high test-retest reliability, internal consistency, internal validity, and good responsiveness to change [4, 35].

The PRWHE is a validated questionnaire [19] with high internal consistency and reliability [26]. The PRWHE assesses the domains of pain and function, with scores ranging from 0 to 50 (higher scores indicate worse outcomes). The total score is calculated as the sum of both domains.

The BCTQ is a validated questionnaire comprising two domains: the Symptom Severity Scale and Functional Status Scale, with scores ranging from 1 to 5 (higher scores indicate more complaints) [17]. The BCTQ has good validity, reliability, and responsiveness [14].

To assess satisfaction with treatment results, we invited all patients to answer an additional question from the Satisfaction with Treatment Result Questionnaire: “How satisfied are you with the treatment result thus far? [7]. Responses were limited to one of the following items, scored on a Likert scale: excellent, good, fair, moderate, or poor. This questionnaire has recently been reported to be reliable and has good construct validity [7].

Primary and Secondary Study Outcomes

The primary aim of this study was to determine the MICs of four PROMs for a variety of condition-treatment combinations of hand and wrist conditions. To achieve this, we estimated the MICs of all PROMs for each condition-treatment combination using two anchor-based methods: the anchor mean change method and the MIC predict method [38, 39]. The weighted mean of both methods was considered to represent the MIC.

Our secondary aim was to examine whether the MIC differs based on the invasiveness of the treatment (nonsurgical treatment or surgical treatment). To achieve this, we performed a Wilcoxon signed rank test to compare the MICs of the PROM scores of nonsurgical and surgical treatments.

Ethical Approval

Ethical approval for this study was obtained from the Erasmus MC University Medical Center, Rotterdam, the Netherlands (MEC-2018-1088).

Study Size and Statistical Methods

To our knowledge, there are no recommendations regarding sample size for calculating an MIC. However, we considered that 100 patients would be sufficient to calculate a condition and treatment-specific MIC. In general, the MIC can be determined with distribution-based, anchor-based, or qualitative methods [33, 37]. Although the ideal method is still under discussion, anchor-based methods are the most frequently used and preferred approaches [15, 29, 33, 40]. We used two anchor-based methods to determine the MICs for each PROM: the anchor mean change method and the MIC predict method [38, 39]. For both methods, we used satisfaction with treatment result as an anchor.

Anchor mean change methods determine the MIC based on the group of patients reporting minimal improvement on the anchor question [29]. Using the satisfaction with the treatment result as anchor question, the response option fair was considered a minimal improvement. Hence, the MIC was defined as the mean change in the PROM of interest of patients rating their satisfaction with treatment results as fair in the anchor mean change method. To determine whether the anchor was suitable for further analyses, we calculated the Spearman correlation coefficient between the anchor question and change on the PROM of interest. Following current standards, an absolute correlation at least 0.3 was considered sufficient [29]. Because MIC values depend on baseline values [1], we also reported the baseline values of all PROMs before nonsurgical treatment, minor surgery, and major surgery (Table 2).

The MIC predict method is a receiver operating characteristic method that provides more accurate estimates when the groups of satisfied and dissatisfied patients are not equal in size [38, 39]. Because receiver operating characteristic curve methods require

dichotomization of the anchor question, patients rating their satisfaction with the treatment results as fair, good, or excellent were classified as satisfied. In contrast, we classified patients as dissatisfied if they rated their satisfaction as moderate or poor. We used the receiver operating characteristic curve method to determine the change in score on the PROM of interest that distinguished between satisfied and dissatisfied patients with the highest sensitivity and specificity, based on the Youden index [45]. Discriminative ability was considered sufficient if the area under the curve was at least 0.75 [10]. As mentioned, the MIC predict method allows for correction of imbalance in group sizes of satisfied and dissatisfied patients, because unequal group sizes may lead to biased MIC estimates [38, 39]. In our study, a larger proportion of patients were satisfied with their treatment (78% (4832 of 6179) for nonsurgical treatment and 86% (15,077 of 17,570) for surgical treatment). We therefore corrected the number of satisfied patients using logistic regression analysis, as described by Terluin et al. [38, 39]. To obtain a single MIC value, we triangulated the MIC estimates of both methods, assigning more weight (2:1) to the MIC predict method. The weighted mean of both methods was determined to represent the MIC.

To examine whether the MIC differs based on the invasiveness of the treatment, we compared the MICs of all PROM scores that were available for both nonsurgical and surgical treatment using a Wilcoxon signed rank test. We performed a nonresponder analysis to compare the characteristics of patients who completed all questionnaires of interest (47% (25,515 of 54,123) responders) and patients who did not (53% (28,608 of 54,123) nonresponders). For normally distributed continuous variables, we used t-tests. For non-normally distributed continuous variables, we used the Wilcoxon test. Chi-square tests were used to compare categorical variables. Additionally, effect sizes for differences were calculated. We calculated the Cohens d for continuous variables and the Cliff delta for categorical variables. In the nonresponder analysis, we found differences in the type of treatment, age, gender, and duration of symptoms (Supplementary Table 2; supplemental materials are available with the online version of CORR®), with absolute effect sizes ranging from 0.00 to 0.14, indicating very small effects [5]. All analyses were performed using R statistical software, version 4.0.1 (R core team). A p-value smaller than 0.05 was considered statistically significant.

RESULTS

MICs for Common Condition-treatment Combinations

We observed substantial variation in the MICs for different condition-treatment combinations. For example, the MICs of VAS pain during load for nonsurgical treatment of hand OA clustered around 10 (interquartile range 8 to 11), for wrist osteotomy/carpectomy the MIC was around 25 (IQR 24 to 27), and for nerve decompression it was 21 (Table 3).

Additionally, the MICs of the MHQ total score ranged from 4 (nonsurgical treatment of CMC1 OA) to 15 (trapeziectomy with LRTI and bone tunnel) (Table 4), for the PRWHE total score it ranged from 2 (nonsurgical treatment of STT OA) to 29 (release of first extensor compartment) (Table 5), and for the BCTQ SSS it ranged from 0.44 (nonsurgical treatment of carpal tunnel syndrome) to 0.87 (carpal tunnel release) (Table 6).

MIC values per calculation method, and triangulated MIC values are freely available in an online application at https://analyse.equipezorgbedrijven.nl/shiny/mic-per-treatment/. This online application allows users to select a treatment for which all MIC values and more details of the MIC calculation for all available PROM subscales are shown.

MICs Vary with the Invasiveness of Treatment

We found that MICs among the combined treatment groups were lower for non-surgical treatment compared to surgical treatment ($p<0.001$). The MICs for nonsurgical treatment can be approximated to be one-ninth (IQR 0.08 to 0.13) of the scale (approximately 11 on a 100-point instrument), while surgical treatment had MICs that were approximately one-fifth (IQR 0.14 to 0.24) of the scale (approximately 19 on a 100-point instrument).

DISCUSSION

Because the use of PROMs has become standard practice in clinical research, further insight into MICs for specific conditions and treatments is essential to accurately interpret whether a change is clinically relevant. In this large, multicenter study, we were able to provide condition-specific and treatment-specific MICs of four commonly used PROMs for hand and wrist conditions. This study shows that MICs differ between condition-treatment combinations. Furthermore, we found higher MICs for surgical treatments than for nonsurgical treatments. These findings indicate that the MIC is context-specific and that it may be misleading if applied inappropriately. Implementation of these condition-specific and treatment-specific MICs in clinical research allows for a better study design to achieve more accurate treatment evaluations. Consequently, this could aid clinicians in better informing patients about the expected treatment results and facilitate shared decision-making in clinical practice.

Limitations

This study also has several limitations. One limitation is that the observational design of this study was associated with a large proportion of missing data. However, although 53% (28,608 of 54,123) of patients did not complete all questionnaires of interest, our nonresponder analysis indicated that the differences between responders and nonresponders

had very small effect sizes. Hence, we are confident that these small differences did not affect our findings.

A second limitation of the observational design with routine outcome measurement data is that the actual treatment may have deviated from these protocols based on the clinician's expertise and patient's preference despite our standardized treatment protocols. For example, there may be differences between patients in the number of hand therapy sessions or steroid injections. Although readers should be aware that the MICs we provided may be influenced by some variation in treatment strategies, these deviations are highly representative for actual daily practice, resulting in more generalizable MIC values.

Third, we only performed analyses on the effect of treatment invasiveness on the MIC. It is plausible that MICs also vary depending on cultural or sociodemographic characteristics [2]. To address this, two previous studies evaluated factors contributing to differences in MICs of several questionnaires in patients with adult spinal deformity [3, 46]. Interestingly, they demonstrated that MICs did not vary based on age or sex, whereas they found substantial differences based on the baseline severity [3]. Moreover, consistent with our findings, they found considerably higher MICs for surgical treatment than nonsurgical treatment on all questionnaires, suggesting that the invasiveness of the treatment may be a more influential factor than the sociodemographic characteristics [46].

Fourth, it is well-known that MICs can vary depending on the method used [24]. In this study, we only used anchor-based methods. Although the ideal method is still under discussion, anchor-based methods are generally preferred as these take relevant changes from the patient's perspective into account [15, 29, 33, 40]. We determined the MIC with an anchor question assessing satisfaction with the treatment result. Some authors recommend using an anchor question that is based on a scale rating the change in outcome measured on that specific outcome domain [33]. However, the MIC is defined as "the smallest change in an outcome measure that patients perceive as important" [9]. Considering this definition, we believe that satisfaction with the treatment result is an accurate measure of patients' perceived improvement. This is also in line with prior studies in hand surgery, in which patient-reported satisfaction was commonly used as an anchor question [13, 18, 20, 21, 35]. However, by using this anchor question, we were unable to determine the MIC for all subdomains of all PROMs for all hand conditions because of a low correlation with the anchor question. This was primarily the case for the MHQ subdomain of aesthetics. Because the most profound complaints in these specific hand conditions are pain and difficulties with hand function, this subdomain might be irrelevant to these specific conditions, accounting for the low correlation with satisfaction.

Finally, we were unable to determine the smallest detectable change because patients were only asked to complete the PROMs once at the last measurement post-treatment. Some MIC values we reported are low, and it is unclear whether these MIC values exceed the

smallest detectable change for this specific outcome. This mostly occurred for nonsurgical treatments in combination with outcome domains that may not be the most relevant for these patients and where little improvement may be expected with treatment (for example, the VAS during load for nonsurgical treatment of cubital tunnel syndrome and MHQ hand function for nonsurgical treatment of CMC1 osteoarthritis). However, as MICs should mainly be used to interpret clinically relevant changes on a group level, comparisons with the smallest detectable changes may be less important, because random measurement error may be canceled out at the group level [2].

MICs for Condition-treatment Combinations

We found large differences in MICs for different condition-treatment combinations, indicating that MICs cannot be reliably applied to other condition-treatment combinations. This is in line with prior literature, demonstrating substantial variation in MICs for different condition-treatment combinations. However, MICs for specific condition-treatment combinations are comparable to those reported in literature, including MICs for the BCTQ in patients undergoing carpal tunnel release [6] or cubital tunnel release [21], and for the MHQ in patients with a PIP prosthesis [22] or trigger finger release [13].

The comprehensive overview of MIC values provided in this study contributes to the application of more accurate MICs in clinical research, resulting in better treatment evaluations which may improve patient counseling and management strategies [40]. For example, MICs can be used as a threshold to determine the percentage of patients reaching the MIC in clinical research. Insight into these percentages could aid clinicians in informing patients about the expected treatment results and may facilitate shared decision-making in clinical practice. However, although the percentage of patients reaching the MIC will probably be correct on a group level, this threshold may not apply to individual patients as all patients have an individual threshold of what they consider an important change. Therefore, it is highly recommended to use the MIC as a probabilistic value, rather than a deterministic cutoff, when applied to individual patients. Future studies may focus on adaptive techniques to achieve individualized MICs as proposed by Zhou et al. [47], which may ultimately aid clinicians in selecting the optimal treatment for individual patients.

MICs Varies with the Invasiveness of Treatment

The finding that MICs differ depending on the invasiveness of the intervention indicates that a more invasive treatment requires a larger improvement in these subdomains for patients to experience satisfaction with the treatment result. Patients receiving surgical treatment might experience more disability and discomfort from their treatment than patients receiving nonsurgical treatment. Consequently, patients treated nonsurgically may be more satisfied with their treatment result when they experience only a small improvement, resulting in a lower MIC.

Although it is plausible that the treatment invasiveness accounts for the variation in MICs, one might also suggest that differences in the groups being compared cause the variation. For example, the follow-up periods differed from 3 months for nonsurgical and minor surgical treatments to 12 months for major surgical treatments. A study in patients receiving decompression of ulnar neuropathy demonstrated that the MIC of the BCTQ was lower at 3 months than at 6 months [21]. However, they found no difference in the MIC at 6 and 12 months postoperatively, suggesting that the MIC is stable if the functional recovery period is reached. The follow-up periods we used in this study align with the ICHOM standard set for hand and wrist conditions, representing the clinical endpoints of the specific treatments [43]. Hence, despite differences in the follow-up periods between treatments, we believe these clinical endpoints are most suitable for assessing the final treatment effects.

In addition, another factor potentially contributing to these differences is variation in baseline characteristics [8, 42]. A study comparing the baseline characteristics of patients with CMC1 osteoarthritis treated surgically and those treated nonsurgically found that surgically treated patients had worse baseline PROM scores, worse illness perceptions and catastrophization, and higher treatment expectations [44]. Although we only observed small differences in baseline scores between surgical and nonsurgical treatment groups, we did not take psychological characteristics and treatment expectations into account. It is plausible that patients with a worse psychological profile or higher treatment expectations may need more improvement in order to be satisfied with treatment results, resulting in a higher MIC. Future studies may examine factors contributing to these differences in MICs to improve expectation management for individual patients.

Conclusion

MICs differ between diagnosis-treatment combinations and particularly differ depending on the invasiveness of the intervention. These findings indicate that the MIC is context-specific and may be misguiding if applied inappropriately. Hence, implementation of these condition-specific and treatment-specific MICs in clinical research allows for a better study design to achieve more accurate treatment evaluations. Consequently, this could aid clinicians in better informing patients about the expected treatment results and facilitate shared decision-making in clinical practice. Future studies may focus on adaptive techniques to achieve individualized MICs reflecting clinically relevant change to individuals instead of groups, which may ultimately aid clinicians in selecting the optimal treatment for individual patients.

ACKNOWLEDGMENTS

The authors thank all patients who participated and allowed their data to be anonymously used for the present study. In addition, we acknowledge all caregivers and personnel of Xpert Clinics, Handtherapie Nederland and Equipe Zorgbedrijven for assisting in the routine outcome measurements that are the basis for this manuscript.

GROUP AUTHORS

Members of the Hand-Wrist Study Group include: Dirk-Johannes Jacobus Cornelis van der Avoort MD, Richard Arjen Michiel Blomme MD, Herman Luitzen de Boer MD, Gijs Marijn van Couwelaar MD, Jan Debeij MD PhD, Jak Dekker MSc, Alexandra Fink PT, Klazina Paulina de Haas MD, Kennard Harmsen MD, Steven Eric Ruden Hovius MD PhD, Rob van Huis PT, Richard Koch MD, Alexander Kroeze MD, Thybout Matthias Moojen MD PhD, Mark Johannes Willem van der Oest PhD, Pierre-Yves Alain Adriaan Pennehouat PT, Willemijn Anna de Ridder PT MSc, Johannes Pieter de Schipper MD, Karin Schoneveld PT MSc, Berbel Jeannee Rinel Sluijter MD PhD, Jeronimus Maria Smit MD PhD, Xander Smit MD PhD, John Sebastiaan Souer MD PhD, Marloes Hendrina Paulina ter Stege MSc, Johannes Frederikes Maria Temming MD, Joris Sebastiaan Teunissen BSc, Jeroen Hein van Uchelen MD PhD, Joris Jan Veltkamp PT, Guus Maarten Vermeulen MD PhD, Erik Taco Walbeehm MD PhD, Oliver Theodor Zöphel MD PhD, Jelle Michiel Zuidam

REFERENCES

1. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. J Clin Epidemiol. 2017;82:128-136.
2. Arima H, Carreon LY, Glassman SD, et al. Cultural variations in the minimum clinically important difference thresholds for SRS-22R after surgery for adult spinal deformity. Spine Deform. 2019;7:627-632.
3. Bahadir S, Yuksel S, Ayhan S, et al. Variation of minimum clinically important difference by age, gender, baseline disability, and change of direction in adult spinal deformity population: Is it a constant value? World Neurosurg. 2021;146:e1171-e1176.
4. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand outcomes Questionnaire. J Hand Surg Am. 1998;23:575-587.
5. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
6. De Kleermaeker F, Boogaarts HD, Meulstee J, Verhagen WIM. Minimal clinically important difference for the Boston Carpal Tunnel Questionnaire: new insights and review of literature. J Hand Surg Eur Vol. 2019;44:283-289.
7. De Ridder WA, van Kooij YE, Vermeulen GM, et al. Test-retest reliability and construct validity of the Satisfaction with Treatment Result Questionnaire in patients with hand and wrist conditions: A prospective study. Clin Orthop Relat Res. 2021;479:2022-2032.
8. de Vet HC, Foumani M, Scholten MA, et al. Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention. J Clin Epidemiol. 2015;68:518-524.
9. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes. 2006;4:54.
10. Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. Cjem. 2006;8 1:19-20.
11. Feitz R, van Kooij YE, Ter Stege MHP, et al. Closing the loop: a 10-year experience with routine outcome measurements to improve treatment in hand surgery. EFORT Open Rev. 2021;6:439-450.
12. Kahl C, Cleland JA. Visual analogue scale, numeric pain rating scale and the McGill pain Questionnaire: an overview of psychometric properties. Physical Therapy Reviews. 2005;10:123-128.
13. Koopman JE, van Kooij YE, Selles RW, et al. Determining the minimally important change of the Michigan Hand outcomes Questionnaire in patients undergoing trigger finger release. J Hand Ther. [Published online ahead of print July 24, 2021]. DOI: 10.1016/j.jht.2021.06.003.
14. Leite JC, Jerosch-Herold C, Song F. A systematic review of the psychometric properties of the Boston Carpal Tunnel Questionnaire. BMC Musculoskelet Disord. 2006;7:78.
15. Leopold SS, Porcher R. Editorial: The minimum clinically important difference - The least we can do. Clin Orthop Relat Res. 2017;475:929-932.
16. Leopold SS, Porcher R. Editorial: Threshold p values in orthopaedic research - We know the problem. What is the solution? Clin Orthop Relat Res. 2018;476:1689-1691.

17. Levine DW, Simmons BP, Koris MJ, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. J Bone Joint Surg Am. 1993;75:1585-1592.

18. London DA, Stepan JG, Calfee RP. Determining the Michigan Hand outcomes Questionnaire minimal clinically important difference by means of three methods. Plast Reconstr Surg. 2014;133:616-625.

19. MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: a reliable and valid measurement tool. J Orthop Trauma. 1998;12:577-586.

20. Maia MV, de Moraes VY, Dos Santos JB, Faloppa F, Belloti JC. Minimal important difference after hand surgery: a prospective assessment for DASH, MHQ, and SF-12. SICOT J. 2016;2:32.

21. Malay S, Group SUNS, Chung KC. The minimal clinically important difference after simple decompression for ulnar neuropathy at the elbow. J Hand Surg Am. 2013;38:652-659.

22. Marks M, Hensler S, Wehrli M, Schindele S, Herren DB. Minimal important change and patient acceptable symptom state for patients after proximal interphalangeal joint arthroplasty. J Hand Surg Eur Vol. 2019;44:175-180.

23. Marks M, Rodrigues JN. Correct reporting and interpretation of clinical data. J Hand Surg Eur Vol. 2017;42:977-979.

24. Mouelhi Y, Jouve E, Castelli C, Gentile S. How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. Health Qual Life Outcomes. 2020;18:136.

25. Ozyurekoglu T, McCabe SJ, Goldsmith LJ, LaJoie AS. The minimal clinically important difference of the Carpal Tunnel Syndrome Symptom Severity Scale. J Hand Surg Am. 2006;31:733-738; discussion 739-740.

26. Packham T, MacDermid JC. Measurement properties of the Patient-Rated Wrist and Hand Evaluation: Rasch analysis of responses from a traumatic hand injury population. J Hand Ther. 2013;26:216-223; quiz 224.

27. Page P. Beyond statistical significance: clinical interpretation of rehabilitation research literature. Int J Sports Phys Ther. 2014;9:726-736.

28. Porter ME. What is value in health care? N Engl J Med. 2010;363:2477-2481.

29. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61:102-109.

30. Rodrigues JN. Different terminologies that help the interpretation of outcomes. J Hand Surg Eur Vol. 2020;45:97-99.

31. Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. J Clin Epidemiol. 2004;57:1008-1018.

32. Schrier VJ, Gelfman R, Amadio PC. Minimal clinically important difference is lower for carpal tunnel syndrome patients undergoing injection versus surgery. J Hand Surg Eur Vol. 2020;45:90-92.

33. Sedaghat AR. Understanding the minimal clinically important difference (MCID) of patient-reported outcome measures. Otolaryngol Head Neck Surg. 2019;161:551-560.

34. Selles RW, Wouters RM, Poelstra R, et al. Routine health outcome measurement: Development, design, and implementation of the Hand and Wrist Cohort. Plast Reconstr Surg. 2020;146:343-354.

35. Shauver MJ, Chung KC. The minimal clinically important difference of the Michigan Hand outcomes Questionnaire. J Hand Surg Am. 2009;34:509-514.

36. Sorensen AA, Howard D, Tan WH, Ketchersid J, Calfee RP. Minimal clinically important differences of 3 patient-rated outcomes instruments. J Hand Surg Am. 2013;38:641-649.

37. Staunton H, Willgoss T, Nelsen L, et al. An overview of using qualitative techniques to explore and define estimates of clinically important change on clinical outcome assessments. J Patient Rep Outcomes. 2019;3:16.

38. Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. J Clin Epidemiol. 2017;83:90-100.

39. Terluin B, Eekhout I, Terwee CB, de Vet HC. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. J Clin Epidemiol. 2015;68:1388-1396.

40. Terwee CB, Peipert JD, Chapman R, et al. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res. 2021;30:2729-2754.

41. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61:344-349.

42. Wang YC, Hart DL, Stratford PW, Mioduski JE. Baseline dependency of minimal clinically important improvement. Phys Ther. 2011;91:675-688.

43. Wouters RM, Jobi-Odeneye AO, de la Torre A, et al. A standard set for outcome measurement in patients with hand and wrist conditions: Consensus by the International Consortium for Health Outcomes Measurement Hand and Wrist working group. J Hand Surg Am. 2021;46:841-855 e847.

44. Wouters RM, Vranceanu AM, Slijper HP, et al. Patients with thumb-base osteoarthritis scheduled for surgery have more symptoms, worse psychological profile, and higher expectations than nonsurgical counterparts: A large cohort analysis. Clin Orthop Relat Res. 2019;477:2735-2746.

45. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3:32-35.

46. Yuksel S, Ayhan S, Nabiyev V, et al. Minimum clinically important difference of the health-related quality of life scales in adult spinal deformity calculated by latent class analysis: is it appropriate to use the same values for surgical and nonsurgical patients? Spine J. 2019;19:71-78.

47. Zhou Z, Zhao J, Bisson LJ. Estimation of data adaptive minimal clinically important difference with a nonconvex optimization procedure. Stat Methods Med Res. 2020;29:879-893.

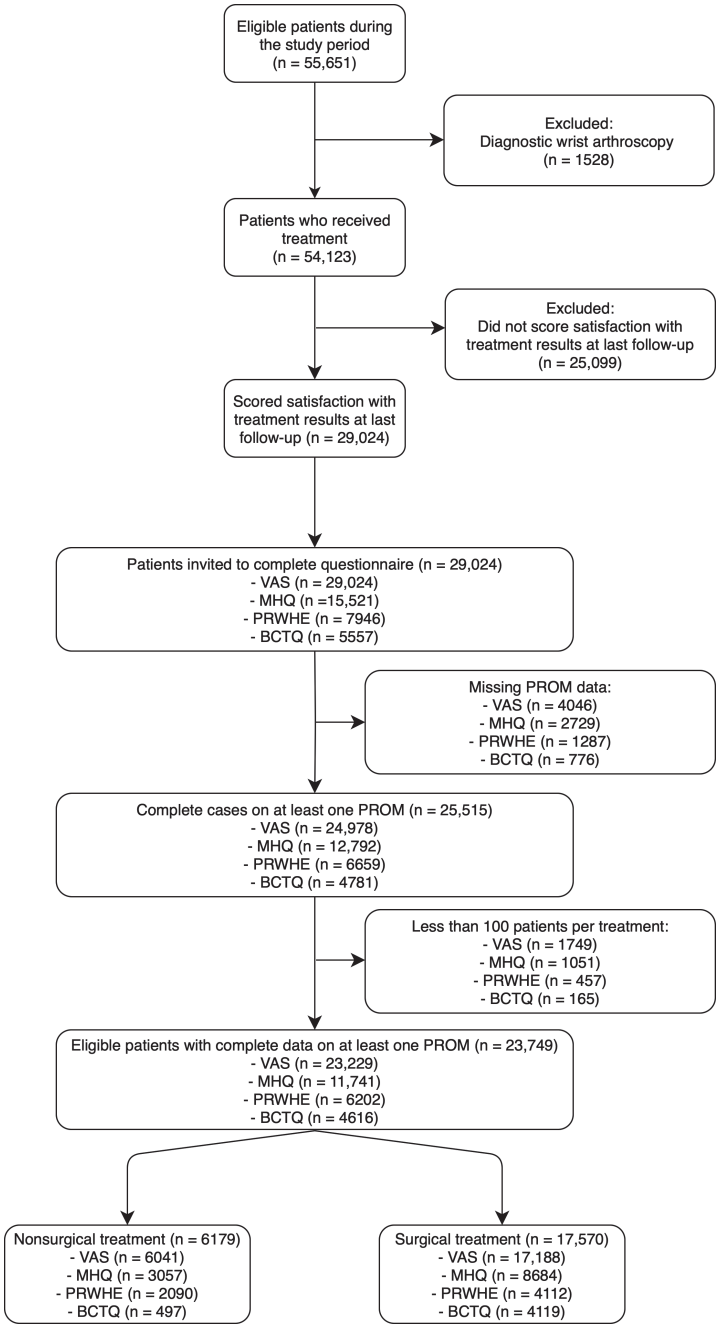


Figure 1. This flowchart shows the patients who were included in our study. As patients were invited to complete the VAS and one other PROM most suitable for their diagnosis, the total number indicated at the top of the box is less than the total number of PROM responses.

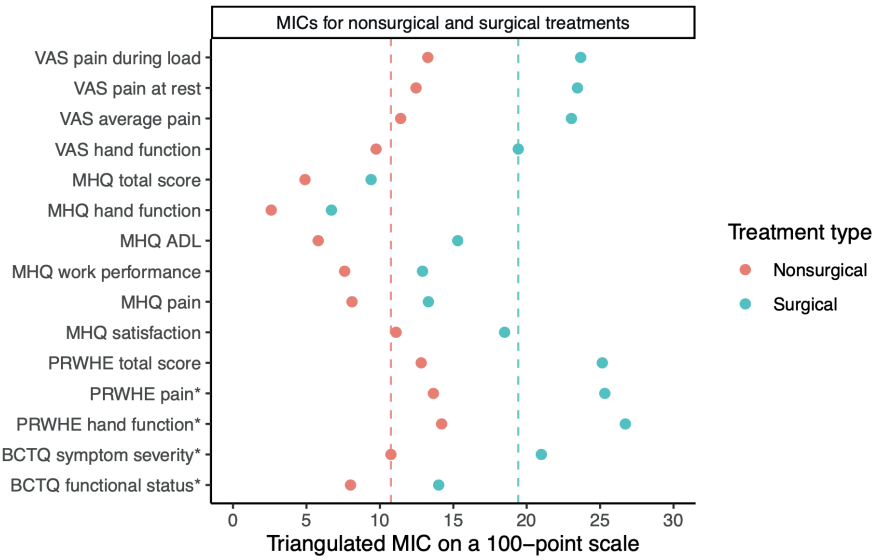


Figure 2. This figure shows triangulated MIC values for the VAS, MHQ, PRWHE, and BCTQ, categorized by treatment group (nonsurgical and surgical). Results are presented as the triangulated MIC value (dots). For interpretability, we converted the PRWHE and BCTQ subscales (depicted with an asterisk) to a 100-point scale. The dashed lines represent the median MIC values for nonsurgical (red) and surgical (blue) treatments. Overall, this figure shows that MIC values are higher for surgical treatments (median 19 points; IQR 14-24) compared to nonsurgical treatments (median 11 points; IQR 8-13).

Table 1. Patient characteristics per treatment group for patients with complete data for at least one PROM and satisfaction with treatment results

Parameter	All included patients (n = 23,749)	Nonsurgical treatment (n = 6179)	Surgical treatment (n = 17,570)
Age in years	55 ± 14	57 ± 15	55 ± 14
Gender			
Women	66 (15,593)	73 (4479)	63 (11,114)
Duration of symptoms in months	11 (5-24)	6 (3-15)	12 (6-24)
Hand dominance			
Right	88 (20,888)	89 (5488)	88 (15,400)
Left	9 (2107)	8 (500)	9 (1607)
Both	3 (754)	3 (191)	3 (563)
Affected side			
Right	55 (12,977)	52 (3239)	55 (9738)
Left	43 (10,242)	40 (2445)	44 (7797)
Both	2 (530)	8 (495)	0.2 (35)
Occupational intensity			
Not employed	38 (8920)	35 (2134)	39 (6786)
Light (such as working in an office)	27 (6346)	28 (1713)	26 (4633)
Moderate (such as working in a shop)	24 (5760)	27 (1635)	23 (4125)
Heavy (such as working in construction)	11 (2723)	11 (697)	12 (2026)
Involved in a personal injury claim[a]			
Yes	2 (311)	2 (102)	2 (209)

Data presented as mean ± SD, % (n), or median (IQR).
[a]This information was available for a selection of patients (all included patients: n = 17,413, nonsurgical treatment: n = 5193, surgical treatment: n = 12,220); PROM = Patient-reported outcome measure.

Table 2. Baseline PROM scores per type of treatment are shown to provide a reference on symptom severity before treatment and the relativity of the MIC

Parameter	PROM range	Nonsurgical treatment, mean ± SD (n = 6179)	Surgical treatment, mean ± SD (n = 17,570)	Effect size
VAS pain during load	0-100	59 ± 25	55 ± 30	0.15
VAS pain at rest	0-100	35 ± 26	38 ± 28	-0.09
VAS average pain	0-100	49 ± 24	48 ± 27	0.07
VAS hand function	0-100	52 ± 25	50 ± 26	0.09
MHQ total score	0-100	62 ± 15	64 ± 18	-0.19
MHQ pain	0-100	50 ± 19	56 ± 24	-0.24
MHQ hand function	0-100	59 ± 18	62 ± 19	-0.12
MHQ work	0-100	63 ± 26	69 ± 28	-0.23
MHQ ADL	0-100	70 ± 20	74 ± 23	-0.20
MHQ aesthetics	0-100	82 ± 20	79 ± 20	0.17
MHQ satisfaction	0-100	46 ± 23	49 ± 26	-0.12
PRWHE total score	0-100	56 ± 21	60 ± 21	-0.20
PRWHE pain	0-50	30 ± 11	32 ± 10	-0.19
PRWHE hand function	0-50	26 ± 12	28 ± 12	-0.19
BCTQ symptom severity scale	1-5	2.6 ± 0.7	2.9 ± 0.7	-0.35
BCTQ functional status scale	1-5	2.3 ± 0.8	2.5 ± 0.8	-0.19

PROM = Patient Reported Outcome Measure; MIC = Minimum Important Change; VAS = Visual Analog Scale; MHQ = Michigan Hand outcomes Questionnaire; ADL = Activities of Daily Living; PRWHE = Patient Rated Wrist/Hand Evaluation; BCTQ = Boston Carpal Tunnel Questionnaire

Table 3. Triangulated MIC values per treatment (group) for VAS subscales pain during load, pain at rest, average pain, and hand function[a]

	VAS pain during load (0-100)	VAS pain at rest (0-100)	VAS average pain (0-100)	VAS hand function (0-100)
All	20	20	19	17
Nonsurgical treatment	13	12	11	10
Surgical treatment	23	23	23	19
Nonsurgical treatment of:				
Carpal tunnel syndrome	13	17	14	NA[b]
Cubital tunnel syndrome	-3	NA[c]	NA[c]	4
Tendinitis/tenosynovitis wrist	19	15	18	16
Trigger finger	13	12	10	10
Trigger thumb	17	11	15	10
CMC-1 OA	8	9	8	6
CMC-1 instability	14	13	13	14
STT OA	10	10	6	8
MCP/PIP/DIP OA	11	12	11	NA[b]
UCL/RCL/VP injury MCP/PIP/DIP	10	NA[c]	16	15
Mallet finger	NA[c]	NA[b]	NA[c]	12
Midcarpal instability/laxity	15	15	14	12
Minor surgical treatments:				
Carpal tunnel release	21	28	25	18
Cubital tunnel release	21	21	21	17
Release of the first extensor compartment	32	27	28	24
Trigger finger release	24	21	21	18
Trigger thumb release	28	23	25	22
Excision of volar wrist ganglion	22	23	21	21
Excision of dorsal wrist ganglion	24	24	21	16
Mucoid cyst excision finger	13	17	12	13
Percutaneous needle aponeurotomy (possibly with lipofilling)	12	14	14	NA[b]
Major surgical treatments:				
Corrective osteotomy distal radius	25	20	23	26
Ulna shortening osteotomy	24	22	22	24
TFCC reinsertion	25	22	22	22
Proximal row carpectomy	27	21	25	18
Osteosynthesis for non-union of scaphoid fracture	NA[b]	NA[c]	NA[b]	22
Pisiformectomy	25	20	23	21
Dorsal capsulodesis wrist (possibly combined with dorsal ganglion excision)	27	23	26	19
Three-ligament tenodesis (Brunelli)	23	17	19	19
Limited fasciectomy (possibly with skingraft)	14	18	14	19
Trapeziectomy with LTRI using the FCR (Weilby technique)	29	26	28	21

	VAS pain during load (0-100)	VAS pain at rest (0-100)	VAS average pain (0-100)	VAS hand function (0-100)
Trapeziectomy with LRTI using the FCR and bone tunnel (Burton-Pellegrini technique)	32	31	27	23
Other surgical treatments for CMC-1 OA	33	27	33	23
CMC-1 instability treated surgically	25	20	23	19
PIP prosthesis	30	NA[b]	NA[b]	13
UCL reinsertion MCP-1	25	NA[c]	NA[c]	23

[a] These scores represent the MIC per condition-treatment combination and for nonsurgical and surgical treatment overall.
[b] Insufficient correlation
[c] Insufficient number of patients
MIC = Minimum important change; CMC-1 = First carpometacarpal joint; OA = Osteoarthritis; STT = Scaphotrapeziotrapezoid joint; MCP = Metacarpal joint; PIP = Proximal interphalangeal joint; DIP = Distal interphalangeal joint; UCL = Ulnar collateral ligament; RCL = Radial collateral ligament; VP = Volar plate; TFCC = Triangular fibrocartilage complex; LRTI = Ligament reconstruction and tendon interposition; FCR = Flexor carpi radialis

Table 4. Triangulated MIC values per treatment (group) for MHQ total score and MHQ subscales

	MHQ total score (0-100)	MHQ hand function (0-100)	MHQ work (0-100)	MHQ ADL (0-100)	MHQ pain (0-100)	MHQ aesthetics (0-100)	MHQ satisfaction (0-100)
All	8	5	11	12	12	11	16
Nonsurgical treatment	5	3	8	6	8	NA[a]	11
Surgical treatment	9	7	13	15	13	12	20
Nonsurgical treatment of:							
Trigger finger	7	5	7	NA[a]	11	8	14
Trigger thumb	7	5	10	7	12	NA[b]	14
CMC-1 OA	4	1	6	4	7	NA[a]	9
CMC-1 instability	5	7	12	6	10	NA[b]	14
MCP/PIP/DIP OA	6	NA[a]	7	NA[a]	9	NA[a]	12
UCL/RCL/VP injury MCP/PIP/DIP	7	NA[a]	NA[a]	12	8	NA[a]	11
Mallet finger	6	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	12
Minor surgical treatments:							
Trigger finger release	8	8	10	14	13	11	19
Trigger thumb release	11	6	NA[a]	16	16	NA[a]	20
Mucoid cyst finger excision	4	1	NA[a]	NA[a]	6	10	6
Percutaneous needle aponeurotomy (possibly with lipofilling)	NA[a]	5	NA[a]	NA[a]	NA[a]	NA[a]	NA[a]
Major surgical treatments:							
Limited fasciectomy (possibly with skingraft)	5	4	16	8	8	11	14
Trapeziectomy with LRTI using the FCR (Weilby technique)	14	9	12	19	19	NA[a]	24

	MHQ total score (0-100)	MHQ hand function (0-100)	MHQ work (0-100)	MHQ ADL (0-100)	MHQ pain (0-100)	MHQ aesthetics (0-100)	MHQ satisfaction (0-100)
Trapeziectomy with LRTI using the FCR and bone tunnel (Burton-Pellegrini technique)	15	9	16	19	18	NA[b]	24
Other surgical treatments for CMC-1 OA	15	10	12	21	22	NA[b]	25
CMC-1 instability treated surgically	13	9	18	13	18	NA[b]	21
PIP prosthesis	13	10	NA[a]	12	24	8	21
UCL reinsertion MCP-1	14	NA[b]	NA[b]	NA[b]	21	NA[b]	19

[a]Insufficient correlation

[b]Insufficient number of patients

MIC = Minimum important change; MHQ = Michigan Hand outcomes Questionnaire; ADL = Activities of daily living; CMC-1 = First carpometacarpal joint; OA = Osteoarthritis; MCP = Metacarpal joint; PIP = Proximal interphalangeal joint; DIP = Distal interphalangeal joint; UCL = Ulnar collateral ligament; RCL = Radial collateral ligament; VP = Volar plate; LRTI = Ligament reconstruction and tendon interposition; FCR = Flexor carpi radialis

Table 5. Triangulated MIC values per treatment (group) for PRWHE total score and subscales

	PRWHE total score (0-100)	PRWHE pain score (0-50)	PRWHE function score (0-50)
All	20	10	11
Nonsurgical treatment	13	7	7
Surgical treatment	24	12	13
Nonsurgical treatment of:			
Tendinitis/tenosynovitis wrist	17	9	10
STT OA	2	2	1
Midcarpal instability/laxity	10	6	5
Minor surgical treatments:			
Release 1st extensor compartment	29	14	16
Excision of volar wrist ganglion	16	9	9
Excision of dorsal wrist ganglion	20	11	10
Major surgical treatments:			
Corrective osteotomy distal radius	26	11	15
Ulna shortening osteotomy	22	10	12
Open TFCC reinsertion	24	12	13
Proximal row carpectomy	21	12	10
Osteosynthesis for non-union of scaphoid fracture	26	12	15
Pisiformectomy	24	11	13
Dorsal capsulodesis wrist (possibly combined with dorsal ganglion excision)	21	11	11
Three ligament tenodesis (Brunelli)	22	12	12

MIC = Minimum important change; PRWHE = Patient-Rated Wrist/Hand Evaluation; STT = Scaphotrapeziotrapezoid joint; OA = Osteoarthritis; TFCC = Triangular fibrocartilage complex

Table 6. Triangulated MIC values per treatment (group) for BCTQ symptom severity scale and BCTQ functional status scale

	BCTQ symptom severity scale (1-5)	BCTQ functional status scale (1-5)
All	0.79	0.52
Nonsurgical treatment	0.43	0.32
Surgical treatment	0.84	0.56
Nonsurgical treatment of:		
Carpal tunnel syndrome	0.44	0.36
Cubital tunnel syndrome	NA[a]	NA[a]
Surgical treatment:		
Carpal tunnel release	0.87	0.57
Cubital tunnel release	0.60	0.47

[a]Insufficient number of patients
MIC = Minimally Important Change ; BCTQ = Boston Carpal Tunnel Questionnaire

3

ESTIMATING MEANINGFUL THRESHOLDS FOR MULTI-ITEM QUESTIONNAIRES USING ITEM RESPONSE THEORY

Terluin, B., Koopman, J. E., Hoogendam, L., Griffiths, P., Terwee, C. B., Bjorner, J. B. (2023). Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 32(6), 1819–1830.

My contribution to this publication was that I did the analyses to calculate the PASS for the MHQ pain score after trigger finger release, and I contributed to the writing and the revision of the manuscript. I have included this publication in my thesis, because the method we introduce here is used to calculate an overview of PASS values for several PROMs and hand-wrist conditions in our study “What are the Patient Acceptable Symptom State values of commonly used multi-item Patient-Reported Outcome Measures for 35 hand and wrist condition-treatment combinations?”, accounting for the methodological considerations we learned from the simulation.

ABSTRACT

Purpose

Meaningful thresholds are needed to interpret patient-reported outcome measure (PROM) results. This paper introduces a new method, based on item response theory (IRT), to estimate such thresholds. The performance of the method is examined in simulated datasets and two real datasets, and compared with other methods.

Methods

The IRT method involves fitting an IRT model to the PROM items and an anchor item indicating the criterion state of interest. The difficulty parameter of the anchor item represents the meaningful threshold on the latent trait. The latent threshold is then linked to the corresponding expected PROM score. We simulated 4500 item response datasets to a 10-item PROM, and an anchor item. The datasets varied with respect to the mean and standard deviation of the latent trait, and the reliability of the anchor item. The real datasets consisted of a depression scale with a clinical depression diagnosis as anchor variable and a pain scale with a patient acceptable symptom state (PASS) question as anchor variable.

Results

The new IRT method recovered the true thresholds accurately across the simulated datasets. The other methods, except one, produced biased threshold estimates if the state prevalence was smaller or greater than 0.5. The adjusted predictive modeling method matched the new IRT method (also in the real datasets) but showed some residual bias if the prevalence was smaller than 0.3 or greater than 0.7.

Conclusions

The new IRT method perfectly recovers meaningful (interpretational) thresholds for multi-item questionnaires, provided the data satisfy the assumptions for IRT analysis.

INTRODUCTION

The use of patient-reported outcome measures (PROMs) has become standard practice in clinical research and daily clinics due to the growing emphasis on patient-centered and value-based care. PROMs typically consist of multi-item questionnaires used to measure constructs (or "traits"), such as "depression" or "pain". However, because PROM scores are often continuous scores without intrinsic meaning, there is a need for (clinically) meaningful thresholds or cutoff points to facilitate interpretation. Examples of meaningful thresholds include a diagnostic cutoff point for depression, and a patient acceptable symptom state (PASS) threshold for pain. Determining a meaningful threshold on a questionnaire requires the comparison with an external criterion indicating the presence or absence of a meaningful "state" (e.g., clinical depression, or an acceptable symptom state). For clarity, we provide some terminology in Box 1.

Given that depression represents a continuous trait in the general population [1], the state clinical depression can be conceptualized as a level of depression above a certain threshold on this trait. Then, making a diagnosis of clinical depression can be seen as estimating a patient's level of depression, based on their history, and to determine whether this level is above or below the threshold of clinical depression [2]. In this example, the threshold is agreed upon by the psychiatric professional community.

The PASS represents a threshold of clinical importance beyond which patients consider their level of symptoms (e.g., pain) as acceptable [3]. A PASS threshold is typically determined using an "anchor" question like "Do you consider your current level of pain acceptable, yes or no?". The question assumes that patients compare their perceived level of pain to a personal threshold (or benchmark) of acceptability. This PASS threshold probably differs across individuals. Thus, the best group-level PASS estimate would be the mean of the individual PASS thresholds in a group of patients.

Given a continuous "test" variable (i.e., a variable holding the PROM scores) and a dichotomous "state" variable (i.e., a variable holding the state scores), the traditional method to determine a meaningful threshold or cutoff point is receiver operating characteristic (ROC) analysis. ROC analysis examines the sensitivity and specificity of all possible test scores with respect to their ability to classify subjects with respect to the meaningful state [4]. As a cutoff point, a test score can be selected based on its desired sensitivity and/or specificity, controlling the type and amount of misclassification. Often a so-called "best" or "optimal" cutoff point is chosen of which the difference between sensitivity and specificity is minimized (top-left criterion) or the sum of sensitivity and specificity is maximized (Youden criterion [5]; in large samples with normally distributed test scores both criteria identify the same threshold [6]). An optimal ROC threshold serves to classify subjects with the least amount of misclassification.

Box 1. Terminology

Trait: The construct of interest (e.g., depression or pain) that is intended to be measured by a PROM, usually a multi-item questionnaire. The construct itself is not directly observable, hence "latent". The latent trait is usually continuous. The PROM score provides an approximation of the true trait level. PROM scores are observed (i.e., manifest).

Perceived trait: The level of the latent trait as being perceived by the patient or by an observer (e.g., a clinician). The perceived trait is equal to the latent trait plus some random (measurement) error.

State (of interest): A clinically meaningful condition that is characterized by a minimum level of a trait of interest. Examples of meaningful states are clinical depression and acceptable symptom state.

Meaningful threshold: The minimum trait level above which a meaningful state is assumed to exist. The meaningful threshold can be thought of as a location on the latent trait (in which case the threshold is latent), or it can be thought of as a particular PROM score (in which case the threshold is manifest, and an approximation of the latent threshold). The term "cutoff point" can be used to indicate a manifest threshold of a PROM.

State assessment: The procedure used to determine whether or not a state of interest is present. The procedure is independent of the PROM of interest. Examples of state assessments are the making of a diagnosis of clinical depression by a trained professional, and the patient response to a targeted question (often called an "anchor" question).

State scores: The results of state assessment. Typically, state scores are dichotomous: "1" for the state of interest is present, and "0" for the state is absent.

State difficulty: The level of a trait (defining a state of interest) where the probability that a state assessment results in establishing that the state of interest is present, is 50%.

A problem with using ROC analysis for identifying meaningful thresholds is that an optimal ROC cutoff point depends on the prevalence of the state. For any given cutoff point, an increase in the state prevalence results in an increase of the cutoff point's sensitivity and a decrease of its specificity, whereas a decrease in the prevalence has the opposite effect [7]. An optimal ROC-based cutoff point with a balanced sensitivity and specificity in one particular situation (with a certain prevalence) will, therefore, not be the optimal cutoff

point with the same sensitivity-specificity balance in another situation. In other words, an optimal ROC cutoff point is context specific [8]. As a meaningful threshold is principally independent of the state prevalence, the optimal ROC cutoff point may not identify the meaningful threshold on a continuous construct [2]. Only if the state prevalence is 50%, the optimal ROC cutoff point will correspond to the meaningful threshold [2]. In other words, whereas the optimal ROC cutoff point will correspond to the meaningful threshold performs excellently in classifying cases and non-cases with minimal misclassification *in specific situations*, it is not suitable to identify the (mean) threshold on a continuous trait, as defined by clinical or patient criteria.

An alternative to ROC analysis is predictive modeling, which involves logistic regression analysis using the state variable as the outcome and the test variable as the predictor variable [9]. The optimal cutoff point is the test score that is equally likely to occur in the state-positive group as in the state-negative group (i.e., the likelihood ratio is 1). Predictive modeling identifies about the same cutoff point as ROC analysis, but with greater precision [9]. However, like the optimal ROC cutoff point, the predictive modeling cutoff point depends on the state prevalence [10]. The prevalence-related bias in the predictive modeling cutoff point depends on the reliability of the state variable, the standard deviation (SD) of the test variable, and the point-biserial correlation between the test variable and the state variable. These parameters can be used to adjust the prevalence-related bias and recover the proper threshold across a wide range of state prevalences [11].

A third method, recently introduced, is based on item response theory (IRT) [2]. This method uses the state prevalence to estimate a meaningful threshold on the latent trait scale and subsequently determines the corresponding test score threshold. However, this method assumes perfect validity and reliability of the state scores, which is arguably questionable. It is currently unknown to which extent the reliability of the state scores affects the threshold estimate.

This paper presents an improved IRT-based method to estimate meaningful thresholds, which is based on the work of Bjorner et al. [12] in estimating meaningful within-individual change thresholds using longitudinal IRT. Like Bjorner et al. [12], the new method uses the IRT difficulty parameter of the state scores, instead of the state prevalence to find the latent trait threshold of interest. We will demonstrate the performance of this method using simulation studies and two real datasets. We will compare the results with the ROC method, the predictive modeling method [9], the adjusted predictive modeling method [11], and the ”old” state prevalence IRT method [2].

METHOD

Item response theory

IRT aims to explain observed item scores by invoking an unobservable variable underlying these item scores [13]. For instance, the responses to the items of a depression scale can be thought of as being driven by an unobservable continuous variable (i.e., a latent trait) called "depression". A popular IRT model is the graded response model (GRM) [14] that defines the probability of scoring in category *c* or above the following way:

$$\ln\left(\frac{P(X_{ij} \geq c|\theta_i)}{P(X_{ij} < c|\theta_i)}\right) = a_j(\theta_i - b_{jc})$$

where, X_{ij} is the response of person *i* to item *j*, θ_i is the score of person *i* on the latent trait. In principle, θ can take values from -∞ to + ∞.

$$\ln\left(\frac{P(X_{ij} \geq c|\theta_i)}{P(X_{ij} < c|\theta_i)}\right)$$

is the natural logarithm of the odds of person *i* scoring *c* or higher on item *j*. a_j is the discrimination parameter for item *j*. The discrimination parameter refers to the slope of the option characteristic curves, and is a measure of how well the item (categories) distinguishes respondents high and low on the trait. b_{jc} is the difficulty parameter for category *c* on item *j*. The difficulty parameter represents the trait level where the probability of endorsing response category *c* or higher is 50%. The difficulty parameter also indicates the level of the trait where the item response option is most informative.

For an item with 4 response options (i.e., 0, 1, 2 and 3), Figure 1 shows the item–trait relationship graphically as modeled using the GRM [14]. As a fitted IRT model mathematically describes the relationship between responses to the items of a scale and the θ values of the underlying trait, the model is not only able to estimate the trait level (θ) for a given set of responses to the items of a questionnaire, but it is also able to estimate the expected (i.e., mean) questionnaire score (i.e., the sum or test score) for a given trait level.

A meaningful threshold can be thought of as a threshold located somewhere on the latent trait. Such a threshold can be estimated by including the dichotomous state variable in the IRT model, effectively treating the state variable as an extra item (Figure 2). The model for such a dichotomous item is:

$$\ln\left(\frac{P(X_{is} = 1|\theta_i)}{P(X_{is} = 0|\theta_i)}\right) = a_s(\theta_i - b_s)$$

where

$$\ln\left(\frac{P(X_{is}=1|\theta_i)}{P(X_{is}=0|\theta_i)}\right)$$

is the natural logarithm of the odds that person *i* is assessed to be in the state of interest *s*, a_s is the discrimination parameter of the state variable *s*, b_s is the difficulty parameter of the state variable *s*.

The logic behind this approach is that, like the questionnaire items, the state variable is an indicator of the latent trait. Adding the state variable to the IRT model yields a single option characteristic curve for the dichotomous state variable. Importantly, the model estimates a single difficulty parameter for the state variable, which represents the trait level where the probability of scoring 1 on the state variable is 50%. Interestingly, this point also represents the mean of the individual thresholds for endorsing the state item [12].[1] Once the meaningful threshold is identified in terms of the latent trait level, the fitted IRT model provides the corresponding threshold in terms of the PROM score using the expected test score function.

Simulations

We simulated datasets with known individual meaningful thresholds to demonstrate how the new IRT method performs, relative to the ROC method, the predictive modeling method [9], the adjusted predictive modeling method [11] and the old IRT method based on the state prevalence [2]. The beauty of simulations is that the true meaningful threshold can be specified and simulated, and the results can be judged with respect to the extent to which the truth can be accurately recovered.

We simulated multiple datasets with 1000 subjects. We used GRM IRT to simulate item responses to a hypothetical 10-item questionnaire, each item having 4 response options, based on a prespecified set of item parameters (see Supplementary File 1) and varying distributions of the latent trait (θ) (the simulation syntax is provided in Supplementary File 1, section 2). We varied the mean of the normally distributed latent trait (θ_{sim}) across the values -1.4, -0.7, 0, 0.7 and 1.4 (thus simulating samples of low to high mean severity of the trait), and the standard deviation (SD) of θ_{sim} across 1, 1.5 and 2 (thus simulating more and less heterogeneous samples). Figure 3 shows the distribution of the latent traits (A-panels) and the resulting distribution of the 10-item scale scores (B-panels) for three example

1 At the location of the mean of the individual thresholds (given a normal distribution of those thresholds) 50% of the individual thresholds are smaller than the θ-value at that location. Hence, for a random sample of respondents who are sitting exactly at the location of the mean of the individual thresholds, the probability that they have passed their individual threshold is 50%. Therefore, the probability that they endorse the state variable at that location is 50%. In other words, the difficulty of the state variable represents the meaningful threshold of interest.

datasets. If the mean θ_{sim}, matches the mean simulated *b*-parameter (Figure 3, dataset 1), the scale score was normally distributed. In case of a mismatch between the mean θ_{sim} and the mean *b*-parameter (Figure 3, datasets 2 and 3), the scale score became skewed and might even demonstrate floor or ceiling effects, despite the underlying latent trait (θ_{sim}) being normally distributed. Figure 3 also shows the expected test function curves obtained from a fitted GRM model (C-panels). By default, a GRM model assumes an underlying latent trait (denoted "modeled theta" or θ_{mod}) with a mean of zero and an SD of 1. Therefore, θ_{mod} is a linear transformation of θ_{sim} and a threshold on the simulated theta scale (θ^T_{sim}) corresponds to a threshold on the modeled theta scale (θ^T_{mod}) according to the following equation:

$$\theta^T_{mod} = (\theta^T_{sim} - \text{mean}(\theta_{sim})) / \text{SD}(\theta_{sim})$$

Note, however, that the expected test score corresponding to $\theta^T_{sim} = 0$ was independent of the distribution of θ^T_{sim}. For illustration, consider dataset 3 in Figure 3. After IRT modeling and fitting the dataset, the threshold θ^T_{mod}, following the equation above, was (0 – 1.4)/2 = –0.7. Panel C shows the expected test score function of the fitted model (i.e., the relationship between θ^T_{mod} and the test score). Based on the expected test score function, the threshold θ^T_{mod} corresponded to an expected test score of 15.1.

The state scores were simulated as follows. We assumed that the state assessment was based on the comparison of a "perceived trait" with the relevant threshold. Professionals making a depression diagnosis compare the perceived level of depression with the professionally defined threshold of clinical depression. Patients answering a PASS anchor question about pain compare their perceived level of pain with their personal thresholds of acceptability. The perceived trait was assumed to consist of the true trait (i.e., the latent trait θ_{sim}) and some "measurement error" (Figure 4) [15]. The measurement error was simulated to have a normal distribution with a variance chosen in such a way as to obtain reliability values of the perceived trait of 0.5, 0.7, or 0.9. The meaningful threshold (θ^T_{sim}) was arbitrarily set to be zero for all datasets. We did not simulate variability of the thresholds across subjects, as this would only add (a little) extra error to the perceived trait. The "observed" dichotomous state scores were then obtained by comparing the continuous perceived trait with the threshold (θ^T_{sim}). Thus, the state scores were a discretization of the underlying perceived trait variable. The observed state prevalence was the proportion of subjects who's perceived trait exceeded the threshold.

The exact true (i.e., as simulated) meaningful threshold in terms of the expected scale score, corresponding to $\theta^T_{sim} = 0$, based on the simulated item parameters (see Supplementary File 1, section 1) was 15.139 (see Supplementary File 1, section 3 for details of the calculation).

Real dataset: diagnostic thresholds

The first real dataset consists of data from a trial involving primary care patients with emotional distress or minor mental disorders [16]. At baseline, 307 patients completed the Hospital Anxiety Depression Scale (HADS), a self-report questionnaire measuring anxiety and depression [17]. In addition, standardized psychiatric diagnoses were obtained by trained interviewers using the Composite International Diagnostic Interview (CIDI) [18]. The original study was approved by the ethical committee of The Netherlands Institute of Mental Health and Addiction and all patients provided written informed consent. We used the HADS depression scale and the CIDI mild, moderate, and severe major depressive disorder (MDD) diagnoses (criteria according to the Diagnostic and Statistical Manual, Fourth Edition; DSM-IV [19]). The HADS depression scale consists of 7 items with 4 response options. Hence, the HADS depression total score ranges from 0 to 21 (0 = no depression, 21 = severe depression). We aimed to establish the clinical thresholds for mild, moderate, and severe MDD. To that end, we constructed 3 dichotomous state variables to be used in separate analyses in conjunction with the HADS items. The first state variable was used to establish the threshold for mild MDD, contrasting mild, moderate, and severe MDD (coded "1") to no MDD (coded "0"). The second state variable was used to establish the threshold for moderate MDD, contrasting moderate and severe MDD (coded "1") to no and mild MDD (coded "0"). The third state variable was used to establish the threshold for severe MDD, contrasting severe MDD (coded "1") to no, mild, and moderate MDD (coded "0").[2]

Real dataset: patient acceptable symptom state (PASS)

The second real dataset was obtained from the Hand-Wrist Study Group cohort and comprised 3522 patients who underwent surgical trigger digit release [20; 21]. All patients were invited to complete the Michigan Hand outcomes Questionnaire (MHQ), a PROM covering six subdomains of hand function [22], three months postoperatively. The study was approved by the local medical ethical review board, and all patients provided written informed consent. We used the MHQ pain subscale, which has a score ranging from 0 to 100 (0 = worst possible pain, 100 = no pain). This score is derived from 5 items, each having five response options. To determine the PASS of the MHQ pain score, we asked patients to answer the following anchor question [23]: "How satisfied are you with your treatment results thus far?" with response options: "excellent", "good", "fair", "moderate", or "poor". Considering that the PASS represents the threshold above which a patient is satisfied with

2 The state variable with 4 diagnostic categories could also be processed as a single-state variable with 4 categories in a single analysis in conjunction with the HADS items. In that case, the state variable provides 3 difficulty estimates representing the thresholds for mild, moderate, and severe MDD on the latent trait metric.

his or her current state [3], we adopted the threshold between "fair" and "good" as the PASS and dichotomized the ratings accordingly.

ANALYSIS

Simulated samples

We calculated thresholds using the ROC method (Youden criterion) [4; 5], the predictive modeling method [9], the adjusted predicted modeling method [11], the old state prevalence IRT method [2], and the new state difficulty IRT method. Bias was calculated as the mean residual (true threshold minus estimated threshold), and the mean square residual (MSR) as the mean of the squared residuals.

Real datasets

As unidimensionality is an important prerequisite for IRT, we checked unidimensionality of the datasets through confirmatory factor analysis. The items were treated as categorical. The following scaled fit indices were taken as indicative of unidimensionality: comparative fit index (CFI) > 0.95, Tucker-Lewis index (TLI) > 0.95, root mean square error of approximation (RMSEA) < 0.06, and standardized root mean square residual (SRMR) < 0.08 [24]. As in the simulated samples, we calculated thresholds using the ROC method [4; 5], the predictive modeling method [9], the adjusted predicted modeling method [11], the old state prevalence IRT method [2], and the new state difficulty IRT method. 95% Confidence intervals were obtained through empirical bootstrap (1000 samples) [25].

Software

We used the statistical program R, version 4.0.3 [26], to organize the data, calculate the predictive and adjusted thresholds, and to perform bootstrapping. The pROC package, version 1.17.0.1 [27], was used to perform ROC analyses. The lavaan package, version 0.6-8 [28], was used to perform confirmatory factor analysis. The mirt package, version 1.33.2 [29], was used to simulate datasets, fit GRMs and calculate expected test scores.

RESULTS

Simulations

The simulated datasets varied in means and standard deviations of the test scores (Table 1). Because of the fixed meaningful threshold ($\theta^{T}_{\text{sim}} = 0$), increasing or decreasing the mean θ_{sim} intentionally lead to increase or decrease of the state prevalence (i.e., the proportion of subjects exceeding the threshold). Moreover, as increasing or decreasing the mean

θ_{sim} caused mismatch between the mean θ_{sim} and the mean item difficulty parameter, this inevitably caused variable degrees of skewness (as illustrated in Figure 3). Figure 5 shows the estimated meaningful thresholds as a function of the state prevalence, by method and state scores reliability. The ROC-based thresholds and the predictive modeling-based thresholds clearly varied with the state prevalence and the state scores reliability. The old state prevalence IRT method [2] also varied with the state prevalence and the state scores reliability, although to a lesser degree. The adjusted predictive modeling method performed significantly better, although some bias remained if the state prevalence was smaller than 0.3 or greater than 0.7. In contrast to the other methods, the new state difficulty IRT method perfectly recovered the true meaningful threshold with almost no bias and high precision (Table 2). Across all simulated samples, the ROC method yielded the most prevalence-related bias and the least precision, whereas the new IRT method yielded the least bias and the greatest precision.

Real dataset: diagnostic thresholds

The sample characteristics are shown in Table 3. The prevalence of any MDD (i.e., mild, moderate, and severe MDD) was 49%. The mean HADS depression score was 10.7. The fit indices showed some violation of the unidimensionality assumption; however, none of the (absolute) residual correlations exceeded 0.2. The reliability of the diagnostic variable, expressed as the variance of the diagnosis explained by the latent depression trait as measured by the HADS [30], was 0.34. The estimated thresholds for mild, moderate, and severe MDD, using different methods, are shown in Table 4. As the prevalence of any MDD was close to 50%, the threshold for mild MDD should be close to the mean score in the sample [10]. This was confirmed for most methods; only the estimated ROC threshold was lower than the mean sample score. For the other thresholds, with state prevalences < 50%, the methods diverged as expected. The new IRT method identified 10.6, 15.4, and 18.2 as the thresholds for mild, moderate, and severe MDD. The adjusted predictive modeling method identified practically the same thresholds for mild and moderate MDD, but, compared to the new IRT method, the adjusted method slightly underestimated the threshold for severe MDD while its precision was slightly less than the new IRT method.

Real dataset: patient acceptable symptom state (PASS)

Complete data at three months postoperatively were available for 2634 patients. The sample characteristics are depicted in Table 5. Sixty-three percent of patients were satisfied with the treatment result. The mean MHQ pain score was 71 with an SD of 23. The distribution of the pain scores was skewed to the left (skewness -0.45, ceiling effect 0.17). Confirmatory factor analysis indicated an RMSEA of 0.109, while the other fit indices and the residual correlations indicated unidimensionality. Therefore, we assumed essential unidimensionality of the scale. The estimated thresholds for the PASS, using different

methods, are shown in Table 6. As expected, the state prevalence greater than 50% resulted in divergent PASS thresholds for the different methods. The new IRT method identified a PASS threshold for MHQ pain of 59.6 (95% CI 57.3; 61.7). Despite the non-normality of the MHQ pain scores, the threshold identified by the adjusted predictive modeling approach was not significantly different and of similar precision. Based on these results, it is safe to assume that the PASS threshold for the MHQ pain score (as anchored by good/excellent satisfaction with treatment results) three months after trigger finger release is around 60 (58-62). All other methods overestimated the PASS threshold due to prevalence-related bias.

DISCUSSION

As the use of PROMs has become standard practice in clinical research and daily clinical practice, there is an increased incentive to develop meaningful thresholds to accurately interpret questionnaire scores and facilitate clinical decision making. In this article, we have introduced a new IRT approach to estimate meaningful thresholds. The method perfectly recovered the true (as simulated) meaningful threshold as a fixed value on the latent trait with practically no bias and high precision, regardless of the state prevalence or the state scores reliability. In contrast, most of the other methods examined produced biased threshold estimates if the state prevalence was $\neq 0.5$.

Importantly, meaningful thresholds or cutoff points are used for two goals that are principally incompatable with each other: interpretation and classification. The first goal, the interpretation of test scores, relates to questions such as the cutoff point for clinical depression on a depression scale, or the minimum level of acceptability on a pain scale. Interpretational thresholds, especially if they are based on relatively subjective criteria, may depend on specific sample characteristics. For instance, more severe patients may be willing to accept higher levels of knee pain and dysfunction as acceptable than less severe patients [31]. If the thresholds vary, they do so on the patient level, affecting the mean threshold in the sample. The thresholds do not vary with the prevalence of the state of interest. Our new state difficulty IRT method identifies these interpretational thresholds.

The second goal is classification of individual patients. For instance, for screening we often want thresholds that ensure the best balance between sensitivity and specificity, in order to minimize misclassification. To that end, classificational thresholds must be prevalence specific, because a cutoff point's sensitivity and specificity change with prevalence [7]. ROC analysis identifies a test's optimal cutoff point in a particular situation, which cannot be generalized to situations with differing prevalence and disease spectrum. Therefore, the ROC cutoff point does not identify the interpretational threshold on the latent trait (unless the prevalence is 0.5) [2].

Apart from the new state difficulty IRT method, the adjusted predictive modeling method also accurately identified the interpretational threshold with high precision, although some bias occurred with state prevalences smaller than 0.3 or greater than 0.7. This bias is at least partly due to the low or high state prevalence [11], but skewness of the test scores might also play a role. However, the observation of highly similar threshold estimates obtained by the adjusted predictive modeling method and the new IRT method, despite profound skewness and ceiling effects in our second dataset, is a promising finding. Nevertheless, future (simulation) studies should determine to what extent non-normality of the test scores affects the results of the adjusted predictive modeling approach.

The new state difficulty IRT method assumes that the state of interest can be regarded as an effect indicator [32] of the latent trait and, therefore, can be included as an additional item in the IRT model. In some cases, states may alternatively be conceptualized as having a causal effect on the latent trait. Use of such causal indicators [32] is beyond the current paper but can be handled by fitting explanatory IRT models [33].

Both the new state difficulty IRT method and the adjusted predictive modeling method can be used to estimate meaningful thresholds, but the methods come with different assumptions. For the new IRT method, the data should be unidimensional enough to allow IRT analysis [34], and the questionnaire should fit an IRT model. Although any IRT model may be employed, the GRM usually provides good fit to PROM data. Furthermore, the IRT method assumes that the latent trait is normally distributed. Skewness of the observed test scores is no problem as long as the latent trait is assumably normal. On the other hand, the adjusted predictive modeling method assumes normality of the test scores [11].

Taking these assumptions into account, the choice of method may depend on the questionnaire's dimensionality, the distribution of the test scores, and the fit of an IRT model. In case of normally distributed test scores, both the adjusted predictive modeling method and the new IRT method may be used. If the data show profound ceiling or floor effects we recommend using the new state difficulty IRT method. The old state prevalence IRT method [2] is clearly inferior to the new IRT method because the state prevalence is affected by the (un)reliability of the state scores. Therefore, we recommend not to use the old state prevalence IRT method [2] anymore. Similarly, ROC analysis should no longer be used to identify interpretational thresholds.

CONCLUSION

We have introduced a new IRT approach to identify meaningful thresholds for multi-item questionnaires through identifying the latent trait level of the threshold of interest and linking this to the corresponding meaningful threshold on the questionnaire scale. The new IRT method is superior to the adjusted predictive modeling method, especially

if the prevalence is <0.3 or >0.7. Therefore, we recommend to use the new IRT method to estimate meaningful (interpretational) thresholds whenever possible. The adjusted predictive modeling method is a feasible alternative in certain circumstances, for example when the PROM score is not unidimensional enough to allow IRT analysis. We provide the R-code for the new IRT method in Supplementary File, section 4.

REFERENCES

1. Angst, J., & Dobler-Mikola, A. (1984). The Zurich study. II. The continuum from normal to pathological depressive mood swings. Eur Arch Psychiatry Neurol Sci, 234(1), 21-29.
2. Terluin, B., Griffiths, P., van der Wouden, J. C., Ingelsrud, L. H., & Terwee, C. B. (2020). Unlike ROC analysis, a new IRT method identified clinical thresholds unbiased by disease prevalence. J Clin Epidemiol, 124, 118-125.
3. Cepeda, N. A., Polascik, B. A., & Ling, D. I. (2020). A primer on clinically important outcome values: Going beyond relying on p values alone. J Bone Joint Surg Am, 102(3), 262-268.
4. Greiner, M., Pfeiffer, D., & Smith, R. D. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med, 45(1-2), 23-41.
5. Youden, W. J. (1950). Index for rating diagnostic tests. Cancer, 3, 32-35.
6. Gallop, R. J., Crits-Christoph, P., Muenz, L. R., & Tu, X. M. (2003). Determination and interpretation of the optimal operating point for ROC curves derived through generalized linear models. Understanding Statistics, 2(4), 219-242.
7. Brenner, H., & Gefeller, O. (1997). Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Stat Med, 16(9), 981-991.
8. Knottnerus, J. A. (1985). Interpretation of diagnostic data: an unexplored field in general practice. J R Coll Gen Pract, 35(275), 270-274.
9. Terluin, B., Eekhout, I., Terwee, C. B., & de Vet, H. C. W. (2015). Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. J Clin Epidemiol, 68, 1388-1396.
10. Terluin, B., Eekhout, I., & Terwee, C. B. (2017). The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. J Clin Epidemiol, 83, 90-100.
11. Terluin, B., Eekhout, I., & Terwee, C. B. (2022). Improved adjusted minimal important change took reliability of transition ratings into account. J Clin Epidemiol, 148, 48-53.
12. Bjorner, J. B., Terluin, B., Trigg, A., Hu, J., Brady, K. J. S., & Griffiths, P. (2022). Establishing thresholds for meaningful change using longitudinal item response theory. Qual Life Res. https://doi.org/10.1007/s11136-022-03172-5
13. Embretson, S. E., & Reise, S. P. (2009). Item response theory for psychologists. Edition 2. Mahwah, NJ: Lawrence Erlbaum.
14. Samejima, F. (1996). The graded response model. In W. J. van der Linden & R. Hambleton (Eds.), Handbook of modern item response theory (pp. 85-100). New York, NY: Springer.
15. Vanier, A., Sebille, V., Blanchin, M., & Hardouin, J. B. (2021). The minimal perceived change: a formal model of the responder definition according to the patient's meaning of change for patient-reported outcome data analysis and interpretation. BMC Med Res Methodol, 21(1), 128.
16. Brouwers, E. P. M., Tiemens, B. G., Terluin, B., & Verhaak, P. F. M. (2006). Effectiveness of an intervention to reduce sickness absence in patients with emotional distress or minor mental disorders: a randomized controlled effectiveness trial. General Hospital Psychiatry, 28, 223-229.
17. Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand, 67, 361-370.

18. World Health Organization. (1990). Composite International Diagnostic Interview. Geneva World Health Organization.

19. American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders. Fourth edition. DSM-IV. Washington DC: American Psychiatric Association.

20. Selles, R. W., Wouters, R. M., Poelstra, R., van der Oest, M. J. W., Porsius, J. T., Hovius, S. E. R., Moojen, T. M., van Kooij, Y., Pennehouat, P. Y., van Huis, R., Vermeulen, G. M., Feitz, R., Slijper, H. P., & Hand-Wrist Study, G. (2020). Routine health outcome measurement: development, design, and implementation of the Hand and Wrist Cohort. Plast Reconstr Surg, 146(2), 343-354.

21. Koopman, J. E., Hundepool, C. A., Wouters, R. M., Duraku, L. S., Smit, J. M., Selles, R. W., Hand-Wrist study, g., & Zuidam, J. M. (2022). Factors associated with self-reported pain and hand function following surgical A1 pulley release. J Hand Surg Eur Vol, 17531934221085401.

22. Chung, K. C., Pillsbury, M. S., Walters, M. R., & Hayward, R. A. (1998). Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg Am, 23(4), 575-587.

23. De Ridder, W. A., van Kooij, Y. E., Vermeulen, G. M., Slijper, H. P., Selles, R. W., Wouters, R. M., & and the Hand-Wrist Study, G. (2021). Test-retest reliability and construct validity of the satisfaction with treatment result questionnaire in patients with hand and wrist conditions: a prospective study. Clin Orthop Relat Res, 479(9), 2022-2032.

24. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1-55.

25. Berrar, D. (2018). Introduction to the non-parametric bootstrap. In M. Cannataro (Ed.), Methods (Vol. Volume I, pp. 766-773). New York: Elsevier.

26. R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

27. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, 77.

28. Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. Journal of Statistical Software, 48, 2.

29. Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. Journal of Statistical Software, 48(6), 1-29.

30. Griffiths, P., Terluin, B., Trigg, A., Schuller, W., & Bjorner, J. B. (2022). A confirmatory factor analysis approach was found to accurately estimate the reliability of transition ratings. J Clin Epidemiol, 141, 36-45.

31. Ingelsrud, L. H., Terluin, B., Gromov, K., Price, A., Beard, D., & Troelsen, A. (2021). Which Oxford Knee Score level represents a satisfactory symptom state after undergoing a total knee replacement? Acta Orthop, 92(1), 85-90.

32. **32.** Bollen, K. A. (1984). Multiple indicators - internal consistency or no necessary relationship. *Quality & Quantity, 18*(4), 377–385.

33. **33.** De Boeck, P., & Wilson, M. (Eds.). (2004). Explanatory item response models: A generalized linear and nonlinear approach. Springer.

34. Cook, K. F., Kallen, M. A., & Amtmann, D. (2009). Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. Quality of Life Research, 18(4), 447-460.

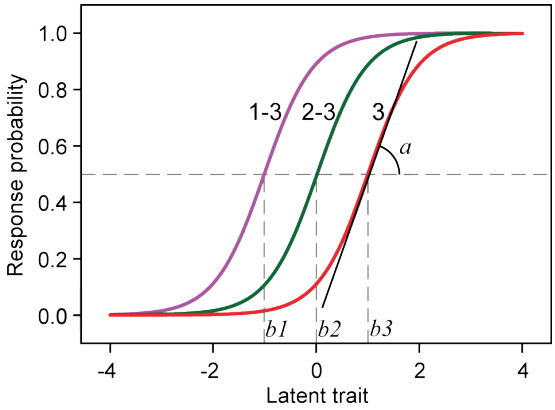


Figure 1. Option characteristic curves of an item with 4 ordered response options (0, 1, 2 and 3) based on the graded response model. Three curves are displayed showing, from left to right, the probability of endorsing options 1, 2 or 3 instead of option 0 (labeled "1-3"), the probability of endorsing options 2 or 3 instead of options 0 or 1 (labeled "2-3"), and the probability of endorsing option 3 instead of options 0, 1 or 2 (labeled "3"), respectively, as a function of the latent trait. The difficulty parameters (labeled "*b1*", "*b2*" and "*b3*") are indicated by vertical dashed lines. The discrimination parameter (labeled "*a*") reflects the slope of the option characteristic curves

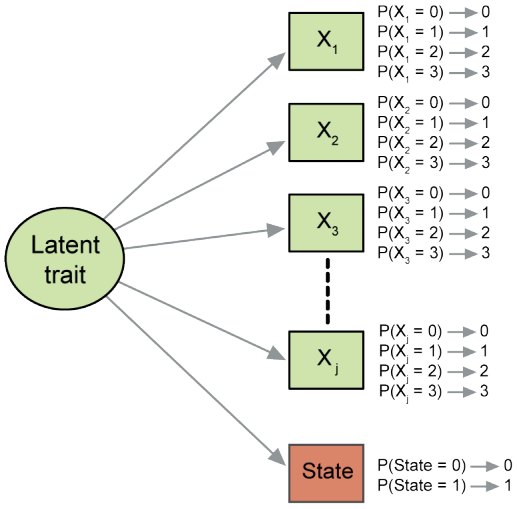


Figure 2. IRT model to estimate a meaningful threshold on a questionnaire with j items. Rectangles represent observed variables: questionnaire items 1 through j, ($X_1 - X_j$), and the state scores (State). The oval represents the latent trait underlying the item scores (and the state scores). The latent trait determines the probabilities of scoring the item response options 0–3 (e.g., $P(X_1 = 0)$, etc.) and the state scores item, according to the item parameters difficulty and discrimination (not shown)

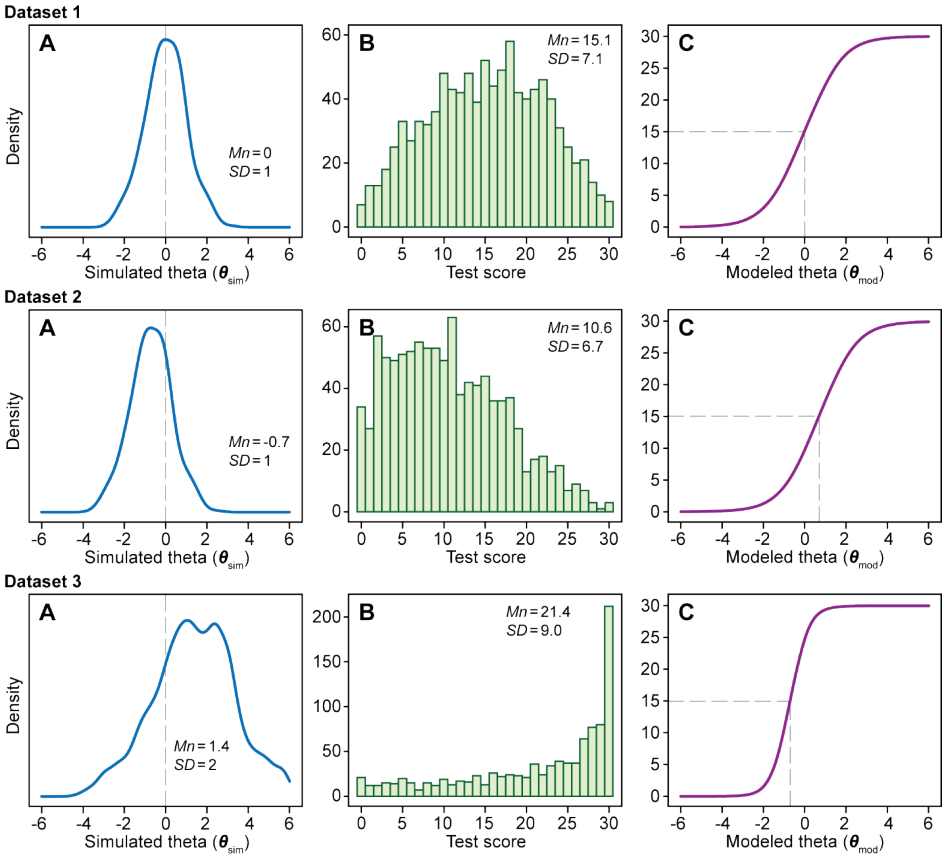


Figure 3. Examples of 3 simulated datasets. A-panels: Density curves showing the simulated theta distributions (*Mn* = mean; *SD* = standard deviation). B-panels: Histograms showing the distribution of the corresponding test scores (i.e., scale scores; *Mn* = mean; *SD* = standard deviation). C-panels: Expected test function curves showing the expected scale score as a function of the modeled theta. Meaningful thresholds defined by $\theta_{sim} = 0$ are indicated by vertical dashed lines (A-panels). The expected test scores corresponding to the meaningful thresholds are indicated by horizontal dashed lines (C-panels)

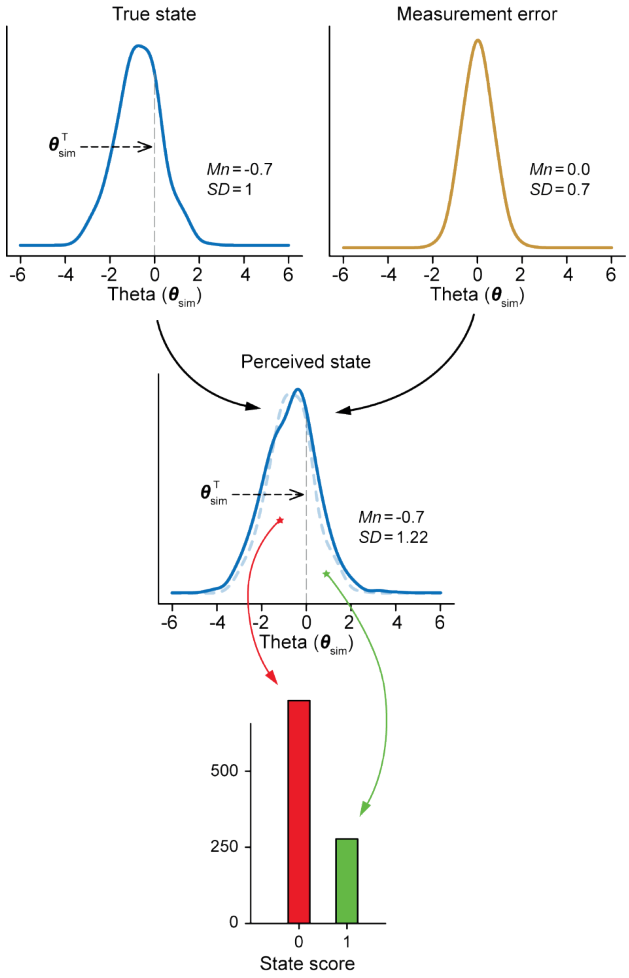


Figure 4. Graphical representation of how the state scores were simulated. The perceived trait is the true trait plus measurement error (all in the theta metric). In the perceived trait graph, the true trait is indicated by a dashed curve. The state scores ("1": state of interest is present; "0": state of interest is absent) are a discretization of the perceived trait relative to the meaningful threshold (θ^{T}_{sim})

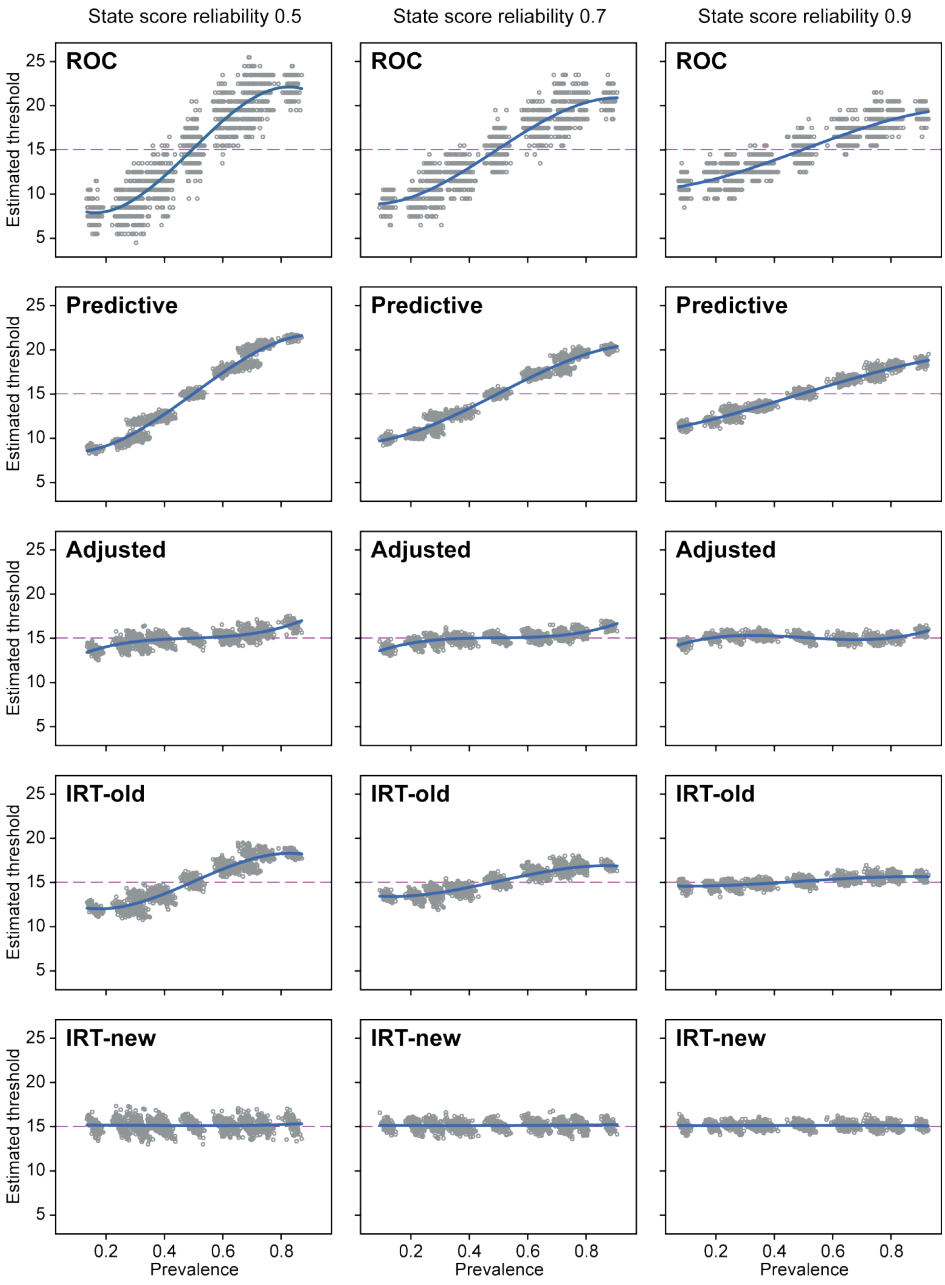


Figure 5. Estimated meaningful thresholds across 4500 simulated datasets by state prevalence, state scores reliability and method (row 1: ROC, row 2: predictive modeling, row 3: adjusted predictive modeling, row 4: old IRT method using state prevalence, row 5: new IRT method using state difficulty parameter. The true threshold was 15.139 in all datasets, indicated by horizontal dashed lines

Table 1. Sample characteristics of the 4500 simulated datasets (mean, range)

Sample characteristic	Mean	Range
Mean test score	15.0	6.2; 23.9
SD test score	8.0	5.2; 10.5
Skewness test score	-0.01	-1.19; 1.21
Kurtosis test score	-0.54	-1.48; 1.25
Floor effects	0.06	0.00; 0.25
Ceiling effects	0.06	0.00; 0.26
State prevalence[a]	0.50	0.07; 0.93

[a] State prevalence based on the proportion of persons passing the threshold on the perceived trait

Table 2. Bias and mean square residual (MSR) by method, state scores reliability, and state prevalence

Method	Prevalence < 0.3		0.3 ≤ Prevalence < 0.5		0.5 ≤ Prevalence < 0.7		Prevalence ≥ 0.7	
Reliability 0.5	**Bias**	**MSR**	**Bias**	**MSR**	**Bias**	**MSR**	**Bias**	**MSR**
ROC	-6.69	47.65	-3.02	16.63	2.94	16.30	6.59	45.84
Predictive modeling	-5.60	32.23	-2.41	8.57	2.15	7.54	5.47	30.76
Adjusted predictive modeling	-0.96	1.42	-0.23	0.34	0.04	0.24	0.95	1.37
Old IRT (state prevalence)	-2.89	8.86	-1.35	3.26	1.28	3.10	2.95	9.09
New IRT (state difficulty)	0.10	0.54	0.02	0.40	-0.06	0.34	0.05	0.49
Reliability 0.7	**Bias**	**MSR**	**Bias**	**MSR**	**Bias**	**MSR**	**Bias**	**MSR**
ROC	-5.10	28.43	-1.84	7.31	1.82	7.26	4.77	25.45
Predictive modeling	-4.20	18.65	-1.45	3.60	1.31	3.13	4.02	17.13
Adjusted predictive modeling	-0.65	0.74	-0.08	0.17	-0.04	0.14	0.56	0.67
Old IRT (state prevalence)	-1.52	2.64	-0.54	0.82	0.56	0.79	1.53	2.63
New IRT (state difficulty)	0.02	0.23	0.04	0.23	-0.01	0.19	0.03	0.25
Reliability 0.9	**Bias**	**MSR**	**Bias**	**MSR**	**Bias**	**MSR**	**Bias**	**MSR**
ROC	-3.39	13.09	-1.10	2.91	0.92	2.59	3.26	12.21
Predictive modeling	-2.88	8.87	-0.92	1.47	0.79	1.24	2.77	8.22
Adjusted predictive modeling	-0.12	0.25	0.13	0.10	-0.21	0.13	0.03	0.21
Old IRT (state prevalence)	-0.49	0.37	-0.13	0.16	0.15	0.16	0.50	0.38
New IRT (state difficulty)	-0.01	0.13	0.03	0.10	-0.03	0.10	0.00	0.13

Table 3. Sample and scale characteristics of the HADS[a] dataset (N = 295)

Characteristics	Values
Gender (proportion females)	0.60
Age, mean (SD)	39.5 (9.2)
Prevalence mild MDD[b]	0.23
Prevalence moderate MDD[b]	0.12
Prevalence severe MDD[b]	0.14
HADS[a] depression score, mean (SD)	10.7 (4.3)
Scaled comparative fit index (CFI)	0.981
Scaled Tucker-Lewis index (TLI)	0.971
Scaled root mean square error of approximation (RMSEA)	0.089
Standardized root mean square residual (SRMR)	0.044
State reliability of the diagnostic variable	0.34

[a] HADS = Hospital Anxiety Depression Scale
[b] MDD = major depressive disorder (DSM-IV)

Table 4. Thresholds for mild, moderate, and severe MDD[a] for the HADS[b] depression scale

Method	Mild MDD[a]		Moderate MDD[a]		Severe MDD[a]	
	Estimate	**95% CI**	**Estimate**	**95% CI**	**Estimate**	**95% CI**
ROC	9.5	9.5; 10.5	10.5	9.5; 13.5	11.5	10.5; 13.5
Predictive modeling	10.8	10.3; 11.2	11.6	11.1; 12.2	12.2	11.7; 12.8
Adjusted predictive modeling	10.8	10.0; 11.7	15.3	14.1; 16.9	17.6	15.8; 20.2
Old state prevalence IRT method	10.7	10.1; 11.3	13.3	12.6; 14.1	15.2	14.3; 16.1
New state difficulty IRT method	10.6	9.7; 11.6	15.4	14.2; 16.8	18.2	16.8; 19.5

[a] MDD = major depressive disorder (DSM-IV)
[b] HADS = Hospital Anxiety Depression Scale

Table 5. Sample and scale parameters of the Hand Wrist Study Group dataset (N=2634)

Characteristics	Values
Gender (proportion females)	0.67
Age, mean (SD)	59 (10)
Satisfaction with treatment results (proportions)	
Poor	0.03
Moderate	0.11
Fair	0.23
Good	0.40
Excellent	0.23
MHQ[a] pain score, mean (SD)	71 (23)
Scaled comparative fit index (CFI)	0.993
Scaled Tucker-Lewis index (TLI)	0.989
Scaled root mean square error of approximation (RMSEA)	0.109
Standardized root mean square residual (SRMR)	0.029
State reliability of the anchor question	0.40

[a] MHQ = Michigan Hand outcomes Questionnaire

Table 6. PASS thresholds for the MHQ[a] pain scale

	Estimate	95% CI
ROC	77.5	72.5; 77.5
Predictive modeling	69.0	68.2; 69.9
Adjusted predictive modeling	60.1	58.4; 61.8
Old state prevalence IRT method	64.3	63.0; 65.8
New state difficulty IRT method	59.6	57.3; 61.7

[a] MHQ = Michigan Hand outcomes Questionnaire

Part II

CLINICALLY IMPORTANT OUTCOMES OF TREATMENT (STRATEGIES) FOR THUMB BASE OSTEOARTHRITIS

5

LONG-TERM OUTCOMES OF NONSURGICAL TREATMENT OF THUMB CARPOMETACARPAL OSTEOARTHRITIS: A COHORT STUDY

Esteban Lopez, L. M. J., Hoogendam, L., Vermeulen, G. M., Tsehaie, J., Slijper, H. P., Selles, R. W., Wouters, R. M., The Hand-Wrist Study Group (2023). *The Journal of bone and joint surgery. American volume*, *105*(23), 1837–1845.

ABSTRACT

Background

Although non-surgical treatment for thumb carpometacarpal (CMC-1) osteoarthritis (OA) provides a short-term improvement, the durability of these effects beyond one year is unknown. This study investigates the patient-reported pain and limitations in ADL at >5 years following non-surgical treatment (i.e., exercise therapy and orthotics) for CMC-1 OA, hypothesizing that pain and limitations in ADL will not worsen after 12 months. Secondary outcomes are satisfaction with treatment results, health-related quality of life, and conversion to surgery at >5 years follow-up.

Methods

This was a multicenter, prospective cohort study using two overlapping samples. The change in Michigan Hand outcomes Questionnaire (MHQ) subscales pain and ADL between 12 months and >5 years were the primary outcomes, measured in the first sample (n=170), which included non-operated patients. Additional time points included baseline and three months. We evaluated conversion to surgery in a second sample, using all patients that responded to the invitation for this follow-up study (n=217).

Results

At a median follow-up of 6.6 years (range 5.1-8.7 years), the MHQ pain showed no significant difference, and the MHQ ADL improved, though not clinically relevant, by 4.4 points (95% CI:1.5-7.2) compared to 12 months. At >5 years, 5% of the patients rated their satisfaction as 'poor', 14% as 'moderate', 26% as 'fair', 39% as 'good', and 16% as 'excellent'. The median EQ-5D-5L index score was 0.852 (range 0.135-1). The conversion to surgery rate was 22% (95% CI:16.4-27.7) at a median follow-up of 7 years (range 5.5-9.0 years).

Conclusions

We found positive outcomes at >5 years follow-up for non-surgical treatment of CMC-1 OA, with no worsening of pain or limitations in ADL after 12 months**.** Our findings support non-surgical treatment as the first treatment choice and suggest that treatment effects are sustainable.

INTRODUCTION

Osteoarthritis (OA) of the thumb carpometacarpal joint (CMC-1) is a common disorder with a symptomatic prevalence of 7% for females and 2% for males aged >50(1–3). Because of the aging population, the number of patients with CMC-1 OA is expected to increase(3,4). CMC-1 OA often results in pain, limitations in activities of daily life (ADL), reduced quality of life, thenar muscle wasting, and/or thumb deformity(1,2).

Most guidelines for CMC-1 OA advise starting with non-surgical treatment such as orthoses, steroid injections, analgesics, exercise therapy, or a combination(5–7). Surgery can be considered when non-surgical treatment fails to sufficiently relieve symptoms(7–9).

Although short-term effectiveness has been demonstrated for the combination of exercise therapy and orthoses(7,10–12), the long-term effects are unknown(7,11,13,14). Exercise therapy can optimize thumb positioning and improve use in daily life, with a known treatment effect up to one year(11). Due to the chronic character of CMC-1 OA, knowledge of long-term outcomes is required to provide patients and clinicians with insights into the long-term perspectives. As the treatment focuses on stable thumb positioning and improved use in daily life, we hypothesized that the treatment effects remain stable after one year up to five years and that this treatment thus offers a sustainable solution. Furthermore, while guidelines advise initial non-surgical treatment to potentially avoid surgery, the actual long-term conversion to surgery has been poorly described. Few studies report conversion to surgery, usually with a follow-up of approximately two years(11,15,16). Avoiding surgery is important due to its lengthy rehabilitation, costs, and outcome variation: the results are not always to the patients' satisfaction(17).

This study investigates the pain and limitations in ADL following non-surgical treatment (consisting of exercise therapy and orthotics) for CMC-1 OA at >5 years and tests the hypothesis that these outcomes will not deteriorate after 12 months in patients who did not convert to surgery. Secondary and exploratory outcomes include satisfaction with treatment results and health-related quality of life at >5 years and the conversion to surgery rate following non-surgical treatment for CMC-1 OA.

MATERIALS AND METHODS

Study Design

This is a multicenter, prospective cohort study reported following the STROBE statement(18). All patients gave informed consent, and the institutional review board approved the study.

Setting

Data were collected at Xpert Clinics, comprising 8 specialized hand clinics with 18 hand surgeons in the Netherlands during the study period. Outcomes up to 12 months were routinely measured using GemsTracker**;** details on our data collection have been published earlier(10,11,19). Based on the diagnosis treatment combination, each patient is assigned a measurement track, which includes predefined measurements on predefined time points(20). For this study, patients were invited to complete prospectively distributed questionnaires >5 years after treatment onset.

Participants

All patients were diagnosed with CMC-1 OA by hand surgeons based on symptoms and physical examination and referred for hand therapy between January 2011 and October 2015. X-rays are usually taken but not routinely recorded. We invited patients to participate in this follow-up study if they completed the Visual Analogue Scale (VAS) and Michigan Hand Outcomes Questionnaire (MHQ) at baseline and three months as part of their routine outcome measurements.

We used two overlapping samples in this study, where sample 1 is a subset of sample 2. In sample 1, we studied pain and limitations in ADL of patients who did not have surgery. We used sample 2 to evaluate conversion to surgery. For both samples, we included adult patients with complete sociodemographics at baseline, the MHQ at baseline and three months, and our follow-up questionnaires at >5 years. Exclusion criteria for both samples were CMC-1 surgery before non-surgical treatment, post-traumatic CMC-1 OA, simultaneous intervention for other hand or wrist comorbidity, and a corticosteroid injection <6 weeks before treatment. For sample 1, additional criteria were that they did not convert to surgery. Information on these criteria was extracted from patient records and our data collection system.

Intervention

Trained hand therapists carried out the treatment at 8 specialized hand clinics. Treatment was based on the local and Dutch treatment guidelines, generally consisting of a custom-made or prefabricated orthosis, weekly 25-minute therapy sessions including exercises and education to achieve a more stable opposition position, and daily exercises(5). A detailed description of the treatment is presented in Appendix 1. Treatment frequency and duration differed per patient and were determined in conjunction by the therapist and the patient, inherent to the study's observational nature.

Measurements

Baseline demographics, including sex, age, type of work, dominance, affected hand, and symptom duration, were retrieved from our database.

The primary outcome was the change in pain and limitations in ADL between 12 months and >5 years. Pain and limitations in ADL were routinely measured with the MHQ subscales pain and ADL at baseline, 3 months, and one year. For this study, additional and prospective MHQ data were collected at >5 years. The MHQ has a high internal consistency, validity, and acceptable reliability(21). The MHQ subscales overall hand function, work performance, aesthetics, and satisfaction with hand function were secondary outcomes(21,22). MHQ scores range from 0-100; higher scores indicate better performance in all subscales except pain. We converted the subscale pain to ease interpretation; thus, higher scores indicate better performance on all subscales. The MHQ subscales pain and ADL have a minimal important change (MIC) of 8 and 6, respectively, for non-surgically treated CMC-1 OA(23).

Apart from the MHQ, other questionnaires were distributed at timepoint >5 years. We used the valid and reliable Satisfaction with Treatment Result Questionnaire, which evaluates satisfaction with treatment results on a 5-point Likert scale (poor, moderate, fair, good, and excellent) and asks whether the patient would undergo the treatment again under similar circumstances(24).

The 5-level EuroQol-5D (EQ-5D-5L) was used to measure Health Related Quality of Life (HRQoL) at >5 years. Scores of the EQ-5D-index are anchored at 1 (full health) and 0 (a state as bad as being dead; lower values indicate even worse health states)(25,26). The EQ-5D-5L had good reliability, validity, and moderate responsiveness in patients with upper extremity disorders(27).

Data on conversion to surgery within our clinics was extracted from patient records and our data collection database of all patients of sample 2. Additionally, we used a survey to ask patients at >5 years whether they underwent surgery for their CMC-1 OA.

Study size

Power analysis based on our four repeated measures, a conventional effect size of 0.25(28), alpha=0.025 (Bonferonni correction, correcting for two primary outcomes, i.e., pain and ADL), and a Power of 0.80 indicated 49 participants were required. This was well below the achieved sample of 170.

Statistical analysis

We used linear mixed models with the MHQ subscales pain and ADL as dependent variables and timepoint as a fixed factor on sample 1. We performed the same analyses

for the other MHQ subscales. Assumptions were checked using residual plots and normal probability plots.

We performed a non-responder analysis to investigate whether patients with missing values systematically differed from those without missing values by comparing demographics and primary outcomes at baseline and 3 months of participants that completed the MHQ at >5 years (defined as responders) with the participants that did not (defined as non-responders) using t-tests and chi-squared tests. We performed a similar non-responder analysis for missing data at 12 months. Additionally, we performed Little's test, which assesses the null hypothesis that the data are missing completely at random(29,30).

We report descriptive statistics for HRQoL and satisfaction with treatment results of sample 1 at >5 years. We report the conversion to surgery rate and use a Kaplan Meier survival curve to display the time in months before deciding to convert to surgery for sample 2. All analyses were performed using R statistical computing (version 4.0.3).

SOURCE OF FUNDING

Not applicable.

RESULTS

We invited 552 patients for this study based on their earlier assigned measurement track. We excluded 87 patients with traumatic CMC-1 OA, hand-related comorbidities, or a corticosteroid injection <6 weeks before treatment initiation (Figure 1). After applying the additional exclusion criteria (no surgery) for sample 1, 170 patients were included in the first sample, of which 134 also had MHQ outcomes at 12 months. We included 217 patients in the second sample. The baseline characteristics of both samples are depicted separately in Table 1.

The two non-responder analyses indicated that at >5 years, only one of twenty-one, and at 12 months, two of twenty-one variables differed between responders and non-responders (Supplementary Tables 1-2). The non-significant Little's test (p=0.175) further suggests that the data were missing completely at random(29,30).

Primary and secondary outcomes

The median follow-up time was 6.6 years (range 5.1–8.7 years). Figure 2 demonstrates the MHQ ADL and pain scores over time in the non-surgical patients of sample 1.

The MHQ pain subscale did not change between 12 months and >5 years. The MHQ ADL improved significantly, although not clinically relevant, by 4.4 points (95% CI:1.5-7.2;

p=<0.0137) (Table 2). The MHQ total score and the subscales overall hand function, and workability improved significantly and clinically relevant between 12 months and >5 years, whereas we observed no difference in the other subscales (Table 2).

163 participants of sample 1 completed the Satisfaction with Treatment Results Questionnaire at >5 years. Five percent responded with 'poor', 14% with 'moderate', 26% with 'fair', 39% with 'good', and 16% with 'excellent' (Figure 3). 71% of the participants were willing to undergo the treatment again under similar circumstances. The median EQ-5D-5L index score was 0.852 (range 0.135-1).

Figure 4 shows the survival curve for conversion to surgery, demonstrating that after a median follow-up of 7.0 years (range 5.5–9.0), 47 participants (22%) converted to surgery (95% CI:16.4 – 27.7). Of these patients, the majority (70%) converted to surgery within the first year after treatment. The median time to decide to convert to surgery was 7.4 months (range 0.7-82.7) after the treatment initiation.

DISCUSSION

We found no change in pain and limitations in ADL between 12 months and >5 years following non-surgical treatment for CMC-1 OA. These findings apply to CMC-1 patients that did not convert to surgery during the follow-up interval. Our findings show that the improvement after non-surgical treatment as measured in the first 12 months after treatment is sustainable. Secondary outcomes at >5 years follow-up indicate high satisfaction with treatment results, and only 22% converted to surgery after a median follow-up of 7.0 years.

The long-term effect of non-surgical treatment of CMC-1 OA has not been well-described. Although most improvements occurred in the active treatment period (i.e., the first three months), there were additional improvements at 12 months and even >5 years in patients that did not convert to surgery. As we hypothesized, there was no deterioration of pain or limitations measured with the MHQ. A possible explanation for the lasting improvements could be that apart from only strengthening the thenar muscles, exercise therapy aims at using a new and more stable position of the thumb, thereby reducing joint loading and inflammation. Thus, patients learn how to prevent pain by using this position and to cope with their OA.

The HRQoL is a standalone cross-sectional outcome at >5 years. Since HRQoL is an underexposed part of outcomes in this patient population, this outcome could be valuable for future research(31,32). A study by Lane et al. reports EQ-5D-index scores of 0.85 and 0.8 at 1 year postoperatively in patients that had CMC-1 OA surgery, which is similar to or slightly lower than the outcome of the present study(33).

Our finding that only 22% converted to surgery is comparable to other studies with shorter follow-ups. Tsehaie et al. found a conversion rate of 15% with a mean follow-up of 2.2 years, which is partially the same cohort of patients as the present study(11). Gravas et al. reports that during a randomized controlled trial, 24% of the patients who received occupational therapy underwent surgery after a follow-up of 2 years(15). Another study, by Schloemann et al., reports a conversion to surgery rate of 9% after a median follow-up of 1.5 years(16). These different conversion rates may be due to differences in populations or treatment protocols. Nevertheless, the findings that the conversion rate at ±2 years is comparable with the conversion rate at >5 years indicates that few people are operated on once the positive treatment effect has been achieved, which further supports non-surgical treatment as a first treatment choice.

Study limitations

A limitation of this study is the missing data, which is inherent to the observational nature of this study. However, our non-responder analyses suggested that the data were missing at random; there was a slight age difference, which seems clinically irrelevant, and no differences in MHQ scores. The non-significant Little's test further suggests that the missing data were missing completely at random. Therefore, we are confident that these missing data did not influence our primary findings.

Related to our observational design and associated missing data is that our estimate of conversion to surgery is suboptimal. As the patients from whom we had data on conversion to surgery may not have been representative of the target population, the actual conversion rate to surgery may be different. This means that some patients may have sought care elsewhere without us knowing (i.e., the conversion rate could actually be higher) or that patients without symptoms who had not undergone surgery did not respond to our survey (i.e., the conversion rate could actually be lower). However, all patients that had no surgery according to our patient records indicated that they had no surgery in the survey on additional treatments, suggesting that we did not underestimate the conversion rate. Still, this is a limitation, and future studies should incorporate this in their design.

Furthermore, insights into factors associated with conversion to surgery may enable faster, well-founded decision-making. Previous studies found that pain improvement, prior surgical experiences, surgeons' attitudes toward CMC-1 OA, previous non-pharmacological treatment, and higher motivation for surgery are influential(11,15,16). Predicting conversion to surgery may be improved by also considering the psychological characteristics. A previous study found that patients scheduled for surgical treatment have a worse psychological profile than their non-surgical counterparts(34). Due to the relatively low number of conversion to surgery events**,** we could not assess that in our study.

Another limitation is that our patient-reported outcomes only represent patients who did not convert to surgery. This may have introduced selection bias because patients who eventually received surgery most likely had worse outcomes. Therefore, these outcomes only apply to those patients that do not end up with surgical treatment at >5 years. However, since most surgeries occurred within the first year after treatment onset, the findings between 12 months and >5 years seem unaffected by this risk of bias.

Finally, a limitation of our observational design is that we cannot prove causality. Oppositely, there is a higher ecological validity as our findings reflect daily practice. Similarly, although therapists are trained the same, there may have been variations in the fulfillment of the treatment. This possible variation is important to acknowledge because it has previously been shown that exercises have a relatively large treatment effect compared to using an orthosis alone(10). However, this variation is also a strength as it reflects daily clinical practice. To prove causality, a no-treatment arm would have to be included as a control group since the natural course of CMC-1 OA may be somewhat self-limited regarding pain.

In conclusion, we found positive outcomes at >5 years follow-up following non-surgical treatment of CMC-1 OA with no worsening of pain or limitations in ADL after 12 months, and only 22% converted to surgery at a median follow-up of 7 years. Our findings support non-surgical treatment as the first treatment choice and indicate that treatment effects are sustainable.

REFERENCES

1. Bijlsma JWJ, Berenbaum F, Lafeber FPJG. Osteoarthritis: An update with relevance for clinical practice. Lancet. 2011;377(9783):2115–26.
2. Haugen IK, Englund M, Aliabadi P, Niu J, Clancy M, Kvien TK, et al. Prevalence, incidence and progression of hand osteoarthritis in the general population: The Framingham Osteoarthritis Study. Ann Rheum Dis. 2011;70(9):1581–6.
3. van der Oest MJW, Duraku LS, Andrinopoulou ER, Wouters RM, Bierma-Zeinstra SMA, Selles RW, et al. The prevalence of radiographic thumb base osteoarthritis: a meta-analysis. Osteoarthr Cartil [Internet]. 2021;(xxxx). Available from: https://doi.org/10.1016/j.joca.2021.03.004
4. Centraal Bureau voor Statistiek. No Title. 2020; Available from: https://www.cbs.nl/nl-nl/cijfers/detail/37296ned
5. Nederlandse vereniging voor handchirurgie. Conservatieve en Chirurgische Behandeling van Primaire Artrose van de Duimbasis. 2014;1–91.
6. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, Hip, and Knee. Arthritis Care Res. 2020;72(2):149–62.
7. Bertozzi L, Valdes K, Vanti C, Negrini S, Pillastrini P, Villafañe JH. Investigation of the effect of conservative interventions in thumb carpometacarpal osteoarthritis: Systematic review and meta-analysis. Disabil Rehabil. 2015;37(22):2025–43.
8. Buhler M, Chapple CM, Stebbings S, Sangelaji B, Baxter GD. Effectiveness of splinting for pain and function in people with thumb carpometacarpal osteoarthritis: a systematic review with meta-analysis. Osteoarthr Cartil [Internet]. 2019;27(4):547–59. Available from: https://doi.org/10.1016/j.joca.2018.09.012
9. Kroon FPB, Rubio R, Schoones JW, Kloppenburg M. Intra-Articular Therapies in the Treatment of Hand Osteoarthritis: A Systematic Literature Review. Drugs and Aging. 2016;33(2):119–33.
10. Wouters RM, Tsehaie J, Slijper HP, Hovius SER, Feitz R, Blomme A, et al. Exercise Therapy in Addition to an Orthosis Reduces Pain More Than an Orthosis Alone in Patients With Thumb Base Osteoarthritis: A Propensity Score Matching Study. Arch Phys Med Rehabil. 2019;100(6):1050–60.
11. Tsehaie J, Spekreijse KR, Wouters RM, Slijper HP, Feitz R, Hovius SER, et al. Outcome of a Hand Orthosis and Hand Therapy for Carpometacarpal Osteoarthritis in Daily Practice: A Prospective Cohort Study. J Hand Surg Am [Internet]. 2018;43(11):1000-1009.e1. Available from: https://doi.org/10.1016/j.jhsa.2018.04.014
12. Villafañe JH, Valdes K, Pedersini P, Berjano P. Thumb carpometacarpal osteoarthritis: A musculoskeletal physiotherapy perspective. J Bodyw Mov Ther. 2019;23(4):908–12.
13. O'Brien VH, Giveans MR. Effects of a dynamic stability approach in conservative intervention of the carpometacarpal joint of the thumb: A retrospective study. J Hand Ther [Internet]. 2013;26(1):44–52. Available from: http://dx.doi.org/10.1016/j.jht.2012.10.005
14. Deveza LA, Robbins SR, Duong V, Bennell KL, Vicenzino B, Hodges PW, et al. Efficacy of a Combination of Conservative Therapies vs an Education Comparator on Clinical Outcomes in Thumb Base Osteoarthritis: A Randomized Clinical Trial. JAMA Intern Med. 2021;181(4):429–38.
15. Gravås EMH, Østerås N, Nossum R, Eide REM, Klokkeide Å, Matre KH, et al. Does occupational therapy delay or reduce the proportion of patients that receives thumb carpometacarpal joint surgery? A multicentre randomised controlled trial. RMD Open. 2019;5(2):1–8.

16. Schloemann D, Hammert WC, Liu S, Bernstein DN, Calfee RP. Risk Factors for Failed Nonsurgical Treatment Resulting in Surgery on Thumb Carpometacarpal Arthritis. J Hand Surg Am [Internet]. 2021;46(6):471-477.e1. Available from: https://doi.org/10.1016/j.jhsa.2021.02.009

17. Vermeulen GM, Brink SM, Slijper H, Feitz R, Moojen TM, Hovius SER, et al. Trapeziometacarpal arthrodesis or trapeziectomy with ligament reconstruction in primary trapeziometacarpal osteoarthritis: A randomized controlled trial. J Bone Jt Surg - Am Vol. 2014;96(9):726–33.

18. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. Int J Surg [Internet]. 2014;12(12):1495–9. Available from: http://dx.doi.org/10.1016/j.ijsu.2014.07.013

19. GemsTracker c, Erasmus MC EZ. No Title. 2017;version 1. Available from: https://gemstracker.org/

20. Selles RW, Wouters RM, Poelstra R, Van Der Oest MJW, Porsius JT, Hovius SER, et al. Routine Health Outcome Measurement: Development, Design, and Implementation of the Hand and Wrist Cohort. Plast Reconstr Surg. 2020;146(2):343–54.

21. Poole JL. Measures of hand function: Arthritis Hand Function Test (AHFT), Australian Canadian Osteoarthritis Hand Index (AUSCAN), Cochin Hand Function Scale, Functional Index for Hand Osteoarthritis (FIHOA), Grip Ability Test (GAT), Jebsen Hand Function Test (JHFT). Arthritis Care Res. 2011;63(SUPPL. 11).

22. Kroon FPB, Boersma A, Boonen A, van Beest S, Damman W, van der Heijde D, et al. Performance of the Michigan Hand Outcomes Questionnaire in hand osteoarthritis. Osteoarthr Cartil [Internet]. 2018;26(12):1627–35. Available from: https://doi.org/10.1016/j.joca.2018.07.018

23. Hoogendam L, Koopman JE, van Kooij YE, Feitz R, Hundepool CA, Zhou C, et al. What Are the Minimally Important Changes of Four Commonly Used Patient-Reported Outcome Measures for 36 Hand and Wrist Condition-Treatment Combinations? Clin Orthop Relat Res. 2021;Publish Ah:1–15.

24. De Ridder WA, van Kooij YE, Vermeulen GM, Slijper HP, Selles RW, Wouters RM. Test-retest Reliability and Construct Validity of the Satisfaction with Treatment Result Questionnaire in Patients with Hand and Wrist Conditions: A Prospective Study. Clin Orthop Relat Res. 2021;Publish Ah:1–11.

25. Versteegh M, M. Vermeulen K, M. A. A. Evers S, de Wit GA, Prenger R, A. Stolk E. Dutch Tariff for the Five-Level Version of EQ-5D. Value Heal [Internet]. 2016;19(4):343–52. Available from: http://dx.doi.org/10.1016/j.jval.2016.01.003

26. Buchholz I, Janssen MF, Kohlmann T, Feng YS. A Systematic Review of Studies Comparing the Measurement Properties of the Three-Level and Five-Level Versions of the EQ-5D. Pharmacoeconomics [Internet]. 2018;36(6):645–61. Available from: https://doi.org/10.1007/s40273-018-0642-5

27. Grobet C, Marks M, Tecklenburg L, Audigé L. Application and measurement properties of EQ-5D to measure quality of life in patients with upper extremity orthopaedic disorders: a systematic literature review. Arch Orthop Trauma Surg [Internet]. 2018;138(7):953–61. Available from: http://dx.doi.org/10.1007/s00402-018-2933-x

28. Cohen J. Statistical power analysis for the behavioral sciences [Internet]. 1988. Available from: https://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf

29. de Groot JAH, Janssen KJM, Zwinderman AH, Bossuyt PMM, Reitsma JB, Moons KGM. Correcting for Partial Verification Bias: A Comparison of Methods. Ann Epidemiol. 2011;21(2):139–48.

30. Little RJA. A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc. 1988;83(404):1198–202.

31. Michon M, Maheu E, Berenbaum F. Assessing health-related quality of life in hand osteoarthritis: A literature review. Ann Rheum Dis. 2011;70(6):921–8.

32. Østerås N, Kjeken I, Smedslund G, Moe RH, Slatkowsky-Christensen B, Uhlig T, et al. Exercise for hand osteoarthritis: A cochrane systematic review. J Rheumatol. 2017;44(12):1850–8.

33. Lane JCE, Rodrigues JN, Furniss D, Burn E, Poulter R, Gardiner MD. Basal thumb osteoarthritis surgery improves health state utility irrespective of technique: a study of UK Hand Registry data. J Hand Surg Eur Vol. 2020;45(5):436–42.

34. Wouters RM, Vranceanu AM, Slijper HP, Vermeulen GM, Van Der Oest MJW, Selles RW, et al. Patients with Thumb-base Osteoarthritis Scheduled for Surgery Have More Symptoms, Worse Psychological Profile, and Higher Expectations Than Nonsurgical Counterparts: A Large Cohort Analysis. Clin Orthop Relat Res. 2019;477(12):2735–46.

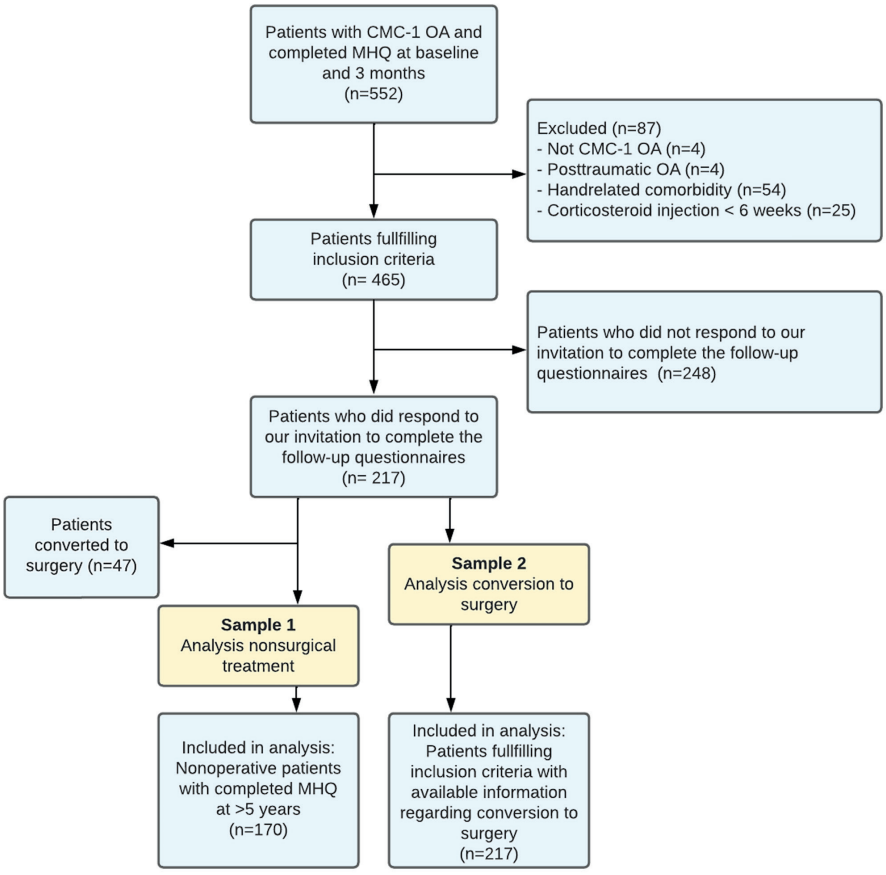


Figure 1. Flowchart of the study. CMC-1 = thumb carpometacarpal, OA = osteoarthritis, MHQ = Michigan Hand Outcomes Questionnaire. Note that sample 1 is a subset of sample 2.

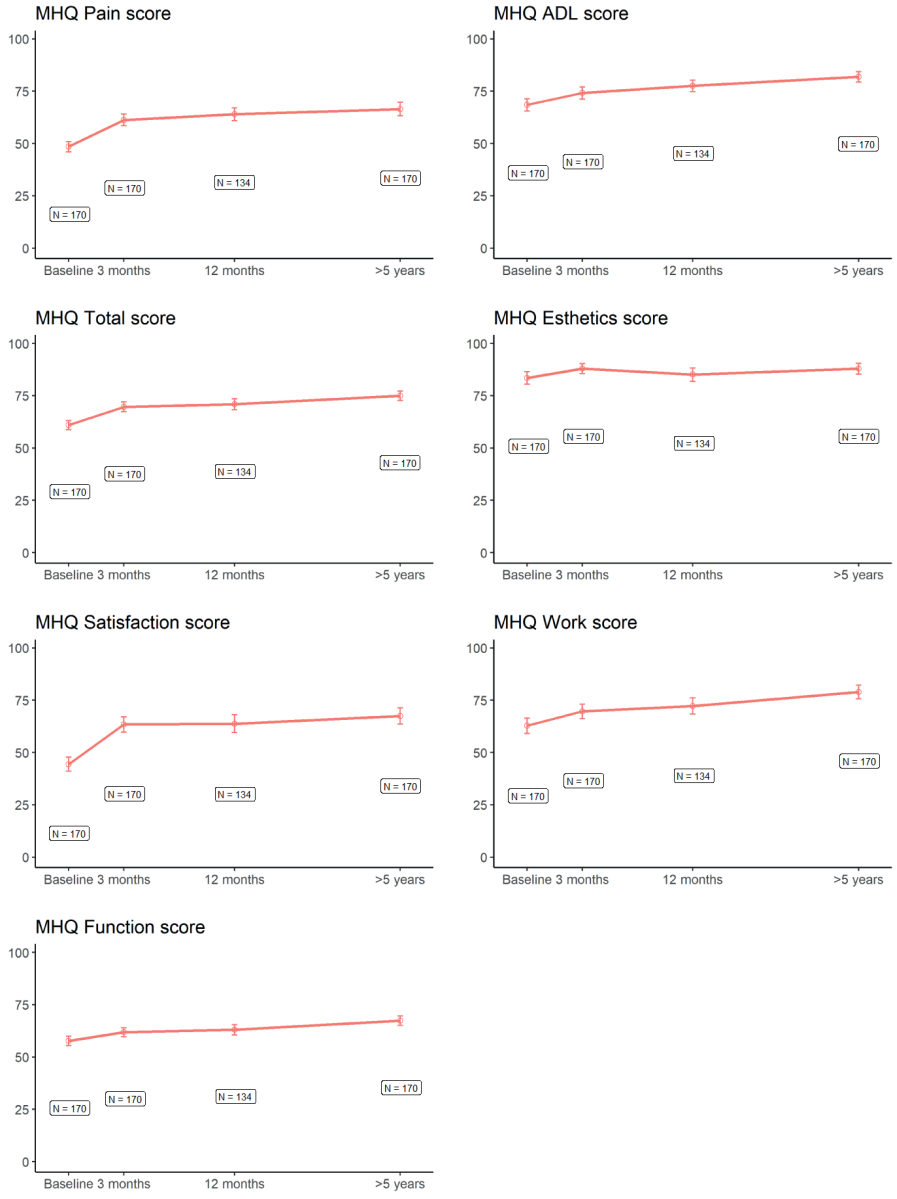


Figure 2. The course of the mean score of the Michigan Hand Outcomes Questionnaire (MHQ) subscales. Most improvement is seen in the first three months. The error bars show the 95% Confidence Interval.

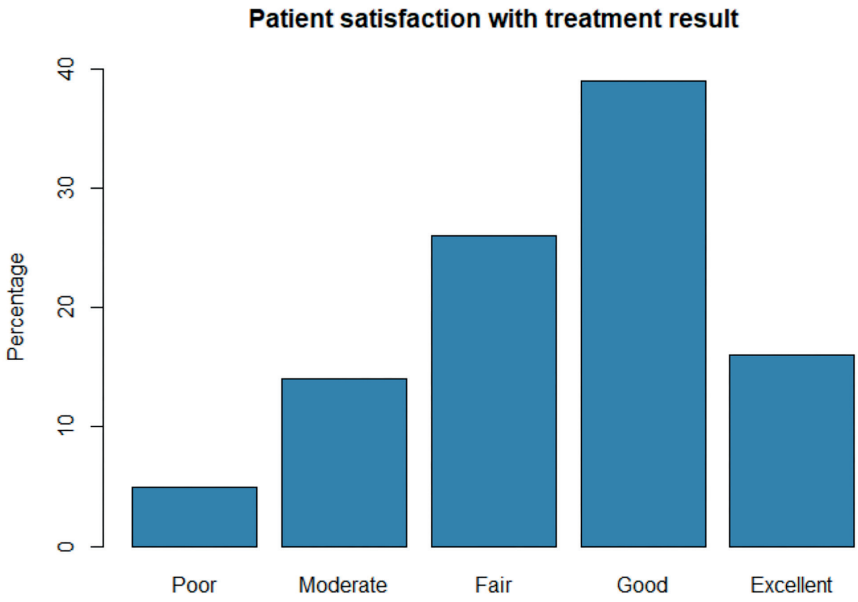


Figure 3. Patient satisfaction with the treatment result (n=163) is expressed in percentages.

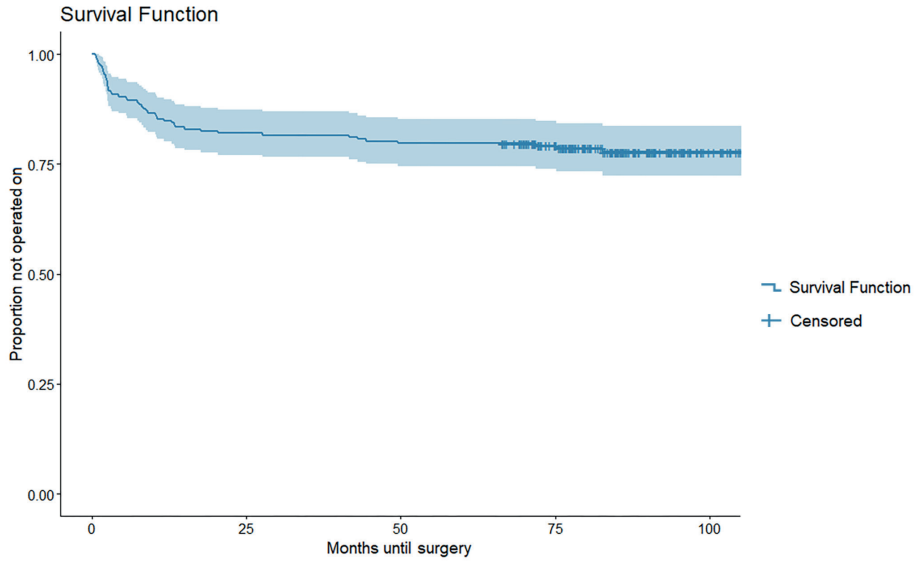


Figure 4. The blue line of the survival curve displays the proportion of patients converted to surgery and the time in months before deciding to convert to surgery. After a median follow-up time of 7.0 years, 21.7% decided to convert to surgery. The median time to decide to convert to surgery was 7.4 months. The light blue space shows the confidence interval. Censoring means that the patient did not convert to surgery at the time of conducting this study.

Table 1. Patient characteristics at baseline

Variables		Sample 1 (N=170)	Sample 2 (N=217)
Age, mean ± SD		59 ± 8	59 ± 8
Sex, n (%)	Female	129 (76)	172 (79)
Symptom duration (months), median [Q1-Q3]		12 [6 – 36]	12 [6 – 36]
Dominant hand, n (%)	Left	14 (8)	19 (9)
	Right	147 (87)	189 (87)
	Both	9 (5)	9 (4)
Treated hand, n (%)	Left	96 (57)	114 (53)
	Right	74 (44)	103 (48)
Type of work, n (%)	Unemployed	70 (41)	83 (38)
	Light physical work	37 (22)	52 (24)
	Moderate physical work	41 (24)	54 (25)
	Heavy physical work	22 (13)	28 (13)
Michigan Hand Outcomes Questionnaire at baseline, mean ± SD	Total score	61 ± 14	59 ± 14
	ADL score	69 ± 19	66 ± 20
	Pain score	49 ± 16	46 ± 16
	Function score	58 ± 15	57 ± 15
	Esthetics score	84 ± 20	83 ± 20
	Satisfaction score	44 ± 22	42 ± 22
	Work score	63 ± 25	61 ± 25

Table 2. Outcomes for the Michigan Hand Outcomes Questionnaire (MHQ) total score and different subscales (score range 0-100, higher scores indicate better function and less pain) of sample 1. Significance testing was performed using linear mixed model analysis. Δ shows the models' estimated marginal mean difference between the given timepoints. NS indicates not significant.* indicates a significant p-value of <0.025. ** indicates a significant p-value of <0.010. * indicates a significant p-value of <0.0001**

Variable	Δ baseline to 3 mo	Δ baseline to 12 mo	Δ baseline to 5 y	Δ 3 mo to 12 mo	Δ 3 mo to 5y	Δ 12mo to 5y
MHQ Total score	8.8 ***	10.0 ***	14.1 ***	1.2 NS	5.3 ***	4.2 **
MHQ ADL score	***5.6*** **	***9.0*** ***	***13.4*** ***	***3.4 NS***	***7.8*** ***	***4.4*** *
MHQ Pain score	***12.7*** ***	***15.2*** ***	***17.9*** ***	***2.5 NS***	***5.2*** **	***2.7 NS***
MHQ Function score	4.1 **	5.1 **	9.7 ***	1.0 NS	5.7 ***	4.6 **
MHQ Esthetics score	4.5 *	1.8 NS	4.4 NS	-2.6 NS	-0.1 NS	2.5 NS
MHQ Satisfaction score	18.9 ***	18.7 ***	23.0 ***	-0.2 NS	4.1 NS	4.3 NS
MHQ Work score	6.8 **	9.8 ***	16.1 ***	3.0 NS	9.3 ***	6.3 *

APPENDIX 1. INTERVENTION

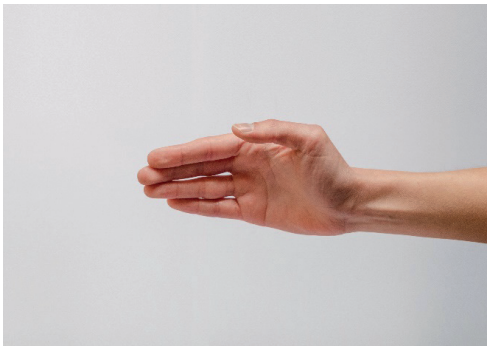
Week 0-6

Instructions on wearing the orthosis 24 hours a day, if possible, and exercise therapy to optimize thumb positioning. This included performing coordinative and isometric exercises 4-6 times a day, 10-15 repetitions for the m. Extensor Pollicis Brevis, m. Abductor Pollicis Brevis, m. Abductor Pollicis Longus, m. Opponens Pollicis, and m. Flexor Pollicis Brevis. Starting with coordinative/selective muscle activation, and continue to build coordination with more functional exercises, such as picking up and holding various light objects while maintaining thumb position.

Examples of exercises:

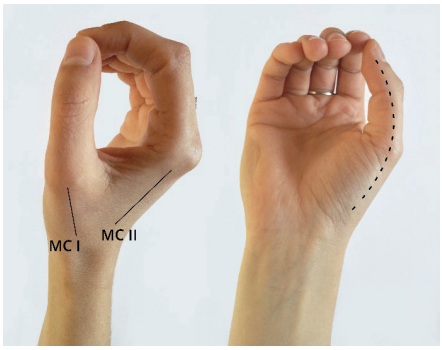
1. M. Abductor Pollicis Brevis/longus and m. Opponens Pollicis – coordination exercise

2. M. Extensor Pollicis Brevis – coordination exercise



3. M. Flexor Pollicis Brevis – coordination exercise

4. Thumb position (MC I= first metacarpal, MC II= second metacarpal)

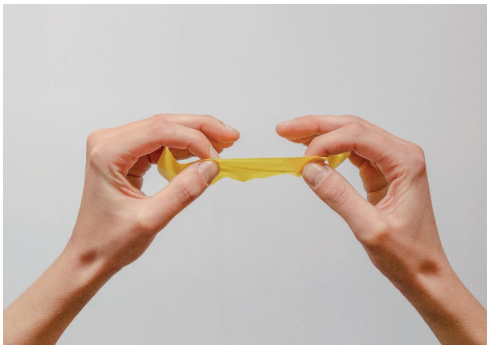


Week 6- 3 months

In the subsequent six weeks, the orthosis was phased out, and the exercises were carried out 2-3 times a day, focused on using the stable position of the thumb during daily activities and improving thenar strength. Building on using the active stability with greater force. Force was applied for 2-3 seconds and build up to multiple sessions such as 3x10-15 repetitions. Exercises were carried out in with 2-point grip and was expanded by using the 3-point grip and lateral key pinch.

Examples of exercises:

5. Pulling a rubber band in closed and correct pinch



6. Applying manual resistance at the proximal phalanx in closed and correct pinch

7. Pulling a rubber band in open chain

8. Applying manual resistance at the proximal phalanx in open chain

9. Active dynamic stability by using a clothespin

6

PREVALENCE OF COMPLICATIONS AND ASSOCIATION WITH PATIENT-REPORTED OUTCOMES AFTER TRAPEZIECTOMY WITH A WEILBY SLING: A COHORT STUDY

Hoogendam, L., van der Oest, M. J. W., Vermeulen, G. M., Feitz, R., Hovius, S. E. R., Zuidam, J. M., Slijper, H. P., Selles, R. W., The Hand-Wrist Study Group, Wouters, R. M. (2023). *The Journal of hand surgery*, *48*(5), 469–478.

ABSTRACT

Purpose

The primary aim of this study was to report complications during the first year after trapeziectomy with Weilby-sling using a standardized tool designed by the International Consortium for Health Outcome Measures (ICHOM). The secondary aim was to determine the association of complications and patient-reported outcomes twelve months postoperatively.

Methods

We included patients treated with a trapeziectomy with Weilby-sling between November 2013 and December 2018. All complications during the first year were scored using the International Consortium for Health Outcomes Measurement Complications in Hand and Wrist conditions (ICHAW) tool. Pain and hand function were measured preoperatively and twelve months postoperatively using the Michigan Hand outcomes Questionnaire. Minimally Important Change (MIC) thresholds of 18.6 for MHQ pain and 9.4 for MHQ function were used to determine clinical importance.

Results

Of 531 patients following trapeziectomy with Weilby sling, 65% had an uneventful recovery, 16% experienced ICHAW Grade 1 deviations only, and 19% experienced Grade 2 or 3 deviations, including requiring antibiotics, corticosteroid injections, or additional surgery. On average, patients improved in pain and hand function, even in the presence of ICHAW events. While all ICHAW grades were associated with poorer patient-reported outcomes twelve months postoperatively, Grade 2 and 3 exceeded the MIC threshold for pain, function, or both.

Conclusions

In 531 patients, 65% had an uneventful recovery, 16% experienced ICHAW Grade 1 deviations only, and 19% experienced Grade 2 or 3 deviations. We recommend describing Grade 1 as “adverse protocol deviations” and Grade 2 and 3 as complications, due to clinically-relevant poorer patient-reported outcomes twelve months postoperatively. The ICHAW is a promising tool in order to systematically evaluate and compare complications in hand surgery, although we recommend further evaluation.

INTRODUCTION

Trapeziectomy with ligament reconstruction and tendon interposition (LRTI) is a common surgical procedure, as it is the preferred surgical option for thumb base osteoarthritis for 72-89% of surgeons (1-3). Several LRTI techniques are available (4), including a flexor carpi radialis (FCR) sling according to Weilby (5), which is often performed (6). In this procedure, an FCR strip is wrapped around the abductor pollicis longus in a figure-of-eight pattern to support the first metacarpal (Figure 1).

While studies consistently show improvement in pain and hand function (4) following trapeziectomy with LRTI, a wide range of complication rates has been reported, ranging from 0.2% (7) to 53% (8), and 18% (9) to 35% (10) specifically for trapeziectomy with Weilby sling. These differences in complication rates are likely due to the use of different definitions and follow-up periods.

To overcome this problem, the International Consortium for Health Outcome Measurement (ICHOM) hand and wrist working group recently defined an international standard for reporting complications in hand surgery (11), based on the Clavien-Dindo complication classification system used in general surgery (12). In the newly developed ICHOM Complications in Hand and Wrist conditions (ICHAW) tool, all events related to the treatment that fall outside the expected recovery during the first twelve months following the initial surgery are considered a complication.

The ICHAW tool has already been used to classify complications following ulnar shortening osteotomy and trigger finger release (13, 14). Notably, it yields higher complication rates than previously reported. While this is in part expected due to the inclusive definition and the follow-up period of one year, it also raises concerns about whether all events classified as complications by the ICHAW tool are actually relevant to patients and influence other outcomes. A change beyond the Minimally Important Change can be considered clinically relevant.

Therefore, the primary aim of the present study was to report complications during the first year after trapeziectomy with a Weilby-sling, using the ICHAW tool (11). As secondary aim, we investigated how complications are associated with patient-reported pain and hand function during the first year after surgery.

METHODS

Study design and setting

We conducted a retrospective study using data of patients that elected trapeziectomy and ligament reconstruction according to Weilby for primary TMC OA (5). Patient-reported

outcome data were collected prospectively, while complication data were collected by retrospective review of the patient records. We included patients treated between December 2011 and December 2018 at Xpert Clinics, comprising 26 locations and 23 European Board certified (FESSH) hand surgeons. The cohort and data collection (15) and their use on daily clinical care (16) were previously described in more detail.

As part of routine outcome measurements, all patients were invited to complete patient-reported outcome measurements (PROMs) at fixed time points, i.e., prior to surgery, three months after surgery, and twelve months after surgery (15). Data were collected between November 2013 and December 2019. All patients provided written informed consent. Institutional Board Review was obtained from the ethics committee of the Erasmus Medical center that approved our study protocol (MEC-2018-1088).

Participants

We included patients who completed the Michigan Hand Outcomes Questionnaire (MHQ) preoperatively and twelve months postoperatively. Patients with posttraumatic TMC OA, previous thumb base surgery, previous major hand or wrist surgery (e.g., proximal row carpectomy), or isolated scaphotrapeziotrapezoid (STT) OA were excluded. We also excluded patients undergoing surgical treatment on the contralateral side within twelve months to avoid interference of treatment of other hand or wrist conditions on the PROM scores.

Surgical technique

Patient records were studied by the first author to confirm that the same modified Weilby technique (Figure 1, reprinted with permission from Xpert Clinics (17)) was used in every patient. A description of the surgical technique is included in Supplement A. During surgery, the STT joint was inspected. When indicated, partial resection of the scaphoid, trapezoid, or both, was performed, depending on surgeons' preference.

Rehabilitation

All hand therapists across all locations were instructed to follow the standard rehabilitation protocol (including immobilization and exercises) in our clinic. During the study period, the duration of wearing a cast was decreased, changing the immobilization regime. These two different immobilization regimens and the exercises have been extensively described by Tsehaie et al. (18), but have been shown to yield similar outcomes; there were no differences in complication rates and there was non-inferiority in all PROMs. The current protocol (from October 2015 onwards) is short immobilization, in which a plaster cast is applied for 3-5 days postoperatively. Afterwards, a thumb spica splint that also provided immobilization of the wrist was applied until four weeks postoperatively. From four to eight

weeks postoperatively, patients received a thumb butterfly splint that was phased out from week eight to ten weeks postoperatively. Routine check-up visits with the surgeon were scheduled at three months and twelve months postoperatively. Additional check-up visits were easily scheduled when the patient or treating hand therapist considered this beneficial.

Measurements

Medical history, Eaton-Glickel score (19), complications, and reoperations within the first year after surgery were obtained from patient records by the first author (not involved in treatment). In accordance with the ICHAW tool, all deviations from the expected treatment course that are related to the treatment during the first year after surgery were considered complications. Treatments for pre-existing conditions that followed another treatment were not considered complications (e.g., treatment for Dupuytren's contracture of the fourth digit following surgery for thumb base osteoarthritis). In addition to the rehabilitation protocol described above, we considered the following as part of the expected treatment course: routine prescription of analgesics (opioids or less strong analgesics for five to ten days, i.e., one prescription), sutures removed at seven to fourteen days, cast removed at three to five days, hand therapy phased out at three months, orthosis phased out at three months, and no additional treatment recommendations from the three months check-up visit onwards.

In the ICHAW guidelines (Supplementary Table 1), the severity of a complication (i.e., grade) is based on the treatment it requires. When a complication was not sufficiently relieved with minimally invasive treatment and more invasive treatment was given, this was only scored as a complication once and the highest grade was reported.

Patient-reported outcomes

The MHQ (20) is a PROM with good reliability, validity, and responsiveness for TMC OA patients (21). The MHQ consists of six domains (pain, hand function, aesthetics, work, activities of daily life, and satisfaction with hand function), each with a score ranging from 0-100 (0 = poorest function, 100 = ideal function). For this study, the MHQ subscales pain and hand function were used. We included the baseline and twelve months postoperative MHQ scores for the analyses. When patients also completed the MHQ three months postoperatively, this data was only included in the figures. Minimally Important Change (MIC) thresholds of 18.6 for MHQ pain and 9.4 for MHQ function were used to determine clinical importance (22).

Study size

The number of patients treated during the study period determined the sample size, making this a convenience sample. In our post-doc effect size calculation, we found that we could detect a small effect size of 0.12 (Cohen's d) with the number of available patients and 80%

power to evaluate whether pain and hand function changed between pre- and postoperative. Additionally, for the linear regression model that included complication grade, we could detect a medium effect size of 0.04 (Cohen's f2) (23).

To examine whether patients completing the MHQ preoperatively and twelve months postoperatively (responders) differed from patients who did not complete these PROMs (non-responder), we performed a non-responder analysis where we compared patient characteristics (i.e., age, sex, symptom duration, hand dominance, affected side, and occupational intensity).

Statistical analysis

T-tests were used to compare normally distributed continuous outcomes. Wilcoxon tests were used to compare non-normally distributed continuous outcomes. Chi-square tests were used for categorical outcomes.

Multivariable linear regression was used to estimate the association between complications and patient-reported outcomes twelve months postoperatively, corrected for patient characteristics, treatment characteristics, and preoperative pain and hand function scores. We checked that the regression analyses complied with the model's assumptions. The adjusted explained variance (R^2), which accounts for the number of variables in the model, was calculated. Additionally, we fitted linear mixed models to again assess clustering within location and we found Intraclass Correlation Coefficients of 0.000 and 0.011 for pain and hand function, respectively. This indicates that the variance in outcome mainly exists within locations (caused by patient variation) and not between locations, therefore we used the multivariable linear regression models.

We considered a p-value <0.05 as statistically significant. As multiple testing correction was not performed; the results of the analyses on the secondary outcomes should be interpreted as exploratory (24).

RESULTS

We included 531 patients (Figure 2) for analysis. Most patients were female (78%), and the mean age was 61 (SD 8) (Table 1). We excluded 384 patients due to non-response on the PROMs of interest. We found no differences in patient characteristics in the non-responder analysis (Supplemental Table 2). In 13%, a concomitant partial resection of the ST joint was performed. Concomitant surgery was performed in 11% of the patients, mostly a carpal tunnel release (38%) (Supplemental Table 3).

Primary outcome

In 65% of all patients, no deviations from the expected course within one year were reported, according to the ICHAW tool (Figure 3). 35% of all patients (95% CI 31%; 39%) experienced any deviation, with 16% of the patients (95% CI 13%; 19%) only experiencing a Grade 1 event and 19% (95% CI 16%; 23%) experiencing at least a Grade 2 or 3 event. We recorded 275 separate ICHAW events in 187 patients (Table 2) (Supplemental figure 1).

A Grade 1 event occurred in 24% of all patients (e.g., pain requiring additional hand therapy, splinting, or analgesics), and 14% experienced a Grade 2 event (e.g., FCR tendinitis requiring corticosteroid injections) (Table 2). 36 patients (7%) required surgical intervention under local (Grade 3A) or regional/general anesthesia (Grade 3B).

Grade 3A events mostly comprised surgical treatment of De Quervain's tenosynovitis (DQ) or surgical treatment of trigger thumb. Grade 3B events mostly comprised revisions within a year (3%), mainly consisting of releases of the Weilby-sling due to pain and tendinitis (n=8) and (partial) resections of the scaphoid, the trapezoid, or both (n=6). The median time to revision surgery was 39 weeks (IQR 31 – 45 weeks). Grade 3C (CRPS) was present in 11 patients (2%). Because the timing of events was an important determinant of whether the event fell outside of the expected recovery, we provide an overview of the number of complications per week per grade (Supplemental figure 2) and the number of Grade 1, 2, and 3 complications per week in more detail (Supplemental figure 3ABC).

As a sensitivity analysis, we excluded treatment for DQ and trigger thumb as ICHAW events. This resulted in 30% of all patients experiencing a deviation and 12% experiencing Grade 2 or 3 events.

Secondary outcomes

Between preoperatively and twelve months postoperatively, MHQ pain scores improved on average by 29 (95% CI 26, 31) from 34 (95% 33, 36) to 63 (95% CI 61, 65) (Figure 4A). MHQ hand function improved, on average, by 15 points (95% CI 13, 17) from 49 (95% CI 48, 51) to 64 (95% CI 63, 66) (Figure 5A). The distributions of MHQ pain and hand function scores pre- and postoperatively are shown in Supplemental figures 4A and 4B, respectively.

To evaluate the impact of complications on recovery, we assessed improvement in MHQ pain and hand function during the first year post-surgery separately for patients with and without complications (Figure 4A and Figure 5A, respectively). Both patients with and patients without ICHAW event experienced significant improvement in pain and hand function ($p<0.05$). After correction for baseline characteristics, we found that patients with an ICHAW event scored 14 points (95% CI 10, 17) worse on MHQ pain, and 11 points worse (95% CI 8, 14) on hand function twelve months postoperatively compared to patients without complications.

Additionally, ICHAW grade was significantly associated with both pain and hand function twelve months postoperatively. Figure 4B and 5B demonstrate how ICHAW grade affected MHQ pain scores and hand function scores, respectively. After correction for baseline characteristics, MHQ pain scores at twelve months were 10 points worse (95% CI -5, -15) for patients with a Grade 1 event increasing up to 28 points worse (95% -16, -41) for patients with a Grade 3C complication compared to patients without complications (Table 3). The impact of ICHAW grade on hand function score was very similar (Table 3). From Grade 2 onwards, there was a clinically relevant worsening in pain, hand function, or both.

DISCUSSION

In our cohort of 531 patients with primary TMC OA treated with trapeziectomy and Weilby sling, we found that, according to the ICHAW, no deviations occurred in 65% of the patients, 16% experienced an ICHAW Grade 1 deviation and 19% experienced an ICHAW Grade 2 or 3 deviation from expected recovery. Although we found that patients improved in pain and hand function following trapeziectomy and Weilby sling, having an ICHAW event was associated with poorer patient-reported outcomes twelve months after surgery, with higher ICHAW grades being associated with worse patient-reported outcomes. From Grade 2 onwards, the difference in patient-reported outcomes exceeds the Minimally Important Change, indicating a clinically relevant difference.

Considering previously reported complication rates of 0.2%-53% for trapeziectomy with LRTI (7, 8), and 18%-35% (9, 10) for the Weilby sling specifically, the overall ICHAW event rate of 35% we found, appears substantial, but not unexpected. This high rate is possibly explained by a large number of Grade 1 deviations, i.e., events requiring additional hand therapy or analgesics, which may be reported less frequently in other studies. Grade 1 deviations may be seen and treated more often by hand therapists or general practitioners, and could therefore be less in the scope of hand surgeons.

Upon applying the ICHAW guidelines, we found Grade 1 deviations particularly difficult to ascertain. The treatments that fall under Grade 1 deviations (e.g., additional hand therapy or silicone gel sheets), are prescribed easily, making it difficult to determine whether these additional treatments were prescribed because they were potentially beneficial of whether they were really necessary. We found that Grade 1 deviations were associated with worse patient-reported outcomes, but it is debatable whether this difference is clinically relevant. We recommend that ICHAW Grade 1 deviations continue to be registered and reported to provide more insight into recovery following hand surgery procedures, but that these may not be considered a complication. Rather, we propose the term “adverse protocol deviation” to classify Grade 1 deviations.

Grade 3 events, including CRPS, revision surgery, or surgical treatment of DQ or trigger thumb, occurred in 8% of all patients in our study. Previously, a CRPS rate of 5% and a revision rate due to scapho-first metacarpal impingement of 1.5% has been reported in patients that underwent trapeziectomy with LRTI (25). We found a slightly lower CRPS rate of 2%, but a higher revision rate of 3%. While there is no conclusive evidence that DQ and trigger thumb are causally related to trapeziectomy with LRTI, they have been reported previously in literature post-trapeziectomy (5, 26). Hypothetically, the FCR suspension with the abductor pollicis longus may evoke friction and edema and subsequent tenovaginitis at the first extensor compartment (i.e., DQ). Similarly, the increase zigzag deformity in the thumb following surgery (27, 28) combined with postoperative edema may evoke a flexor pollicis longus tenovaginitis (i.e., trigger thumb). Future studies may investigate whether the development of DQ and trigger thumb are truly related to the surgery and whether treatment should therefore be considered a complication following trapeziectomy and LRTI.

Our study has several limitations, including the amount of non-response due to the observational setting of this study. However, in the non-responder analysis, we found no differences in patient characteristics between responders and non-responders, suggesting that the patients included in this study are similar to patients who did not complete all PROMs. Our observational setting may also be considered a strength, as it may better reflect daily clinical hand surgery practice.

In contrast to the prospective collection of PROMs, complications were scored retrospectively based on electronic patient records because the ICHAW tool has only recently been developed. Despite our standardized procedures, collecting data retrospectively is more sensitive to bias than prospective data collection. To obtain more reliable complication estimates, we recommend recording complications prospectively, preferably by clinicians. Still, as the ICHAW tool uses the administered treatment to determine complication grade, this will likely be reasonably well-reported in patient records, particularly for Grades 2 and 3. The inter-rater reliability of the ICHAW needs to be studied in future research.

Finally, we included a single surgical treatment option for TMC OA in our study because most surgeons in our clinic prefer this technique. This is in line with recent survey studies, showing that most surgeons (72-89%) prefer to treat TMC OA with trapeziectomy and LRTI (1, 2, 6), despite that current evidence suggests no benefit from LRTI in addition to trapeziectomy (4, 29). Since the ICHAW tool has not yet been applied to other surgical techniques for treating TMC OA, comparing our complication rate to, for example isolated trapeziectomy or implants, is challenging. Future comparative studies using the ICHAW tool are needed to assess this.

In conclusion, this study reports complications using the ICHAW tool, following trapeziectomy with a Weilby-sling. Of 531 patients, 65% had an uneventful recovery, 16% experienced ICHAW Grade 1 deviations only, and 19% experienced Grade 2 or 3 deviations.

Considering that Grade 2 and 3 deviations were associated with poorer patient-reported outcomes twelve months postoperatively, we propose to classify Grade 1 deviations as "adverse protocol deviations" instead of complications and Grade 2 and 3 as complications in future studies. While the ICHAW tool is promising to systematically evaluate and compare complications in hand surgery, we recommend psychometric evaluation of the ICHAW tool and possibly further defining the guidelines to optimally define, register, and compare complications in hand surgery.

ACKNOWLEDGMENTS

The authors thank all patients who participated and allowed their data to be anonymously used for the present study.

REFERENCES

1. Ottenhoff JSE, Teunis T, Janssen SJ, Mink van der Molen AB, Ring D. Variation in Offer of Operative Treatment to Patients With Trapeziometacarpal Osteoarthritis. J Hand Surg Am. 2020;45(2):123-30 e1.
2. Yuan F, Aliu O, Chung KC, Mahmoudi E. Evidence-Based Practice in the Surgical Treatment of Thumb Carpometacarpal Joint Arthritis. J Hand Surg Am. 2017;42(2):104-12 e1.
3. Wilcke M, Roginski M, Astrom M, Arner M. A registry based analysis of the patient reported outcome after surgery for trapeziometacarpal joint osteoarthritis. BMC Musculoskelet Disord. 2020;21(1):63.
4. Wajon A, Carr E, Edmunds I, Ada L. Surgery for thumb (trapeziometacarpal joint) osteoarthritis. Cochrane Database Syst Rev. 2009(4):CD004631.
5. Weilby A. Tendon interposition arthroplasty of the first carpo-metacarpal joint. J Hand Surg Br. 1988;13(4):421-5.
6. Deutch Z, Niedermeier SR, Awan HM. Surgeon Preference, Influence, and Treatment of Thumb Carpometacarpal Arthritis. Hand (N Y). 2018;13(4):403-11.
7. Lane JC, Craig R, Rees JL, Gardiner M, Mikhail MM, Riley N, et al. Low rates of serious complications and further procedures following surgery for base of thumb osteoarthritis: analysis of a national cohort of 43 076 surgeries. BMJ Open. 2021;11(7):e045614.
8. Salem H, Davis TR. Six year outcome excision of the trapezium for trapeziometacarpal joint osteoarthritis: is it improved by ligament reconstruction and temporary Kirschner wire insertion? J Hand Surg Eur Vol. 2012;37(3):211-9.
9. Nylen S, Juhlin LJ, Lugnegard H. Weilby tendon interposition arthroplasty for osteoarthritis of the trapezial joints. J Hand Surg Br. 1987;12(1):68-72.
10. Naram A, Lyons K, Rothkopf DM, Calkins ER, Breen T, Jones M, et al. Increased Complications in Trapeziectomy With Ligament Reconstruction and Tendon Interposition Compared With Trapeziectomy Alone. Hand (N Y). 2016;11(1):78-82.
11. ICHOM Hand and Wrist Conditions. 2020 [updated April 2020. Available from: https://www.ichom.org/portfolio/hand-and-wrist-conditions/.
12. Clavien PA, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD, et al. The Clavien-Dindo classification of surgical complications: five-year experience. Ann Surg. 2009;250(2):187-96.
13. Koopman JE, Zweedijk BE, Hundepool CA, Duraku LS, Smit J, Wouters RM, et al. Prevalence and Risk Factors for Postoperative Complications Following Open A1 Pulley Release for a Trigger Finger or Thumb. J Hand Surg Am. 2022;47(9):823-33.
14. Teunissen JS, Wouters RM, Al Shaer S, Zophel OT, Vermeulen GM, Hovius SER, et al. Outcomes of ulna shortening osteotomy: a cohort analysis of 106 patients. J Orthop Traumatol. 2022;23(1):1.
15. Selles RW, Wouters RM, Poelstra R, van der Oest MJW, Porsius JT, Hovius SER, et al. Routine health outcome measurement: Development, design, and implementation of the Hand and Wrist Cohort. Plast Reconstr Surg. 2020.
16. Feitz R, Van Kooij YE, Ter Stege MHP, Van der Oest MJW, Souer JS, Wouters RM, et al. Closing the loop: a 10-year experience with routine outcome measurements to improve treatment in hand surgery. EFORT Open Reviews. 2021;Vol. 6(No. 6).
17. Xpert Clinics. Trapezium excisie en Weilby sling [Available from: https://www.xpertclinic.nl/aandoeningen/duimbasis-artrose#process-tab-3.

18. Tsehaie J, Wouters RM, Feitz R, Slijper HP, Hovius SER, Selles RW, et al. Shorter vs Longer Immobilization After Surgery for Thumb Carpometacarpal Osteoarthritis: A Propensity Score-Matched Study. Arch Phys Med Rehabil. 2019;100(11):2022-31 e1.

19. Eaton RG, Glickel SZ. Trapeziometacarpal osteoarthritis. Staging as a rationale for treatment. Hand Clin. 1987;3(4):455-71.

20. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg Am. 1998;23(4):575-87.

21. Marks M, Audige L, Herren DB, Schindele S, Nelissen RG, Vliet Vlieland TP. Measurement properties of the German Michigan Hand Outcomes Questionnaire in patients with trapeziometacarpal osteoarthritis. Arthritis Care Res (Hoboken). 2014;66(2):245-52.

22. Hoogendam L, Koopman JE, van Kooij YE, Feitz R, Hundepool CA, Zhou C, et al. What Are the Minimally Important Changes of Four Commonly Used Patient-reported Outcome Measures for 36 Hand and Wrist Condition-Treatment Combinations? Clin Orthop Relat Res. 2022;480(6):1152-66.

23. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, N.J.: L. Erlbaum Associates; 1988.

24. Li G, Taljaard M, Van den Heuvel ER, Levine MA, Cook DJ, Wells GA, et al. An introduction to multiplicity issues in clinical trials: the what, why, when and how. Int J Epidemiol. 2017;46(2):746-55.

25. Saab M, Chick G. Trapeziectomy for trapeziometacarpal osteoarthritis. Bone Jt Open. 2021;2(3):141-9.

26. Vermeulen GM, Brink SM, Sluiter J, Elias SG, Hovius SE, Moojen TM. Ligament reconstruction arthroplasty for primary thumb carpometacarpal osteoarthritis (weilby technique): prospective cohort study. J Hand Surg Am. 2009;34(8):1393-401.

27. Komura S, Hirakawa A, Masuda T, Nohara M, Kimura A, Matsushita Y, et al. Chronological changes in surgical outcomes after trapeziectomy with ligament reconstruction and tendon interposition arthroplasty for thumb carpometacarpal osteoarthritis. Orthop Traumatol Surg Res. 2020;106(2):357-64.

28. Van Heest AE, Kallemeier P. Thumb carpal metacarpal arthritis. J Am Acad Orthop Surg. 2008;16(3):140-51.

29. Vermeulen GM, Slijper H, Feitz R, Hovius SE, Moojen TM, Selles RW. Surgical management of primary thumb carpometacarpal osteoarthritis: a systematic review. J Hand Surg Am. 2011;36(1):157-69.

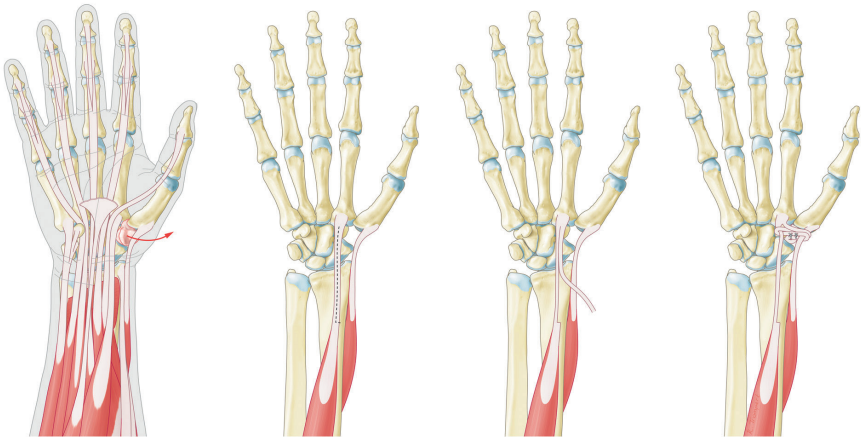


Figure 1. Surgical technique of the Weilby procedure. Reprinted with permission from Xpert Clinics(17)). Copyright by Xpert Clinics.

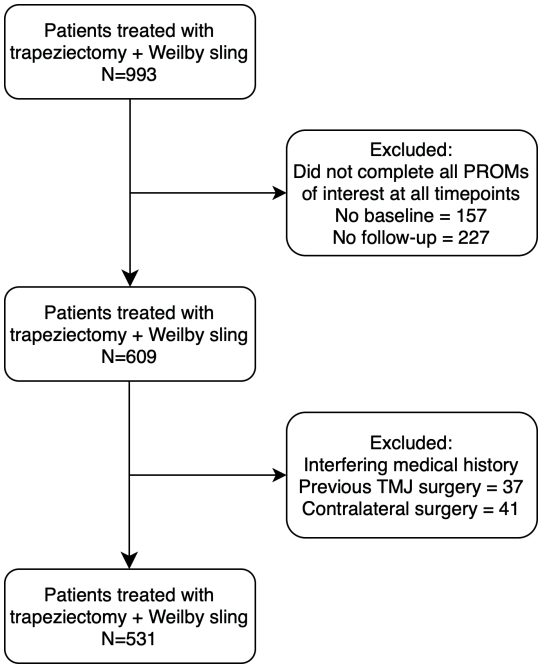


Figure 2. Flowchart of included patients

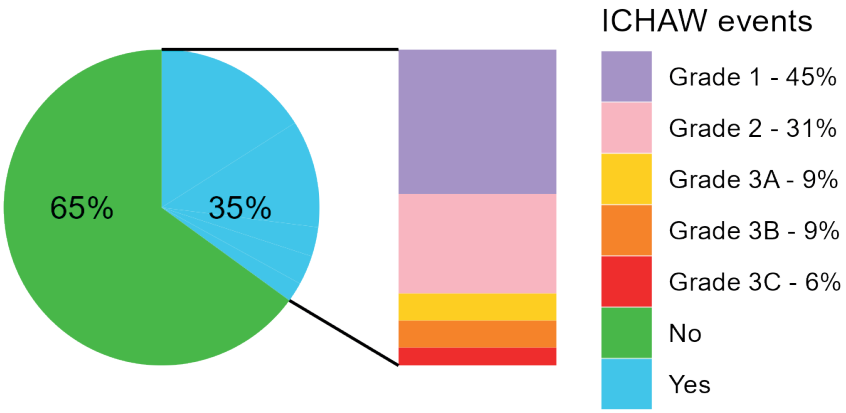


Figure 3. Distribution of ICHAW events during the first year after trapeziectomy with a Weilby sling. The pie chart depicts the proportion of patients with a complication. The bar chart shows the distribution of complication severity for patients with a complication. Because a small number of patients experienced multiple events, we report the most severe ICHAW grade per patient.

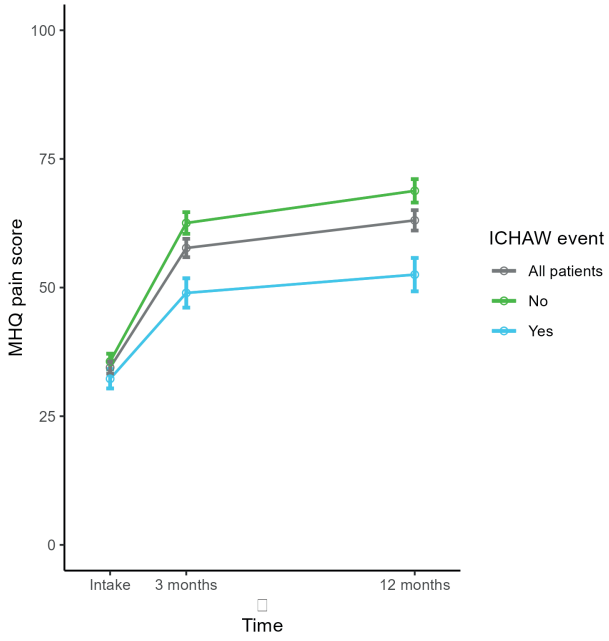


Figure 4A. Mean MHQ pain during the first year after trapeziectomy with a Weilby sling, categorized by the occurrence of an ICHAW event. The error bars represent the 95% confidence interval.

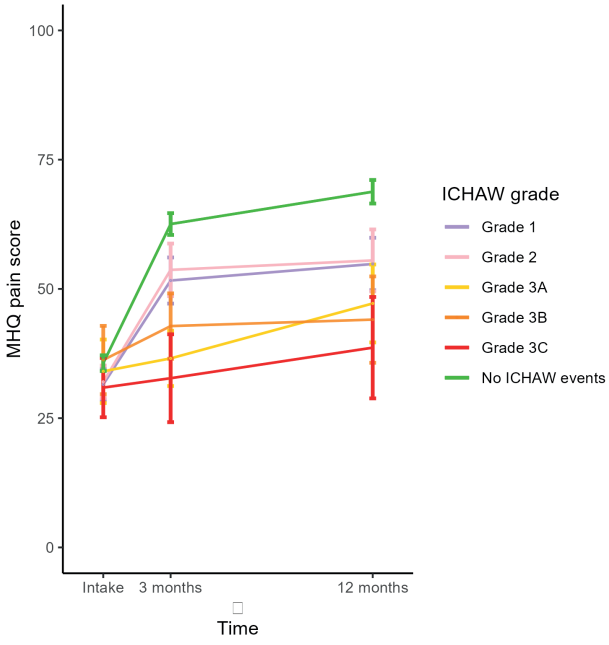


Figure 4B. Mean MHQ pain during the first year after trapeziectomy with a Weilby sling, categorized by ICHAW grade. The error bars represent the 95% confidence interval.

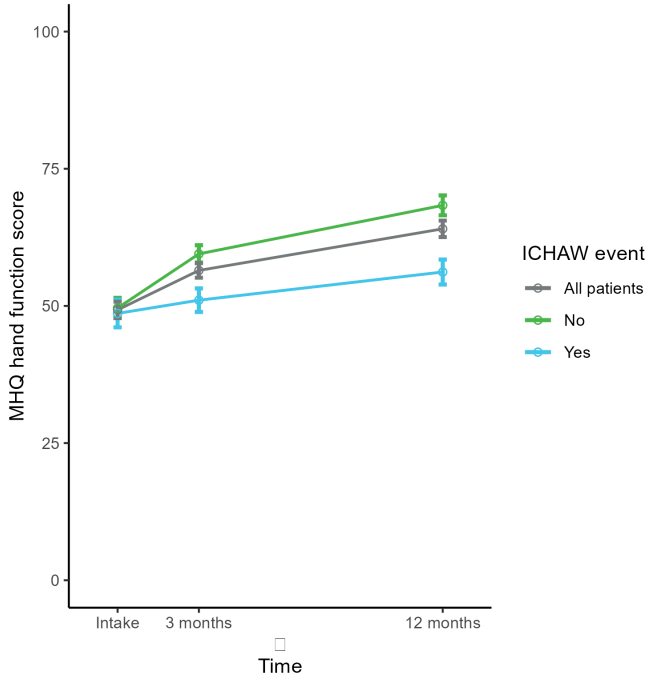


Figure 5A. Mean MHQ hand function during the first year after trapeziectomy with a Weilby sling, categorized by the occurrence of an ICHAW event. The error bars represent the 95% confidence interval.

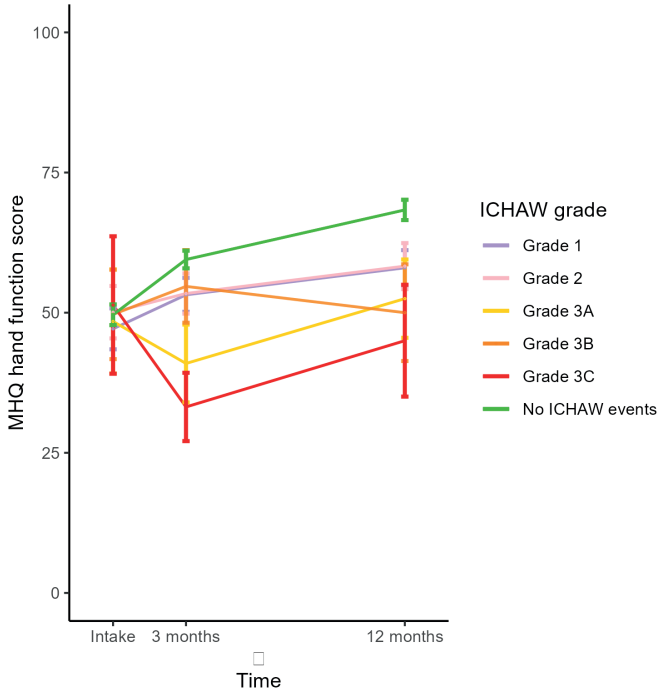


Figure 5B. Mean MHQ hand function during the first year after trapeziectomy with a Weilby sling, categorized by ICHAW grade. The error bars represent the 95% confidence interval.

Table 1. Patient characteristics and patient reported outcomes prior to surgery of the included patients. Additionally, patient characteristics and patient reported outcomes prior to surgery are shown for patients with and without an event according to the ICHAW tool in the first year after trapeziectomy and Weilby-sling.

	All (N=531)	Patients with an ICHAW event (N=187)	Patients without an ICHAW event (N=344)
Age (in years)	61 (8)	60 (9)	62 (8)
Sex (%)			
Female	78	82	76
Duration of symptoms (in months) (median ± IQR)	24 (12 – 48)	24 (12 – 48)	24 (12 – 50)
Hand dominance (%)			
Right	84	83	85
Left	10	9	10
Both	6	8	5
Affected side (%)			
Right	49	54	46
Occupational intensity (%)			
Not employed	47	47	48
Light (e.g., working in an office)	18	15	19
Moderate (e.g., working in a shop)	22	21	23
Severe (e.g., working in construction)	13	17	10
BMI (in kg/m^2)	26 (4)	27 (4)	26 (4)
Diabetes (%)			
Yes	6	7	6
Smoking (%)			
Yes	13	16	11
Eaton Glickel grade (%)			
1	1	1	0
2	6	7	6
3	23	23	23
4	43	41	44
Not available	27	28	27
Concomitant partial resection of trapezoid, scaphoid, or both (%)			
Yes	13	11	13
Partial scaphoid resection	1		
Partial trapezoid resection	7		
Partial scaphotrapezoid resection	5		
Concomitant minor surgery (%)			
Yes	11	15	9
MHQ pain	34 (14)	32 (13)	36 (14)
MHQ hand function	49 (18)	49 (18)	50 (17)

Values are reported as mean (SD) unless stated otherwise.

Table 2. Overview of the number of ICHAW events and the number of patients who experienced that event, according to the ICHAW tool, during the first year after trapeziectomy and Weilby-sling. Because a selection of patients experienced multiple events, the number of complications does not equal the number of patients.

	ICHAW events (N)
Grade 1	140 events in 125 patients (i.e., 23.5% (95% CI 20.0; 27.3) of all patients had a Grade 1 event)
Pain requiring hand therapy, splinting or analgesia, additional to the normal rehabilitation protocol	86
Instability[a]	26
Prolonged or excessive swelling[b]	11
Sensory change[c]	6
Scar tenderness[d]	5
Tendon rupture (no intervention required)	4
Stenosing tenosynovitis of the thumb requiring hand therapy or splinting	2
Grade 2	81 events in 72 patients (i.e., 13.6% (95% CI 10.8; 16.8) of all patients had a Grade 2 event)
FCR tendinitis requiring corticosteroid injection	26
De Quervain's tenosynovitis requiring corticosteroid injection	25
Stenosing tenosynovitis of the thumb requiring corticosteroid injection	15
Infections requiring antibiotics	8
Pain requiring corticosteroid injection[e]	7
Grade 3	54 events in 43 patients (i.e., 8.1% (95% CI 5.9; 10.8) of all patients had a Grade 3 event)
Grade 3A	25 events in 25 patients (i.e., 4.7% (95% CI 3.1; 6.9) of all patients had a Grade 3A event)
Surgical treatment of de Quervain's tenosynovitis	11
Surgical treatment of stenosing tenosynovitis of the thumb	11
Pain requiring pain rehabilitation or second opinion	3
Grade 3B	18 events in 17 patients (i.e., 3.2% (95% CI 1.9; 5.1) of all patients had a Grade 3B event)
Revision surgery	17
Neuroma treatment	1
Grade 3C	11 events in 11 patients (i.e., 2.1% (95% CI 1.0; 3.7) of all patients had a Grade 3C event)
Complex regional pain syndrome (CRPS)	11
Overall	275 events, 187 patients (i.e., 35.2% (95% CI 31.2; 39.5) of all patients had an event)

[a] = Instability including first metacarpophalangeal joint collapse, first metacarpophalangeal hyperextension, Z-deformity, or trapezoidal collapse requiring a (permanent) splint and/or additional hand therapy

[b] = Excessive postoperative swelling requiring early removal of the plaster cast or prolonged swelling requiring bandaging

[c] = Persistent sensory change (such as numbness and tingling) not related to carpal, cubital, or radial tunnel syndrome

[d] = Scar tenderness for which silicone gel sheets were prescribed

[e] = Persistent pain requiring corticosteroid injections at the base of the first metacarpal or in the scaphotrapezial joint

Table 3. Regression coefficients and 95% confidence intervals for the association between ICHAW grade and MHQ pain score and MHQ hand function score twelve months postoperatively, respectively. In the regression analysis, we corrected for patient characteristics and preoperative MHQ pain score and MHQ hand function score, respectively. These regression coefficients can be interpreted as the mean difference in MHQ pain or MHQ hand function twelve months postoperatively for patients with an ICHAW event of a specific grade compared to a patient without an ICHAW event when all other variables (patient characteristics and preoperative MHQ score) remain the same

	Outcome			
	MHQ pain twelve months postoperatively		MHQ hand function twelve months postoperatively	
ICHAW grade (Ref = no event)	B (95% CI)	P-value	B (95% CI)	P-value
Grade 1 event	-10.4 (-5.3, -15.5)	<0.001	-8.7 (-4.8, -12.7)	<0.001
Grade 2 event	-11.6 (-5.8, -17.5)	<0.001	-10.5 (-6.0, -15.1)	<0.001
Grade 3A event	-19.2 (-8.6, -29.8)	<0.001	-15.4 (-7.2, -23.6)	<0.001
Grade 3B event	-23.5 (-13.0, -34.0)	<0.001	-18.0 (-9.8, -26.2)	<0.001
Grade 3C event	-28.2 (-15.6, -40.9)	<0.001	-23.7 (-13.9, -33.6)	<0.001
Adjusted R^2	0.22		0.18	

7

WHICH TENDON PLASTY HAS THE BEST OUTCOME? A COMPARISON OF FOUR TENDON PLASTY TECHNIQUES IN A LARGE COHORT OF PATIENTS WITH SYMPTOMATIC TRAPEZIOMETACARPAL OSTEOARTHRITIS

Hoogendam, L.*, Bink, T.*, de Lange, J., Selles, R. W., Colaris, J. W., Zuidam, J. M., Hovius, S. E. R., van der Heijden, B., The Hand-Wrist Study Group (2022). *Plastic and reconstructive surgery*, *150*(2), 364e–374e.

* Both authors contributed equally

ABSTRACT

Introduction

Trapeziometacarpal osteoarthritis (TMC OA) is commonly treated with a trapeziectomy combined with a form of tendon plasty. The type of tendon plasty used is based on surgeons' preference. The purpose of this observational study is to compare the outcomes of four different tendon plasties combined with trapeziectomy used to treat TMC OA: the Weilby, Burton-Pellegrini, Zancolli and Anchovy plasty.

Patients and Methods

Patients treated with a trapeziectomy followed by a tendon plasty completed patient-reported outcome measures at baseline and twelve months postoperatively. The primary outcome is the Michigan Hand outcomes Questionnaire (MHQ) pain subscale. Secondary outcomes are the minimal clinically important difference (MCID) of MHQ pain, MHQ hand function, satisfaction, and complication rate.

Results

793 patients received a trapeziectomy with a tendon plasty between November 2013 and December 2018. There was no difference in pain score after twelve months between the four tendon plasty techniques. Patients undergoing an Anchovy plasty had a higher chance of reaching the MCID for MHQ pain compared to the other techniques (OR 2.3 (95% CI 1.2-4.6)). Overall, more than 80% of the patients were satisfied with the treatment outcome, independent of which technique was used. Complication rates of the different techniques were similar.

Conclusion

Surgical treatment of TMC OA reduced pain after twelve months, independent of which tendon plasty was used. Patients receiving an Anchovy plasty were more likely to experience a clinically relevant improvement in pain, whilst having a similar hand function, satisfaction, and complication rate. This suggests that Anchovy plasty is the preferred tendon plasty.

INTRODUCTION

If conservative treatment of osteoarthritis of the trapeziometacarpal joint (TMC OA) fails, several surgical techniques are available. Surveys show surgeons prefer to combine trapeziectomy with a form of interposition and/or suspension plasty.[1-4] The choice for a specific plasty after trapeziectomy is arbitrary and based on surgeon preferences, as no technique seems to be superior.[5,6]

There have been both observational and randomized controlled trials comparing different types of tendon plasties following trapeziectomy to treat TMC OA. No differences in outcome have been found after at least one year of follow-up.[7-14]These studies only compared two types of tendon plasty techniques at the same time.

A recent registry based analysis from Sweden evaluated patient reported outcome measures (PROM) after trapeziectomy with or without ligament reconstruction and tendon interposition for TMC OA.[15] Improvement was similar for both trapeziectomy with or without ligament reconstruction and tendon interposition. This substantiates the already available evidence.[5,6] Unfortunately it was not possible to distinguish the type of tendon plasty performed. It is still unknown if any of the tendon plasty techniques used is superior. As tendon plasties are preferred, this would be an addition to current evidence on TMC OA treatment.[2]

The tendon plasties mainly used for TMC OA in our clinics are the Weilby technique, the Burton-Pellegrini technique, the Anchovy plasty technique and the Zancolli technique. The Zancolli technique stabilizes the base of the first metacarpal by wrapping either a strip of the abductor pollicis longus (APL) or an accessory APL around the flexor carpi radialis (FCR) tendon and sutures it onto itself, minimizing the space left by the trapeziectomy.[16] The Weilby and Burton-Pellegrini techniques both use a strip of the FCR tendon for stabilization. With the Weilby technique the strip of the FCR is intertwined with the APL tendon in a figure-of-eight fashion and locked together with sutures.[17] The Burton-Pellegrini techniques passes a strip of the FCR tendon through a bone tunnel in the base of the proximal first metacarpal and fixes this firmly to the periosteum.[18] Both techniques suture the remaining end of the FCR strip together to fill the trapezial void. The Anchovy plasty technique is the least extensive technique of the four, where only the trapezial void is filled with a rolled-up graft of either the palmaris longus (PL) tendon or a strip of the FCR tendon.[19]

This study will evaluate if one of the tendon plasty techniques is superior in terms of patient reported outcomes or complications. The primary aim is to compare the change in postoperative pain scores after twelve months between four different tendon plasties following trapeziectomy for the treatment of TMC OA. The secondary aims are to report the percentage of patients reaching the minimal clinically important difference (MCID) for pain

and to describe hand function, satisfaction with the treatment outcome, and postoperative complications.

PATIENTS AND METHODS

Study design

This is an observational cohort study using prospectively collected data in a consecutive, population-based sample, reported following the STROBE statement.[20] The ethics committee of the Erasmus University Medical Centre approved our study protocol. All patients provided consent for their data to be used in this study.

Setting

Patients treated at Xpert Clinic and Handtherapie Nederland were invited to complete several patient-reported outcome measures (PROMS) preoperatively and postoperatively. The cohort and data collection have previously been reported.[21] Xpert Clinic and Handtherapie Nederland comprise of 27 outpatient clinics for hand surgery and therapy in The Netherlands. Patient inclusion took place between November 2013 and December 2018.

Patients

Patients included in this study had symptomatic TMC OA and failed to improve despite hand therapy and splinting for at least three months according to protocol. All patients were surgically treated with a trapeziectomy, followed by a tendon plasty. The technique used for tendon plasty was based on surgeon's preference. The Weilby technique is performed most often at our centers, followed by respectively the techniques of Burton-Pellegrini, Anchovy plasty, and Zancolli.[16-19] Detailed descriptions of the techniques are stated in Supplement A. Patient records were studied by the first author to confirm that the surgical technique was performed in accordance with the described techniques.

As simple trapeziectomy is scarcely performed in our clinic, we could not include these patients in our study.

We excluded patients receiving corticosteroid injections within six weeks prior to the date of surgery, as this may affect baseline pain scores. Other exclusion criteria were posttraumatic TMC OA, former surgery to the trapeziometacarpal joint, concomitant partial excision of the proximal trapezoid or surgical treatment on the contralateral side within one year. However, patients undergoing minor additional interventions like a trigger finger release, a carpal tunnel release, or a ganglion excision were not excluded.

Rehabilitation

Standardized treatment protocols were used in postoperative rehabilitation. The thumb was immobilized in a spica cast for 3-5 days, after which the cast was replaced by a removable protective orthosis and a hand therapist started standardized hand therapy focused on reducing oedema and regaining functionality by increasing mobility, stability, and strength of the thumb. The removable protective orthosis was worn for at least six weeks, after which its use was phased out. Further details of the rehabilitation protocol have been previously reported.[22]

Michigan Hand Outcomes Questionnaire, Dutch version

Patients completed the Dutch version of the Michigan Hand outcomes Questionnaire (MHQ-Dutch). [23,24] This is a hand-specific patient-reported outcomes measurement, which contains six subscales; (1) overall hand function, (2) activities of daily living, (3) pain, (4) work performance, (5) aesthetics and (6) patient satisfaction with hand function. All 57 questions are answered on a five-point Likert scale. Outcome values range from 0 to 100, where a low score signifies poor function and a high score indicates good hand function. As primary outcome, the pain subscale is used, as this is most relevant for TMC OA.[25] In the pain subscale, a low score indicates severe pain, while a high score indicates little pain. The minimal clinically important change (MCID) of MHQ pain has been described to be 11 by Shauver et al. for patients with rheumatoid arthritis undergoing arthroplasty of the metacarpophalangeal joints.[26]

Patient satisfaction

Patient satisfaction was assessed by a questionnaire on the perceived treatment effect (5-point Likert scale: "poor", “mediocre”, “fair”, “good”, and “excellent”) and the patient's willingness to undergo treatment again (“yes”, “no”).[27] For analysis, we dichotomized patient satisfaction, considering “good” and “excellent” as satisfied and “fair”, “poor” and “mediocre” as not satisfied. As sensitivity analysis, the analysis was repeated with “fair”, “good”, and “excellent” classified as satisfied.

Patients were invited to complete the MHQ prior to surgery. Three months and twelve months postoperatively we invited patients to complete the MHQ and the patient satisfaction questionnaire.

Other data recording: patient characteristics and complications

Additional variables that were routinely collected in the database included age, sex, type of work, symptom duration, treatment side, and dominance. We reviewed the medical records to collect data on possible exclusion criteria and the occurrence of complications.

As complications we scored the following events when they occurred within a year after surgery: infections requiring antibiotics, neuropathic pain, tendon rupture, tendinitis, stenosing tenosynovitis of the thumb (treated with splint/hand therapy, corticosteroid injection or surgery), de Quervain's tenosynovitis (treated with splint/hand therapy, corticosteroid injection or surgery) and reoperations.

Statistics

Paired t-tests were used to compare preoperative and postoperative continuous scores. Univariable comparisons of continuous values between groups were done with ANOVA tests; for categorical outcomes, Chi-Square tests were used. When for categorical data less than five observations per category were present, Fisher's exact tests were used. No significance level correction was applied. We expected a low number of complications and, therefore, mainly focused on the total percentage of complications, including reoperations, between the four groups. A non-responder analysis was performed to compare patient characteristics of patients completing all PROMs of interest (responders) and non-responders.

We used multivariable linear and logistic mixed model analyses to assess the association between surgical techniques and continuous and binary outcomes, while accounting for clustering per surgeon. Patient characteristics (i.e., age, gender, type of work, duration of symptoms, treated side), concomitant surgical procedure, and all baseline MHQ subscale scores were included in the multivariable mixed model analyses to account for potential differences at baseline. No variable selection was performed. From the linear mixed effects analysis, an estimate per surgical technique can be obtained, which can be interpreted as the difference in outcome score for patients undergoing that technique compared to patients undergoing the reference technique (Burton-Pellegrini) when all other variables are the same. Following the recommendation by Howard et al.[28], no correction for family-wise type I error rate due to comparing the Anchovy plasty, Zancolli, and Weilby techniques to the same reference group (Burton-Pellegrini), was applied. As measure of goodness of fit, we report the explained variance based on multilevel variance partitioning.[29] For logistic mixed models, odds ratios with 95% confidence intervals are reported, again comparing patients undergoing that technique to patients undergoing the reference technique when all other variables are the same. As measure of goodness of fit for the logistic mixed models, we report the area under the receiver operating curve (AUC). To determine whether this study was sufficiently powered, we performed a post-hoc power calculation for the mixed effects model with MHQ pain twelve months postoperatively as outcome. More details on this calculation are reported in Supplement B.

The variance inflation factor was calculated to assess the correlation between the variables in the mixed effects models. A value of 10 or higher indicates that the variables are

too strongly correlated and that the effect of individual predictors cannot be assessed accurately.[30]

All analyses were performed using R statistical computing, version 3.6.3. P-values for the mixed models were derived from the lmerTest package.[31] For all tests, a p-value smaller than 0.05 was considered statistically significant. We chose to report a primary outcome and multiple secondary outcomes. However, multiple testing correction was not performed; therefore, the results of the analyses on the secondary outcomes should be interpreted as merely exploratory.[32]

RESULTS

From November 2013 until December 2018, 1718 patients were treated with trapeziectomy and a form of tendon plasty. The loss of follow-up concerns 732 patients that did not complete all PROMS of interest, i.e. non-responders, having incomplete data at baseline and/or incomplete data at twelve months postoperatively (Figure 1). After applying the exclusion criteria, another 193 were excluded; therefore, we included 793 patients. We compared patient characteristics and surgical technique (Supplementary Table 1), and found a difference in surgical techniques used for responders and non-responders (p=0.040).

Most patients, 465 (59%), were treated with the Weilby technique. Patient characteristics and baseline MHQ scores per surgical technique are reported in Table 1 and 2, respectively. From the univariable comparisons we concluded that the four groups were not comparable at baseline, as we found differences in the MHQ work score (p=0.038) between the groups. This indicated that correction for baseline differences between the groups was required.

Primary outcome

For all groups, there was a significant improvement ranging from 28-34 points in MHQ pain score between baseline and twelve months postoperatively (p <0.001) (Figure 2). The MHQ pain score at baseline was not different between the groups. In the post-hoc power calculation for a mixed effects model with MHQ pain twelve months postoperatively as outcome, we aimed to detect a difference of -5 in MHQ pain score for patients undergoing a Weilby plasty and +5 for patients undergoing an Anchovy plasty. The simulations indicated that our study was sufficiently powered to detect this difference (power = 96.1%, 95% CI (94.7; 97.2)).

Table 3 demonstrates the results of the mixed effects analysis of MHQ pain score at twelve months. The variance inflation factor was below 10, which indicates that the variables are not too strongly correlated, and the effects of the individual predictors can be assessed. We found no significant difference in MHQ pain score at twelve months between the surgical techniques when accounting for baseline differences.

Secondary outcomes

MHQ scores

In the mixed effects analysis, we found no differences in MHQ hand function (Supplementary Table 2) and MHQ total score (Supplementary Table 3) after accounting for baseline differences at twelve months between surgical techniques. In the Burton-Pellegrini group, 71% of the patients achieved the MCID for MHQ pain, compared to 77% of the Zancolli group, 73% of the Weilby group, and 86% of the Anchovy plasty group (Figure 3). In logistic mixed effects modelling, we found that patients undergoing Anchovy plasty have significantly higher odds of achieving the MCID for MHQ pain after accounting for baseline differences, with an odds ratio of 2.3 (95% CI 1.2 - 4.8). The AUC of this model was 0.65.

Satisfaction with treatment outcomes

In Figure 4, satisfaction with treatment outcomes twelve months postoperatively per surgical technique is shown. 84% of all patients rated their satisfaction with treatment outcomes as fair, good or excellent twelve months postoperatively. In logistic mixed effects modelling, we found no significant difference in patients' satisfaction between surgical techniques after accounting for baseline differences regardless of whether patients scoring "fair" were classified as satisfied or not satisfied (AUC 0.62-0.68).

Complications

In Table 4, the number of complications and reoperations within a year, as well as univariable comparisons, are reported for each surgical technique. In total, 141 patients (18%) had a complication of which 16 had a reoperation (2% of total population). The majority of reoperations consisted of (partial) resection of the scaphoid, trapezoid, or both (N=5) and release of the tendon plasty due to tendinitis or adhesions (N=7). In the Anchovy plasty group, 7% of the patients had complications, compared to 22% in the Weilby group. In univariable analysis, we found differences between the surgical techniques in occurrence of tendinitis and stenosing tenosynovitis of the thumb. However, in multivariable analysis where we accounted for baseline differences, we found no significant differences in the proportion of patients having complications between surgical techniques. This model has an AUC of 0.72.

DISCUSSION

The main aim of this observational study was to evaluate if one of the tendon plasty techniques is superior in terms of patient reported outcomes. We compared the postoperative MHQ pain scores of four different techniques of tendon plasty following trapeziectomy,

performed for patients with primary TMC OA. All four techniques resulted in a considerable MHQ pain reduction at twelve months, but no technique was superior.

A notable difference was that patients treated with Anchovy plasty were more likely to experience a clinically relevant improvement in pain compared to the other three techniques: Burton-Pellegrini, Weilby and Zancolli.

We believe a comparison between multiple types of tendon plasty after trapeziectomy is relevant, as previous studies have shown that trapeziectomy with a tendon plasty is still preferred over simple trapeziectomy by many surgeons.[1-3,15] Till now only two tendon plasty were compared. Several studies have compared different techniques of arthroplasties for the treatment of TMC OA. Although different outcome parameters are used which makes comparison difficult, almost all studies analyzed the change in pain pre- and postoperatively (VAS, DASH, or an alternative pain score). Until now, no technique has been proven to be superior in pain reduction. For instance, Esenwein et al. reported no differences in pain outcome at eight months between the Burton-Pellegrini procedure and a trapeziectomy with an APL sling.[10] Vermeulen et al. found no difference in pain scores twelve months postoperatively, comparing Burton-Pellegrini to the Weilby procedure.[13] In our study, we have used the MHQ pain instead of the DASH or VAS pain score, but in correspondence with the former studies, a significant reduction in pain was seen after surgery and the change in pain was similar for all four techniques.

The MCID for MHQ pain of arthroplasties of the TMC joint remains underexposed. Therefore, the comparison of MCID of the four different techniques in this study is a valuable addition to the current available evidence. Unfortunately, the MCID threshold has not been calculated yet specifically for TMC OA. We therefore used the MCID for MHQ pain that was determined for patients undergoing arthroplasty for rheumatoid arthritis (value 11).[26] Alternatively, the MCID of MHQ pain for atraumatic unilateral single hand/forearm diagnoses (value 14)[33] could have been chosen. This does not change our finding that Anchovy plasty patients have a higher chance of experiencing clinically relevant improvement in pain compared to the other three tendon plasties.

Other outcome measurements of interest of the present study were hand function twelve months postoperatively, postoperative complications and patients' satisfaction with the treatment outcome. No differences were found between the four tendon plasties for these outcomes. As for the complications, mainly tendinitis was seen after the Weilby technique. This is similar to the study of Vermeulen et al., that showed no difference in complication rate between the Weilby and Burton Pellegrini technique.[13]

The strength of this study is the collection of data, which occurred prospectively in real-life conditions. Surgery was performed by surgeons from twenty-seven different clinics. This diversity, combined with the size of our cohort, adds to the strength of the comparison.

One could argue there the lack of randomization between the treatment options induces selections bias. However, all surgeons mainly used one technique for all their patients, and we therefore advocate that treatment was mostly based on surgeon preference instead of selection bias (Supplementary Figure 1). As we corrected for potential differences in patient characteristics and baseline patient reported outcomes of the patients in all treatment groups, the non-randomized comparison seems valid.

A second limitation is the relatively high number of patients lost to follow-up (45%), which is inherent to the registration system used. In our practices patients voluntarily fill in questionnaires on set times during their treatment follow up and reminders are sent twice. Analysis of the patient characteristics of the non-responders showed that the only difference with the responders was in surgical technique used (Supplementary Table 1), but since this was only a small difference, we do not believe that this has an impact on our findings.

A final limitation is that this study did not include patients that were treated with isolated trapeziectomy. This would be a valuable addition to this observational study, however the number of patients treated with simple trapeziectomy at our clinics was extremely low.

In correspondence with former studies, we found that the majority of the patients treated in our clinic experience a clinical important improvement after trapeziectomy combined with a form of tendon plasty. However, there is still room for improvement. Dependent on the technique used, 14-29% of patients did not reach the MCID (Figure 3). Besides surgical technique, patient characteristics or illness perception might be associated with negative outcomes.[34,35] Further analysis into which factors are predictive of less successful results should be performed in the future.

CONCLUSION

We found that MHQ pain scores twelve months postoperatively were similar between the four tendon plasty techniques following trapeziectomy as treatment for TMC OA.

The Weilby; Burton-Pellegrini; Zancolli and the Anchovy plasty all showed significant pain reduction over time. Satisfaction rates (over 80%) and complication rates were not different between the groups. Overall, more than 70% of patients experienced a clinically relevant improvement in pain. A notable distinction was that patients who received an Anchovy plasty following trapeziectomy, had the highest chance to reach this threshold. Therefore, when using a tendon plasty, we would advise to use the Anchovy plasty, being the least invasive and fastest technique with the highest chance of obtaining clinical improvement.

ACKNOWLEGDMENTS

The authors thank all patients who participated and allowed their data to be anonymously used for the present study.

GROUP COLLABORATORS

The collaborators of the Hand-Wrist Study Group are: RAM Blomme, BJR Sluijter, DJJC van der Avoort, A Kroeze, J Smit, J Debeij, ET Walbeehm, GM van Couwelaar, GM Vermeulen, JP de Schipper, JFM Temming, JH van Uchelen, HL de Boer, KP de Haas, K Harmsen, OT Zöphel, R Feitz, JS Souer, R Koch, TM Moojen, X Smit, R van Huis, PY Pennehouat, K Schoneveld, YE van Kooij, RM Wouters, P Zagt, FJ van Ewijk, JJ Veltkamp, A Fink, WA de Ridder, HP Slijper, JT Porsius, J Tsehaie, R Poelstra, MC Jansen, MJW van der Oest, PO Sun, JS Teunissen, J Dekker, ML Jansen-Landheer, MHP ter Stege.

REFERENCES

1. Brunton LM, Wilgis EF. A survey to determine current practice patterns in the surgical treatment of advanced thumb carpometacarpal osteoarthrosis. *Hand (N Y).* 2010;5(4):415-422.
2. Deutch Z, Niedermeier SR, Awan HM. Surgeon Preference, Influence, and Treatment of Thumb Carpometacarpal Arthritis. *Hand (N Y).* 2018;13(4):403-411.
3. Ottenhoff JSE, Teunis T, Janssen SJ, Mink van der Molen AB, Ring D. Variation in Offer of Operative Treatment to Patients With Trapeziometacarpal Osteoarthritis. *Journal of Hand Surgery.* 2020;45(2):123-130.e121.
4. Wolf JM, Delaronde S. Current trends in nonoperative and operative treatment of trapeziometacarpal osteoarthritis: a survey of US hand surgeons. *J Hand Surg Am.* 2012;37(1):77-82.
5. Vermeulen GM, Slijper H, Feitz R, Hovius SE, Moojen TM, Selles RW. Surgical management of primary thumb carpometacarpal osteoarthritis: a systematic review. *J Hand Surg Am.* 2011;36(1):157-169.
6. Wajon A, Vinycomb T, Carr E, Edmunds I, Ada L. Surgery for thumb (trapeziometacarpal joint) osteoarthritis. *Cochrane Database Syst Rev.* 2015(2):CD004631.
7. Avant KR, Nydick JA, White BD, Vaccaro L, Hess AV, Stone JD. Basal joint osteoarthritis of the thumb: comparison of suture button versus abductor pollicis longus suspensionplasty. *Hand (N Y).* 2015;10(1):80-84.
8. Belcher HJ, Nicholl JE. A comparison of trapeziectomy with and without ligament reconstruction and tendon interposition. *J Hand Surg Br.* 2000;25(4):350-356.
9. Catalano L, Horne LT, Fischer E, Barron OA, Glickel SZ. Comparison of ligament reconstruction tendon interposition and trapeziometacarpal interposition arthroplasty for basal joint arthritis. *Orthopedics.* 2008;31(3):228.
10. Esenwein P, Hoigne D, Zdravkovic V, Sanchez T. [Resection, interposition and suspension arthroplasty for treatment of Basal joint arthritis of the thumb: a randomized and prospective comparison of techniques using the abductor pollicis longus- and the flexor carpi radialis tendon]. *Handchir Mikrochir Plast Chir.* 2011;43(5):289-294.
11. Gangopadhyay S, McKenna H, Burke FD, Davis TR. Five- to 18-year follow-up for treatment of trapeziometacarpal osteoarthritis: a prospective comparison of excision, tendon interposition, and ligament reconstruction and tendon interposition. *J Hand Surg Am.* 2012;37(3):411-417.
12. Park MJ, Lichtman G, Christian JB, et al. Surgical treatment of thumb carpometacarpal joint arthritis: a single institution experience from 1995-2005. *Hand (N Y).* 2008;3(4):304-310.
13. Vermeulen GM, Spekreijse KR, Slijper H, Feitz R, Hovius SE, Selles RW. Comparison of arthroplasties with or without bone tunnel creation for thumb basal joint arthritis: a randomized controlled trial. *J Hand Surg Am.* 2014;39(9):1692-1698.
14. Reissner L, Marks M, Schindele S, Herren DB. Comparison of clinical outcome with radiological findings after trapeziectomy with ligament reconstruction and tendon interposition. *J Hand Surg Eur Vol.* 2016;41(3):335-339.
15. Wilcke M, Roginski M, Astrom M, Arner M. A registry based analysis of the patient reported outcome after surgery for trapeziometacarpal joint osteoarthritis. *BMC Musculoskelet Disord.* 2020;21(1):63.
16. Sigfusson R, Lundborg G. Abductor pollicis longus tendon arthroplasty for treatment of arthrosis in the first carpometacarpal joint. *Scand J Plast Reconstr Surg Hand Surg.* 1991;25(1):73-77.

17. Weilby A. Tendon interposition arthroplasty of the first carpo-metacarpal joint. *J Hand Surg Br.* 1988;13(4):421-425.

18. Burton RI, Pellegrini VD, Jr. Surgical management of basal joint arthritis of the thumb. Part II. Ligament reconstruction with tendon interposition arthroplasty. *J Hand Surg Am.* 1986;11(3):324-332.

19. Froimson AI. Tendon arthroplasty of the trapeziometacarpal joint. *Clin Orthop Relat Res.* 1970;70:191-199.

20. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet.* 2007;370(9596):1453-1457.

21. Selles RW, Wouters RM, Poelstra R, et al. Routine health outcome measurement: Development, design, and implementation of the Hand and Wrist Cohort. *Plast Reconstr Surg.* 2020.

22. Tsehaie J, Wouters RM, Feitz R, Slijper HP, Hovius SER, Selles RW. Shorter vs Longer Immobilization After Surgery for Thumb Carpometacarpal Osteoarthritis: A Propensity Score-Matched Study. *Arch Phys Med Rehabil.* 2019;100(11):2022-2031.e2021.

23. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am.* 1998;23(4):575-587.

24. Marks M, Audige L, Herren DB, Schindele S, Nelissen RG, Vliet Vlieland TP. Measurement properties of the German Michigan Hand Outcomes Questionnaire in patients with trapeziometacarpal osteoarthritis. *Arthritis Care Res (Hoboken).* 2014;66(2):245-252.

25. Frouzakis R, Herren DB, Marks M. Evaluation of expectations and expectation fulfillment in patients treated for trapeziometacarpal osteoarthritis. *J Hand Surg Am.* 2015;40(3):483-490.

26. Shauver MJ, Chung KC. The minimal clinically important difference of the Michigan hand outcomes questionnaire. *J Hand Surg Am.* 2009;34(3):509-514.

27. De Ridder WA, van Kooij YE, Vermeulen GM, et al. Test-retest Reliability and Construct Validity of the Satisfaction with Treatment Result Questionnaire in Patients with Hand and Wrist Conditions: A Prospective Study. *Clin Orthop Relat Res.* 2021.

28. Howard DR, Brown JM, Todd S, Gregory WM. Recommendations on multiple testing adjustment in multi-arm trials with a shared control group. *Stat Methods Med Res.* 2018;27(5):1513-1530.

29. LaHuis DM, Hartman MJ, Hakoyama S, Clark PC. Explained Variance Measures for Multilevel Models. *Organizational Research Methods.* 2014;17(4):433-451.

30. Applied linear statistical models. 2005.

31. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. *2017.* 2017;82(13):26.

32. Li G, Taljaard M, Van den Heuvel ER, et al. An introduction to multiplicity issues in clinical trials: the what, why, when and how. *Int J Epidemiol.* 2017;46(2):746-755.

33. London DA, Stepan JG, Calfee RP. Determining the Michigan Hand Outcomes Questionnaire minimal clinically important difference by means of three methods. *Plast Reconstr Surg.* 2014;133(3):616-625.

34. Hoogendam L, van der Oest MJW, Tsehaie J, et al. Psychological factors are more strongly associated with pain than radiographic severity in non-invasively treated first carpometacarpal osteoarthritis. *Disabil Rehabil.* 2019:1-6.

35. van der Oest MJW, Poelstra R, Feitz R, et al. Illness Perceptions of Patients With First Carpometacarpal Osteoarthritis, Carpal Tunnel Syndrome, Dupuytren Contracture, or Trigger Finger. *J Hand Surg Am.* 2020;45(5):455.e451-455.e458.

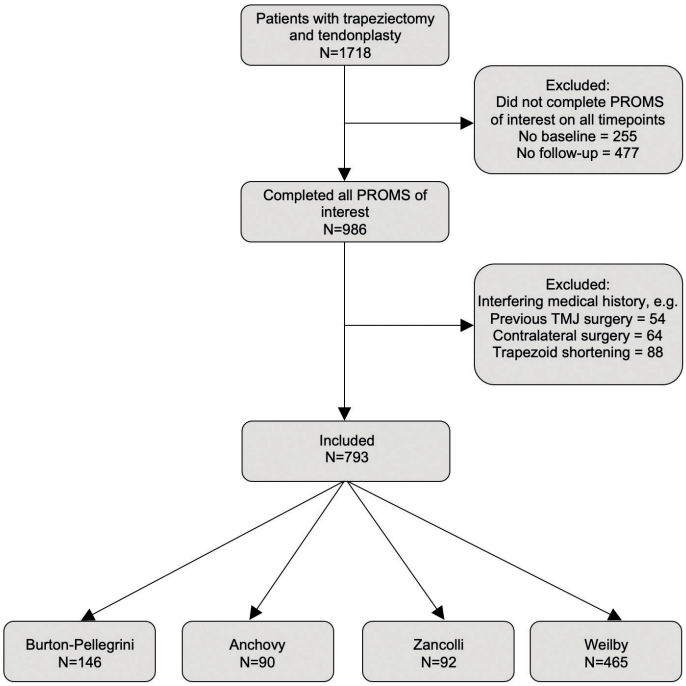


Figure 1. Flowchart of patient inclusion.

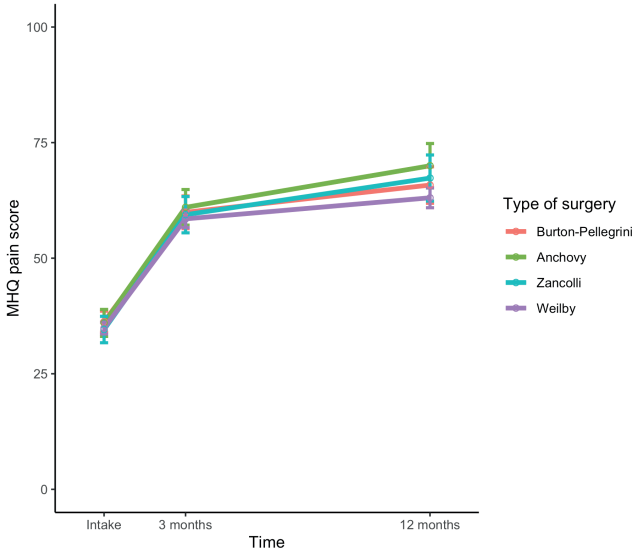


Figure 2. Mean MHQ pain scores in the first year after surgery per surgical technique. The error bars represent the 95% confidence interval.

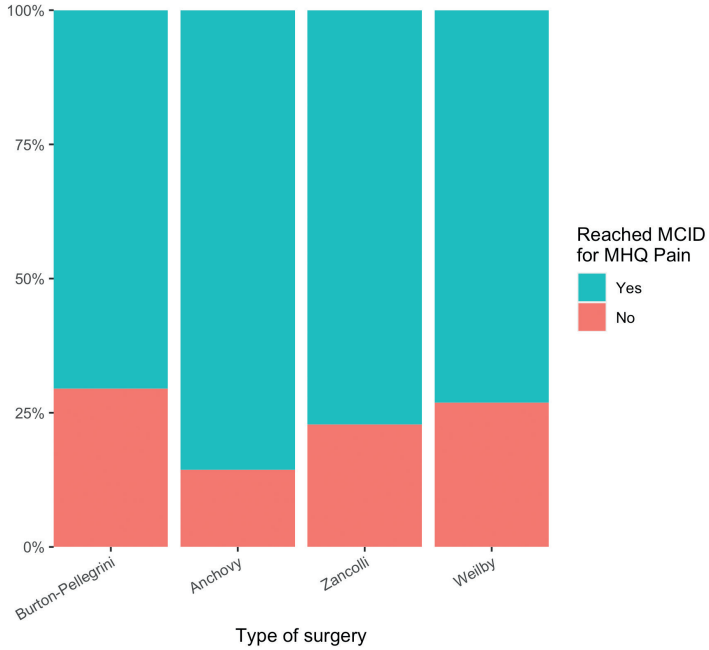


Figure 3. Bar chart of patients reaching the MCID for MHQ pain twelve months postoperatively per surgical technique.

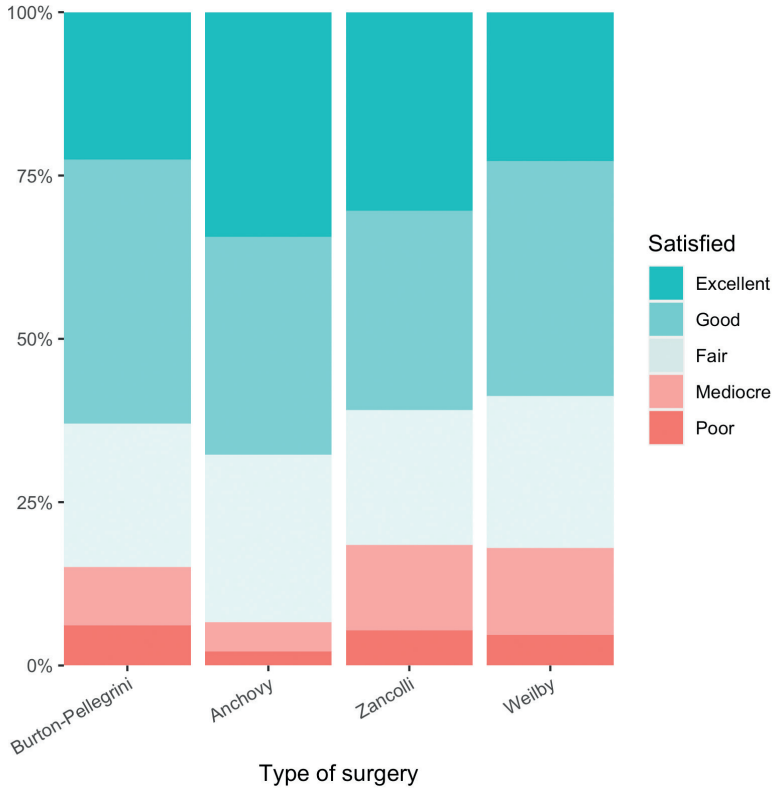


Figure 4. Bar chart of patients' satisfaction with treatment outcomes twelve months postoperatively per surgical technique.

Table 1. Patient characteristics at baseline per treatment group

	Burton-Pellegrini (N=146)	Anchovy (N=90)	Zancolli (N=92)	Weilby (N=465)	P-value
Age (years)	60 (8)	60 (7)	62 (8)	61 (8)	0.092
Gender, female (N, %)	111 (76)	68 (76)	75 (82)	360 (77)	0.747
Type of work (N, %)	_	_	_	_	0.262
Unemployed	56 (38)	40 (44)	47 (51)	217 (47)	_
Light physical labor	39 (27)	12 (13)	17 (18)	84 (18)	_
Moderate physical labor	35 (24)	24 (27)	20 (22)	107 (23)	_
Heavy physical labor	16 (11)	14 (16)	8 (9)	57 (12)	_
Duration of symptoms (months), median (IQR)	24 (12-36)	24 (12-46.5)	24 (12-36)	24 (12-48)	0.165
Treated side, right (N, %)	63 (43)	39 (43)	47 (51)	225 (48)	0.509
Concomitant surgical procedure (N, %)	28 (19)	11 (12)	15 (16)	53 (11)	0.088

Table 2. MHQ scores at baseline per treatment group

	Burton-Pellegrini (N=146)	Anchovy (N=90)	Zancolli (N=92)	Weilby (N=465)	P-value
MHQ-Pain	36 (14)	36 (14)	35 (14)	35 (14)	0.670
MHQ-Hand function	50 (16)	51 (17)	49 (18)	49 (18)	0.832
MHQ-ADL	56 (20)	53 (23)	52 (21)	54 (21)	0.572
MHQ-Work	53 (25)	49 (26)	44 (26)	49 (25)	0.038*
MHQ-Aesthetics	79 (20)	77 (22)	74 (22)	79 (20)	0.106
MHQ-Satisfaction	33 (18)	30 (19)	29 (18)	31 (18)	0.354
Total MHQ score	51 (14)	49 (14)	47 (14)	50 (14)	0.134

Mean (SD) unless stated otherwise
* $p \leq 0.05$

Table 3. Results from linear mixed model analysis with MHQ pain score twelve months postoperatively as outcome. The beta coefficients (B), p-values and 95% confidence intervals (95% CI) are reported

Explanatory variables	B	P-value	95% CI
Surgical technique (Ref = Burton-Pellegrini)	_	_	_
Anchovy plasty	4.98	0.163	-2.0; 12.0
Zancolli	3.95	0.229	-2.5; 10.4
Weilby	-1.18	0.644	-6.2; 3.8
Age (years)	0.26	0.045*	0.01; 0.5
Gender, female	-2.59	0.198	-6.5; 1.4
Type of work (Ref = unemployed)	_	_	_
Light physical labor	0.61	0.801	-4.1; 5.3
Moderate physical labor	5.64	0.013*	1.2; 10.1
Heavy physical labor	1.90	0.510	-3.8; 7.6
Duration of symptoms (months)	0.02	0.251	-0.02; 0.06
Affected side, right hand	1.08	0.504	-2.1; 4.2
Concomitant surgical procedure, no	6.52	0.006**	1.9; 11.1
MHQ-Pain	0.44	<0.001***	0.3; 0.6
MHQ-Hand function	-0.01	0.914	-0.1; 0.1
MHQ-ADL	-0.06	0.242	-0.2; 0.04
MHQ-Work	0.11	0.007**	0.03; 0.2
MHQ-Aesthetics	0.03	0.549	-0.1; 0.1
MHQ-Satisfaction	0.06	0.370	-0.1; 0.2
Explained variance R2		0.16	
(multilevel variance partitioning)			

* $p \leq 0.05$
** $p \leq 0.01$
*** $p \leq 0.001$

Table 4. Complications and reoperations within the first year after surgery per surgical technique

	Burton-Pellegrini (N=146)	Anchovy (N=90)	Zancolli (N=92)	Weilby (N=465)	P-value
Infections	2 (1)	0 (0)	0 (0)	8 (2)	0.533
Neuropathic pain	0 (0)	3 (3)	4 (4)	10 (2)	0.054
Tendon rupture	1 (1)	0 (0)	1 (1)	4 (1)	0.924
Tendinitis	4 (3)	0 (0)	2 (2)	31 (7)	0.007
Stenosing tenosynovitis of the thumb	7 (5)	1 (1)	5 (5)	39 (8)	0.040
De Quervain's tenosynovitis	8 (5)	2 (2)	4 (4)	35 (8)	0.239
Reoperation	3 (2)	0 (0)	0 (0)	13 (3)	0.199
Total number of patients with at least one complication	21 (14)	6 (7)	14 (15)	100 (22)	0.004

N (%) unless stated otherwise

Part III

PROGNOSTIC FACTORS FOR CLINICALLY IMPORTANT TREATMENT OUTCOMES (FOR THUMB BASE OSTEOARTHRITIS)

9

PSYCHOLOGICAL FACTORS ARE MORE STRONGLY ASSOCIATED WITH PAIN THAN RADIOGRAPHIC SEVERITY IN NON-INVASIVELY TREATED FIRST CARPOMETACARPAL OSTEOARTHRITIS

Hoogendam, L., van der Oest, M. J. W., Tsehaie, J., Wouters, R. M., Vermeulen, G. M., Slijper, H. P., Selles, R. W., Porsius, J. T., The Hand-Wrist Study Group (2021). *Disability and rehabilitation*, *43*(13), 1897–1902.

ABSTRACT

Background

The aim of this study was to investigate to what extent psychological factors are related to pain levels prior to non-invasive treatment in patients with osteoarthritis of the first carpometacarpal joint.

Methods

We included patients (n=255) at the start of non-invasive treatment for osteoarthritis of the first carpometacarpal joint who completed the Michigan Hand Outcome Questionnaire. Psychological distress, pain catastrophizing behavior and illness perception was measured. X-rays were scored on presence of scaphotrapeziotrapezoid osteoarthritis. We used hierarchical linear regression analysis to determine to what extent pain levels could be explained by patient characteristics, X-ray scores and psychological factors.

Results

Patient characteristics and X-ray scores accounted for only 6% of the variation in pre-treatment pain levels. After adding the psychological factors to our model, 47% of the variance could be explained.

Conclusions

Our results show that psychological factors are more strongly related to pain levels prior to non-invasive treatment in patients with osteoarthritis of the first carpometacarpal joint than patient characteristics and X-ray scores, which implies the important role of these factors in the reporting of symptoms. More research is needed to determine whether psychological factors will also affect treatment outcomes for patients treated non-invasively for osteoarthritis of the first carpometacarpal joint.

INTRODUCTION

Osteoarthritis of the first carpometacarpal joint (CMC-1 OA) is a degenerative disease that causes pain and loss of function [1]. Patients are initially treated non-invasively [2] with hand therapy, occupational therapy, an orthosis, or a combination of treatment modalities [3]. Non-invasive treatment is an effective treatment that may prevent the need for surgical treatment [4] and reduces pain in a selection of patients [5].

At the start of the treatment, considerable variation in pain levels between patients is seen [5]. However, traditional patient and disease attributes, e.g. age, grip strength and X-ray scores, only explain a small amount of the variation in reported pain and disability, which suggests other factors are at play [6,7]. It is currently unclear which factors are associated with pain for CMC-1 OA patients.

Several studies on surgical treatment of OA, including total knee or hip replacement [8-11] and surgery for CMC-1 OA [12,13] found that psychological factors (e.g. depression, pain catastrophizing behavior and illness perception) are associated with worse patient reported outcomes, both before and after treatment. Moreover, recent studies suggested that interventions improving catastrophizing behavior [14] and negative illness perception [15] have a beneficial effect on OA symptoms.

Although there is evidence for the association between psychological factors and symptom severity in knee and hip OA , little is known regarding this association for patients treated non-invasively for CMC-1 OA [16,17]. In particular, the association between psychological factors and pain, which is the primary complaint of CMC-1 OA patients [18], is currently unknown. Moreover, while illness perceptions have been shown to be important factors in other conditions, like knee OA, no studies have investigated to what extent illness perceptions are associated with pain in this patient population [11]. Therefore, the aim of this study is to investigate to what extent psychological distress, pain catastrophizing behavior and illness perceptions are associated with pain levels prior to non-invasive treatment in CMC-1 OA patients.

METHODS

Setting and study population

This cross-sectional study was performed at Xpert Clinic in The Netherlands. Xpert Clinic is a specialized private treatment center for hand and wrist conditions. Xpert Clinic has 20 different locations, with 20 European Board certified (Federation of European Societies for Surgery of the Hand) hand surgeons and over 150 hand therapists.

All patients who received non-invasive treatment, consisting of orthosis and/or hand therapy, for CMC-1 OA at Xpert Clinic between September 2017 and July 2018 were invited to complete several questionnaires as part of routine clinical care to measure symptom severity, psychological status, understanding of disease and quality of life prior to treatment. These questionnaires were e-mailed after the first consultation and before non-invasive treatment started. Three reminders were e-mailed to non-responders. Furthermore, baseline demographics, including age, sex, hand dominance and occupational intensity were collected. Occupational intensity was classified by the hand therapist in one of the following categories: not employed, light occupational intensity (e.g. working in an office), moderate occupational intensity (e.g. working in a shop) or severe occupational intensity (e.g. construction work). All patients provided written informed consent.

Michigan Hand Outcomes Questionnaire

The Michigan Hand Outcomes Questionnaire (MHQ) [19] is a patient reported outcome measure with six domains (pain, aesthetics, hand function, performance of activities of daily living, work performance and satisfaction) with good validity, reliability and responsiveness in CMC-1 OA patients [20]. Scores range from 0-100 (0 = poorest function, 100 = ideal function). In the present study the pain scores were reversed (0 = no pain, 100 = extreme pain).

The MHQ pain subscale was used as primary outcome measure, because pain is the primary complaint and the main reason to seek treatment for CMC-1 OA patients [18]. To also see if there is an association between psychological effects and patient reported hand performance, we used the total MHQ score as secondary outcome measure.

Patient Health Questionnaire-4

The Patient Health Questionnaire-4 (PHQ) [21] is a generic depression- and anxiety-screening tool and was used to measure psychological distress. This questionnaire is a combination of two brief screening tools (Patient Health Questionnaire-2 and Generalized Anxiety Disorder-2). The PHQ contains two domains (anxiety and depression) with two questions each. The total score ranges from 0-12 (0 = no indication for psychological distress; 12 = strong indication for psychological distress). It has a good reliability and validity in the general population [22].

Pain Catastrophizing Scale

The Pain Catastrophizing Scale (PCS) [23] was used to assess catastrophizing behavior in response to pain. This questionnaire consists of 13 questions and 3 domains (helplessness, magnification and rumination). It has been demonstrated to have good validity, reliability

and responsiveness in patients with pain related problems [24,25]. The total score ranges from 0-52 (0 = no catastrophizing behavior; 52 = severe catastrophizing behavior).

Brief Illness Perception Questionnaire

The Brief Illness Perception Questionnaires (B-IPQ) [26] was used to assess the patients' perceptions of illness. This questionnaire is a short version of the Revised Illness perception Questionnaire [27]. In the B-IPQ each subscale of the Revised Illness perception Questionnaire is summarized by one question. Five questions assess cognitive illness representation, two questions assess emotional representations, one question assesses understanding of disease and in the final question patients are asked to list the factors they believe to have caused their illness. This last question was not part of our questionnaire. Reliability and validity has been demonstrated for multiple conditions, including low back pain, cardiovascular disease and chronic obstructive pulmonary disease [28-31].

CMC-1 joint X-rays

The patients' records were searched for X-rays of the first carpometacarpal joint. If multiple X-rays were present, we selected the X-ray in which both the CMC-1 joint and the scaphotrapeziotrapezoid joint (STT) were most clearly visible. The Eaton-Glickel classification [32] ranges from stage I to stage IV. Stage III is defined as excessive CMC-1 degeneration and subluxation. Stage IV is defined as stage III with additional presence of STT OA. According to this classification, presence of STT OA indicates the most advanced stage of structural damage. Therefore, we used this feature as indication of radiographic severity of disease. The first 100 X-rays were independently scored by both a European Board-certified hand surgeon (G.V.) and a junior scientist (L.H.). The Intraclass Correlation Coefficient was 0.58 (95% CI 0.49-0.65). This is in agreement with the study by Dela Rosa et al. [33], who reported fair to moderate inter-observer agreement for the Eaton-Glickel classification, with similar agreement rates for stage I, III and IV. The scores of the junior scientist were used for all patients. Patients without an X-ray of the CMC-1 joint were excluded.

Statistical analyses

A complete case analysis was performed with patients who completed all previously mentioned questionnaires. To see whether patients with missing data differed from patients with complete data, two non-responder analyses were performed; both for patients who completed the MHQ, but did not complete the psychological questionnaires and for patients who completed all questionnaires, but without X-ray of the CMC-1 joint. For these analyses, T-tests were used for normally distributed continuous data and Mann-Whitney-Wilcoxon tests were used for continuous data that was not normally distributed. Chi square statistics

were used for categorical data. We calculated Pearson correlation coefficients to determine to what extent the psychological variables were correlated.

Data were analyzed using hierarchical linear regression analyses with baseline MHQ-pain levels as a dependent variable. In the first step of the analysis, age, sex and occupational intensity were entered into the model. Presence of STT OA was added in the second step of analysis. In the third step, we entered the total PHQ score, as well as the total PCS score. In the fourth step, all B-IPQ subscales were added in order to determine the effect of illness perceptions on pain, after correcting for psychological distress and pain catastrophizing behavior.

For all variables, the regression coefficients (B) are reported, which represents the increase in the dependent variable for one unit increase in the independent variable, when all other variables remain constant. In order to compare the relative contribution of each explanatory variable on the outcome, the standardized regression coefficients (β) are also reported. Standardized regression coefficients allow for comparison of the strength of associations when the independent variables were measured on different scales. For all models both the multiple explained variance (R^2) and the explained variance adjusted for number of variables in the model (adjusted R^2) is calculated.

All analyses were performed using R statistical computing, version 3.4.1. For all tests a p-value smaller than 0.05 was considered statistically significant.

RESULTS

Non-responder analysis

We identified 475 patients at the start of non-invasive treatment for CMC-1 OA who had completed the MHQ. 5.2% of these patients did not complete all psychological questionnaires. Of the patients who completed all questionnaires, 40.4% did not have an X-ray of the CMC-1 joint. Supplemental tables 1-2 show demographic characteristics and MHQ scores for responders and non-responders, indicating no significant differences in any patient characteristic or MHQ-pain scores between responders and non-responders.

Patient characteristics

255 patients were included in the analysis (figure 1). Their mean age was 60 ± 8 years (mean ± SD) and 75% of the patients were female. The mean MHQ-pain score was 52.9 ± 17.3 and the mean total MHQ score was 59.7 ± 15.2. Table 1 shows baseline characteristics for all included patients. Correlations between the psychological factors ranged from -0.24 to 0.60 (supplemental table 3), which can be interpreted as weak to moderate correlations [34].

Hierarchical linear regression

Table 2 shows the outcomes of the hierarchical regression analysis. In model 1 and 2, female sex was statistically significantly associated with higher MHQ-pain scores. However, after adding psychological distress and catastrophizing behavior to the model, sex was no longer statistically significantly related to pain, while higher PHQ and PCS scores were statistically significantly associated with higher MHQ-pain scores. After adding the B-IPQ subscales to the model, B-IPQ subscales 'consequences' and 'identity' were, in addition to PHQ score and PCS score, also statistically significantly associated with pain. Figure 2 shows the increase in explained variance per model. Model 1 and 2 had an explained variance of 5% and 6% respectively. After adding psychological distress and catastrophizing behavior, the explained variance increased to 35%, and after adding illness perceptions, the explained variance increased to 47%. In this model, more psychological distress (PHQ score, B = 0.79), more pain catastrophizing behavior (PCS score, B = 0.46), experiencing more consequences (B-IPQ 'consequences', B = 1.36) and more symptoms (B-IPQ 'identity', B = 1.11) were statistically significantly associated with higher MHQ-pain scores. Of all significant variables in the final model, total PCS score had the largest standardized regression coefficient (β = 0.27) in this model, indicating that pain catastrophizing behavior has the largest independent effect on pre-treatment pain of all variables investigated in this study.

We found similar associations in the hierarchical linear regression analysis on total MHQ score. Further details are reported in supplemental table 4.

DISCUSSION

The aim of this study was to investigate to what extent psychological factors are related to pre-treatment pain levels in patients receiving non-invasive treatment for CMC-1 OA. After controlling for patient characteristics and radiographic severity of OA, we found that higher psychological distress, more pain catastrophizing behavior and experiencing more consequences and symptoms from the illness were independently associated with higher pre-treatment pain levels. Pain catastrophizing behavior has the strongest association with pre-treatment pain. Patient characteristics and radiographic severity of the disease had no additive predictive value for pre-treatment pain.

Several previous studies focused on the association between X-ray scores and pain in patients with CMC-1 OA. However, different methods are available to score X-rays [35] and the available literature is conflicting [6,7,36-39]. Several radiographic OA features including erosions and sclerosis have been linked to pain levels [36,39], while Dahaghin et al. [7] in a large population study (n=3906) reported that radiographic OA was poorly correlated with pain. In our study, we found that presence of STT OA could only explain a

limited amount of variance in pain scores. Possibly, other radiographic OA features have a stronger association with pain and would therefore be more informative. However, as presence of STT OA is a clear indication of advanced CMC-1 OA in the Eaton-Glickel classification [32], we expected a stronger association between presence of STT OA and pain in CMC-1 OA patients. While X-rays may still have an important role in clinical practice, our study indicates that their value for understanding pain scores is limited.

To our knowledge this is the first study that assessed pre-treatment pain in CMC-1 OA patients that included both radiographic severity and psychological factors in the analysis. Murphy et al. [40] performed a similar study for women with knee OA. They found that fatigue, sleep quality and depression explained additional variance in pain severity after correcting for age and X-ray scores. This is in line with our findings that psychological factors explained additional variance in pre-treatment pain. However, in our study we found that psychological factors explained an additional variance of 40% compared to 10% in Murphy et al., which may be explained by use of different definitions of psychological variables in both studies.

Becker et al. [17] reported that symptom severity could largely be explained by whether or not the patients sought treatment for his symptoms and by pain catastrophizing behavior. This is in agreement with our finding that pain catastrophizing behavior has the strongest association with pre-treatment pain of all variables included in our study.

While no studies reported the association between pain and illness perceptions for patients with CMC-1 OA, Hill et al. [41] found that higher pain levels were associated with reporting more frustration, experiencing more consequences and expecting a chronic timeline in people with musculoskeletal hand problems.

The strength of this study is the large population where we combined psychological distress, pain catastrophizing as well as illness perceptions in explaining pre-treatment pain in non-invasively treated CMC-1 OA. Moreover, we included presence of STT OA in our analyses as measure for structural damage to the CMC-1 joint. To our knowledge this is the first study that combined psychological factors and radiographic severity of disease to explain pre-treatment pain levels of CMC-1 OA patients.

Study limitations

A limitation of our study is the quality of the X-rays that we used to score presence of STT OA. These X-rays were taken as part of daily clinical practice and therefore not taken in a standardized way, making it difficult to score all radiographic OA features. For that reason we only scored presence of STT OA, because presence of STT OA is an indication that radiologically the disease has reached an advanced stage [32]. Still our study clearly demonstrates that presence of STT OA, on X-rays taken as part of routine care, is not

related to pre-treatment pain, whereas psychological factors show a strong association with pre-treatment pain.

In addition, it is not possible to draw any causal implications from our research findings due to the cross-sectional design of our study. While our study demonstrates a strong association between pre-treatment pain levels and psychological factors, more research is needed to determine the direction of this association.

Future research

Based on the strong association between pre-treatment pain and treatment outcomes in our study and in previous studies [17, 23, 31], the question arises whether psychological factors will affect treatment outcomes and consequently, whether treatment results will improve when patients receive psychological support in addition to usual care. A large prospective study may provide valuable knowledge of the longitudinal association between psychological factors and pain and can be used to answer the research questions of interest.

Moreover, future studies have to demonstrate what type of psychological intervention would improve pain levels most in non-invasively treated CMC-1 OA patients, while also being feasible to offer in addition to usual care. This study suggests that pain catastrophizing behavior is the most important factor to target with a psychological intervention, but psychological distress and/or negative illness perceptions may be relevant targets for intervention as well.

In conclusion, the present study demonstrates that patient characteristics and X-rays have limited value for understanding pain in patients with CMC-1 OA. In contrast, we found a strong association between psychological factors and pain levels prior to non-invasive treatment. Clinicians should be aware of the strong relation between pain and psychological factors and should look beyond X-ray scores to judge pain intensity in patients with CMC-1 OA.

ACKNOWLEDGMENTS

The authors thank all patients who participated and allowed their data to be anonymously used for the present study.

GROUP AUTHORS

The members of the Hand-Wrist Study Group are: Arjen Blomme, Berbel Sluijter, Corinne Schouten, Dirk- Jan van der Avoort, Erik Walbeehm, Gijs van Couwelaar, Guus Vermeulen, Hans de Schipper, Hans Temming, Jeroen van Uchelen, Luitzen de Boer, Nicoline de Haas,

Oliver Zöphel, Reinier Feitz, Sebastiaan Souer, Steven Hovius, Thybout Moojen, Xander Smit, Rob van Huis, Pierre-Yves Pennehouat, Karin Schoneveld, Yara van Kooij, Robbert Wouters, Paul Zagt, Folkert van Ewijk, Frederik Moussault, Rik van Houwelingen, Joris Veltkamp, Arenda te Velde, Alexandra Fink, Harm Slijper, Ruud Selles, Jarry Porsius, Kim Spekreijse, Chao Zhou, Jonathan Tsehaie, Ralph Poelstra, Miguel Janssen, Mark van der Oest, Stefanie Evers, Jak Dekker, Matijs de Jong, Jasper van Gestel, Marloes ter Stege, Menno Dekker, Roel Faber, Frank Santegoets, Monique Sieber-Rasch and Ton Gerritsen.

REFERENCES

1. Anakwe RE, Middleton SD. Osteoarthritis at the base of the thumb. BMJ 2011;343:d7122.
2. van Uchelen J, Beumer A, Brink SM, Hoogvliet P, Moojen TM, Spaans AJ, Vermeulen GM, Schreuders TAR, Selles RW. Nederlandse Vereniging voor Plastische Chirurgie (NVPC); 2014 -]; Available from: https://www.nvpc.nl/uploads/stand/150416DOC-MB-Definitieve_richtlijn_Conservatieve_en_Chirurgische_behandeling_duimbasisartrose_28-10-2014_aangenomen_ALV_14_april_2015149.pdf.
3. Aebischer B, Elsig S, Taeymans J. Effectiveness of physical and occupational therapy on pain, function and quality of life in patients with trapeziometacarpal osteoarthritis - A systematic review and meta-analysis. Hand Ther 2016;21:5-15.
4. Berggren M, Joost-Davidsson A, Lindstrand J, Nylander G, Povlsen B. Reduction in the need for operation after conservative treatment of osteoarthritis of the first carpometacarpal joint: a seven year prospective study. Scand J Plast Reconstr Surg Hand Surg 2001;35:415-7.
5. Tsehaie J, Spekreijse KR, Wouters RM, Slijper HP, Feitz R, Hovius SER, Selles RW. Outcome of a Hand Orthosis and Hand Therapy for Carpometacarpal Osteoarthritis in Daily Practice: A Prospective Cohort Study. J Hand Surg Am 2018.
6. Jones G, Cooley HM, Bellamy N. A cross-sectional study of the association between Heberden's nodes, radiographic osteoarthritis of the hands, grip strength, disability and pain. Osteoarthr Cartilage 2001;9:606-11.
7. Dahaghin S, Bierma-Zeinstra SM, Ginai AZ, Pols HA, Hazes JM, Koes BW. Prevalence and pattern of radiographic hand osteoarthritis and association with pain and disability (the Rotterdam study). Ann Rheum Dis 2005;64:682-7.
8. Bierke S, Petersen W. Influence of anxiety and pain catastrophizing on the course of pain within the first year after uncomplicated total knee replacement: a prospective study. Arch Orthop Trauma Surg 2017;137:1735-42.
9. Hernandez C, Diaz-Heredia J, Berraquero ML, Crespo P, Loza E, Ruiz Iban MA. Pre-operative Predictive Factors of Post-operative Pain in Patients With Hip or Knee Arthroplasty: A Systematic Review. Reumatol Clin 2015;11:361-80.
10. Lewis GN, Rice DA, McNair PJ, Kluger M. Predictors of persistent pain after total knee arthroplasty: a systematic review and meta-analysis. Br J Anaesth 2015;114:551-61.
11. Hanusch BC, O'Connor DB, Ions P, Scott A, Gregg PJ. Effects of psychological distress and perceptions of illness on recovery from total knee replacement. Bone Joint J 2014;96-B:210-6.
12. London DA, Stepan JG, Boyer MI, Calfee RP. The impact of depression and pain catastrophization on initial presentation and treatment outcomes for atraumatic hand conditions. J Bone Joint Surg Am 2014;96:806-14.
13. Oh Y, Drijkoningen T, Menendez ME, Claessen FM, Ring D. The Influence of Psychological Factors on the Michigan Hand Questionnaire. Hand (N Y) 2017;12:197-201.
14. Broderick JE, Keefe FJ, Bruckenthal P, Junghaenel DU, Schneider S, Schwartz JE, Kaell AT, Caldwell DS, McKee D, Reed S and others. Nurse practitioners can effectively deliver pain coping skills training to osteoarthritis patients with chronic pain: A randomized, controlled trial. Pain 2014;155:1743-54.
15. de Raaij EJ, Pool J, Maissan F, Wittink H. Illness perceptions and activity limitations in osteoarthritis of the knee: a case report intervention study. Man Ther 2014;19:169-72.

16. Becker SJ, Bot AG, Curley SE, Jupiter JB, Ring D. A prospective randomized comparison of neoprene vs thermoplast hand-based thumb spica splinting for trapeziometacarpal arthrosis. Osteoarthr Cartilage 2013;21:668-75.

17. Becker SJ, Makarawung DJ, Spit SA, King JD, Ring D. Disability in patients with trapeziometacarpal joint arthrosis: incidental versus presenting diagnosis. J Hand Surg Am 2014;39:2009-15 e8.

18. Frouzakis R, Herren DB, Marks M. Evaluation of expectations and expectation fulfillment in patients treated for trapeziometacarpal osteoarthritis. J Hand Surg Am 2015;40:483-90.

19. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg Am 1998;23:575-87.

20. Marks M, Audige L, Herren DB, Schindele S, Nelissen RG, Vliet Vlieland TP. Measurement properties of the German Michigan Hand Outcomes Questionnaire in patients with trapeziometacarpal osteoarthritis. Arthritis Care Res (Hoboken) 2014;66:245-52.

21. Kroenke K, Spitzer RL, Williams JB, Lowe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50:613-21.

22. Lowe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, Schneider A, Brahler E. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. J Affect Disord 2010;122:86-95.

23. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychol Assessment 1995;7:524-32.

24. Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. The Pain Catastrophizing Scale: Further psychometric evaluation with adult samples. J Behav Med 2000;23:351-65.

25. Osman A, Barrios FX, Kopper BA, Hauptmann W, Jones J, O'Neill E. Factor structure, reliability, and validity of the pain catastrophizing scale. J Behav Med 1997;20:589-605.

26. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res 2006;60:631-7.

27. Moss-Morris R, Weinman J, Petrie KJ, Horne R, Cameron LD, Buick D. The revised Illness Perception Questionnaire (IPQ-R). Psychol Health 2002;17:1-16.

28. Broadbent E, Wilkes C, Koschwanez H, Weinman J, Norton S, Petrie KJ. A systematic review and meta-analysis of the Brief Illness Perception Questionnaire. Psychol Health 2015;30:1361-85.

29. de Raaij EJ, Schroder C, Maissan FJ, Pool JJ, Wittink H. Cross-cultural adaptation and measurement properties of the Brief Illness Perception Questionnaire-Dutch Language Version. Man Ther 2012;17:330-5.

30. Hallegraeff JM, van der Schans CP, Krijnen WP, de Greef MH. Measurement of acute nonspecific low back pain perception in primary care physical therapy: reliability and validity of the brief illness perception questionnaire. BMC Musculoskelet Disord 2013;14:53.

31. Timmermans I, Versteeg H, Meine M, Pedersen SS, Denollet J. Illness perceptions in patients with heart failure and an implantable cardioverter defibrillator: Dimensional structure, validity, and correlates of the brief illness perception questionnaire in Dutch, French and German patients. J Psychosom Res 2017;97:1-8.

32. Eaton RG, Glickel SZ. Trapeziometacarpal osteoarthritis. Staging as a rationale for treatment. Hand Clin 1987;3:455-71.

33. Dela Rosa TL, Vance MC, Stern PJ. Radiographic optimization of the Eaton classification. J Hand Surg Br 2004;29:173-7.

34. Hinkle DE, Wiersma W, Jurs SG. Boston, Mass.; [London]: Houghton Mifflin ; [Hi Marketing] (distributor); 2003 -]; Available from: http://catalog.hathitrust.org/api/volumes/oclc/50716608.html.

35. Visser AW, Boyesen P, Haugen IK, Schoones JW, van der Heijde DM, Rosendaal FR, Kloppenburg M. Radiographic scoring methods in hand osteoarthritis--a systematic literature search and descriptive review. Osteoarthr Cartilage 2014;22:1710-23.

36. Haugen IK, Slatkowsky-Christensen B, Boyesen P, van der Heijde D, Kvien TK. Cross-sectional and longitudinal associations between radiographic features and measures of pain and physical function in hand osteoarthritis. Osteoarthr Cartilage 2013;21:1191-8.

37. Hoffler CE, 2nd, Matzon JL, Lutsky KF, Kim N, Beredjiklian PK. Radiographic Stage Does Not Correlate With Symptom Severity in Thumb Basilar Joint Osteoarthritis. J Am Acad Orthop Surg 2015;23:778-82.

38. Kroon FPB, van Beest S, Ermurat S, Kortekaas MC, Bloem JL, Reijnierse M, Rosendaal FR, Kloppenburg M. In thumb base osteoarthritis structural damage is more strongly associated with pain than synovitis. Osteoarthr Cartilage 2018;26:1196-202.

39. Sonne-Holm S, Jacobsen S. Osteoarthritis of the first carpometacarpal joint: a study of radiology and clinical epidemiology. Results from the Copenhagen Osteoarthritis Study. Osteoarthr Cartilage 2006;14:496-500.

40. Murphy SL, Lyden AK, Phillips K, Clauw DJ, Williams DA. Association between pain, radiographic severity, and centrally-mediated symptoms in women with knee osteoarthritis. Arthritis Care Res (Hoboken) 2011;63:1543-9.

41. Hill S, Dziedzic K, Thomas E, Baker SR, Croft P. The illness perceptions associated with health and behavioural outcomes in people with musculoskeletal hand problems: findings from the North Staffordshire Osteoarthritis Project (NorStOP). Rheumatology (Oxford) 2007;46:944-51.

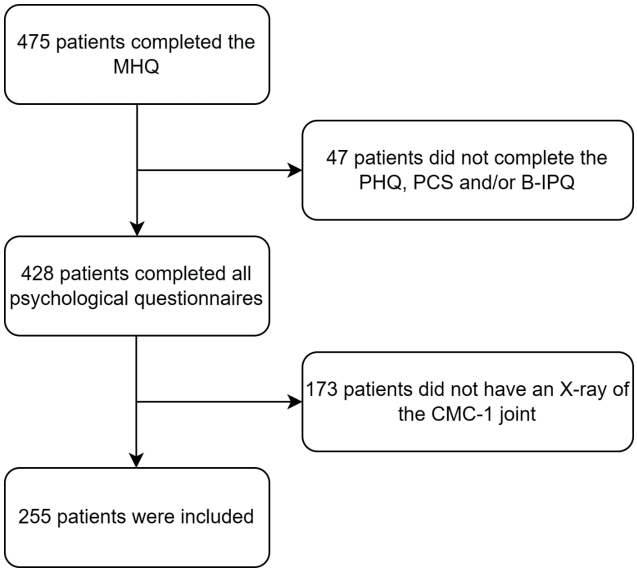


Figure 1. Flowchart

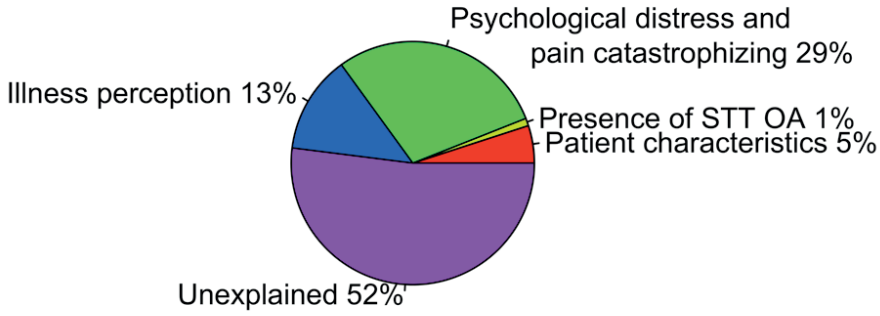


Figure 2. Increase in explained variance (increase in multiple R^2) of pre-treatment MHQ-pain per step in the hierarchical linear regression model

Table 1. Baseline characteristics of the patients included for analysis

	Total (n = 255)
Age in years	59.8 (7.8)
Sex (%)	
Female	74.9%
Hand dominance (%)	
Right	91.4%
Left	6.3%
Both	2.4%
Laterality of the affected hand (%)	
Right	47.8%
Left	50.2%
Both	2.0%
Dominant hand affected (%)	51.8%
Duration of symptoms in months (median, interquartile range)	12 (6-24)
Occupational intensity (%)	
Not employed	40.0%
Light	21.6%
Moderate	27.5%
Severe	11.0%
STT OA present (%)	13.3%
PHQ score	1.6 (2.6)
PCS score	12.2 (10.0)
B-IPQ Consequences	6.4 (2.3)
B-IPQ Timeline	7.8 (2.3)
B-IPQ Personal control	5.3 (2.3)
B-IPQ Treatment control	6.8 (1.8)
B-IPQ Identity	6.1 (2.5)
B-IPQ Concern	6.1 (2.7)
B-IPQ Understanding	8.4 (2.0)
B-IPQ Emotional response	4.3 (2.8)
MHQ-pain	52.9 (17.3)

* Values reported as mean (SD) unless otherwise stated.

Table 2. Beta coefficients and explained variance (R^2) for hierarchical linear regression models explaining pre-treatment pain levels. In each additional model, more variables potentially explaining pre-treatment pain levels are included. Both the unstandardized (B) and standardized (β) regression coefficients are reported.

	Model 1 (Patient characteristics)		Model 2 (Model 1+ Presence of STT OA)		Model 3 (Model 2+ Psychological distress and pain catastrophizing)		Model 4 (Model 3 + Illness perception)		Univariable models	
	B	**β**	**B**	**β**	**B**	**β**	**B**	**β**	**B**	**β**
Explanatory variables										
Age	-0.09	-0.04	-0.08	-0.03	-0.003	-0.001	-0.01	-0.005	-0.15	-0.07
Sex, male	-8.48***	-0.49***	-8.36**	-0.48**	-5.13*	-0.30*	-3.25	-0.19	-8.18	-0.47
Physical activity at work (ref = no work)										
Light	-1.85	-0.11	-1.94	-0.11	0.38	0.02	1.68	0.10	0.08	0.005
Moderate	-2.09	-0.12	-2.14	-0.12	-1.54	-0.09	-1.95	-0.11	-0.29	-0.02
Severe	3.05	0.18	4.03	0.23	2.52	0.15	1.29	0.07	3.07	0.18
STT OA, present			-4.87	-0.28	-3.62	-0.21	-2.55	-0.15	-4.56	-0.26
PHQ score					1.10*	0.16*	0.79*	0.12*	2.91	0.43
PCS score					0.75***	0.44***	0.46***	0.27***	0.94	0.55
B-IPQ Consequences							1.36**	0.19**	3.84	0.52
B-IPQ Timeline							-0.76	-0.10	0.69	0.09
B-IPQ Personal control							-0.56	-0.08	-1.29	-0.17
B-IPQ Treatment control							0.30	0.03	-0.52	-0.05
B-IPQ Identity							1.11**	0.16**	2.88	0.41
B-IPQ Concern							0.40	0.06	3.03	0.47
B-IPQ Understanding							0.25	0.03	0.33	0.04
B-IPQ Emotional response							0.72	0.12	3.12	0.51
Multiple R^2	0.051		0.060		0.346		0.474			
Adjusted R^2	0.032		0.037		0.325		0.439		___	

* $p \leq 0.05$
** $p \leq 0.01$
*** $p \leq 0.001$

10

ASSOCIATIONS BETWEEN POSITIVE TREATMENT OUTCOME EXPECTATIONS, ILLNESS UNDERSTANDING, AND OUTCOMES: A COHORT STUDY ON NON-OPERATIVE TREATMENT OF FIRST CARPOMETACARPAL OSTEOARTHRITIS

van der Oest, M. J. W., Hoogendam, L., Wouters, R. M., Vermeulen, G. M., Slijper, H. P., Selles, R. W., Vranceanu, A. M., Porsius, J. T., The Hand-Wrist Study Group (2022). *Disability and rehabilitation*, *44*(19), 5487–5494.

ABSTRACT

Purpose

More positive outcome expectations and illness perceptions are associated with better outcomes for patients with several osteoarthritic orthopedic conditions. However, it is unknown whether these factors also influence outcomes of non-operative treatment for first carpometacarpal osteoarthritis (CMC-1 OA). Therefore, we assess the role of pre-treatment outcome expectations and illness perceptions in reports of pain and hand function three months after non-operative treatment for CMC-1 OA.

Materials and methods

We conducted a cohort study with 219 patients treated non-operatively for CMC-1 OA between September 2017 and October 2018. Patients were included in the study if they completed measures of pain and hand function, illness perceptions (scale: 0-10) and expectations (scale: 3-27) as part of routine outcome measurements. Pain and hand function were measured before treatment and three months after starting treatment using the Dutch version of the Michigan Hand Outcome Questionnaire. Multivariable linear regression analysis was used to assess the influence of outcome expectations and illness perceptions on pain and hand function.

Results

Both positive outcome expectations (B = 0.64; 95%CI[0.1-1.2]) and a better illness understanding (an illness perception subdomain; B = 1.53; 95%CI[0.2-2.9]) at baseline were associated with less pain at three months. For hand function, similar estimated were found.

Conclusions

We found that positive outcome expectations and a better illness understanding, were associated with a better outcome of non-operative treatment for CMC-1 OA.

INTRODUCTION

Carpometacarpal osteoarthritis of the thumb (CMC-1 OA) is a disabling illness where non-operative treatment can often be successful[1]. The incidence of CMC-1 OA is estimated at 7.5%, of which 20 % seeks treatment [2]. Usually, treatment is started non-operatively with orthosis, injections, or hand therapy. As a recent study indicated, a good outcome of non-operative treatment for CMC-1 OA may delay and often prevent surgical treatment[3]. Given the considerable variation in treatment outcome for CMC-1 OA and modest patient satisfaction [4], it is important to understand which pre-treatment factors are associated with a better outcome of non-operative treatment for CMC-1 OA.

To date, it is mostly unknown which factors are associated with a better outcome of non-operative treatment for CMC-1 OA. One study found no predictive factors for treatment outcome over and above baseline pain and function [1]. However, in this study, the psychological mindset of patients at the start of the treatment was not assessed.

Several aspects of the influence of patient mindset on the outcome of treatment have been assessed before. For example, studies have found that patients who experience more psychological distress and who have a stronger tendency to catastrophize pain may benefit less from CMC-1 OA treatments [5-8]. However, these studies are limited as they did not adjust for baseline disease severity.

Two additional and potentially-important aspects of the patient mindset are the extent to which a patient has positive expectations about the efficacy of the treatment, and the illness perceptions of the patients. Several studies in other conditions have demonstrated a positive association between outcome expectations and better patient-reported treatment outcomes across a variety of medical conditions including osteoarthritis [9-14]. However, no prior studies have prospectively examined the role of expectations and illness perception on outcomes of non-operative treatment in CMC-1 OA.

Therefore, we investigated the association between treatment outcome expectations and illness perceptions and patient-reported pain and hand function three months after starting non-operative treatment of CMC-1 OA patients, while adjusting for baseline pain, function, psychological distress and catastrophic thinking about pain.

METHODS

Context

The study was performed at Xpert Clinic and Handtherapie Nederland, comprising of 22 outpatient clinics for hand surgery and therapy in the Netherlands, and took place between September 2017 and October 2018, after approval by the local Medical Ethical

Committee (Rotterdam, NL/sl/MEC-2018-1088). Patients were treated by hand therapists who received the same internal training on how to treat CMC-1 OA with hand therapy. Participants received treatment under the supervision of (generally) the same therapist, using a standardized protocol. All therapist are certified physical therapist with extensive experience as a hand therapist.

Patients

All patients receiving non-operative treatment for CMC-1 OA who completed psychological screening questionnaires before treatment and the Michigan Hand Outcome Questionnaire (MHQ) before treatment and three months after treatment were included in the cohort. Details of the data collection have been published earlier[15]. Patients and therapists were not blinded to the treatment.

Intervention

All patients were clinically diagnosed with CMC-1 OA based on presenting complaints and clinical signs. As defined in the Dutch guideline[16] for primary CMC-1 OA, all patients were offered non-operative treatment first, including an orthosis and/or hand therapy. In general, treatment consisted of prescribing a custom-made or prefabricated thumb orthosis (usually including CMC-1 palmar and radial abduction and slight metacarpophalangeal flexion) and one to two 25-minute sessions of hand therapy per week for a total of 12 weeks. The first six weeks of treatment aimed at correcting the position of the CMC-1 into a more stable position of palmar and radial abduction and prevention of flexion and adduction. This included coordination and mobility exercises[4] (4-6 times/day, 10-15 repetitions). The last six weeks of treatment were mainly focused on improving active stability and pinch strength and also included functional exercises[4] (2-3 times/day up until 50-100 repetitions). Additional or fewer sessions could be planned based on the therapist's judgment and the availability of the participant. A more detailed description of the treatment has been reported in a previous paper [1, 4].

Outcome measures

The primary endpoint for this study was three months after starting non-operative treatment. As the primary outcome, we used the pain subscale of the Dutch version of the MHQ[17], measured as part of routine outcome measurements. The MHQ is a validated patient-reported outcome measure to assess patients' pain and hand functioning [17, 18]. The secondary outcome was the hand function subscale of the MHQ. The questions of these subscales result in a 0 (severe pain or dysfunction) – 100 (no pain or dysfunction) score. All data was collected as part of routine outcome measurement using GemsTracker electronic data capture tools[19]. After the initial diagnosis, a hand therapist assigns prespecified routine outcome measures to all patients in the clinic. Patients receive emails to complete online

questionnaires at preset time points. This system has been described in more detail in an earlier publication[15]. Completing questionnaires is encouraged and facilitates evaluation of progress throughout the treatment, but completing questionnaires is not required for any part of the treatment.

Baseline demographics

Baseline characteristics of all patients (including age, sex, workload, duration of complaints, hand dominance, smoking status, and body mass index) were collected by the therapist during the first consultation.

Patient mindset

To assess patients' baseline mindset, patients completed four questionnaires: i) the Patient Health Questionnaire (PHQ-4)[20], a screening tool for depression and anxiety that results in a score from 0 (no psychological distress) – 12 (high psychological distress), ii) the Pain Catastrophizing Scale (PCS)[21], a questionnaire to assess a patient's tendency to catastrophize pain. The PCS ranges from 0 (no pain catastrophizing) – 52 (high pain catastrophizing), iii) the Credibility/ Expectancy Questionnaire (CEQ)[22], a questionnaire that measures patients' outcome expectations and credibility of the treatment and results in a score from 3 (low expectations and credibility) – 27 (high expectations and credibility). The CEQ specifically asks patients how much they "feel" or "think" the treatment will reduce symptoms and the physical limitation due to symptoms of their CMC-1 OA. In this study, we only evaluated the 3-item expectancy subscale of the CEQ. iv) The Brief Illness Perceptions Questionnaire (B-IPQ)[23], a questionnaire that measures how patients perceive their illness on eight separate domains, using a single 0-10 question for each domain. The fourth question of the IPQ asks patients how they think the treatment will affect their illness. Since this construct is evaluated using the expectancy domain of the CEQ as well, this item of the B-IPQ was not used in our analyses. We used the validated Dutch versions of all questionnaires [24-26].

Data analysis

First, we assessed the univariable associations of all baseline characteristics with pain and hand function at three months. However, in this analysis we adjusted for baseline MHQ, because we previously found that baseline pain and function were strongly related to the outcome[3]. Then, to asses which patient characteristics and mindset variables were independently associated with outcome, we performed two multivariable linear regressions with pain and hand function as outcomes. In addition to an overall multivariable model, we developed a stepwise multivariable model to assess the contribution of the different sets of variables to the explained variance (R^2). First, we added patient and disease characteristics to the model, second, we added psychological distress and pain catastrophizing, and finally,

we added outcome expectations and illness perceptions. We added psychological distress and pain catastrophizing first, because previous literature has shown associations between these variables and pain and hand function. The relationship between outcome expectations, illness perceptions and outcomes is unknown. Therefore, we added these variables last.

All analyses were conducted using R statistical computing, version 3.3.4. For all tests, a p-value ≤ 0.05 was considered statistically significant. We assumed that all relationships were linear and the model residuals were normally distributed. We confirmed these assumptions. A power calculation using G Power (version 3.1), based on a multivariable model, indicated that a sample of 159 patients was needed to detect a small to medium effect ($f = 0.1$) of the patient mindset on outcomes given a power of 0.8 and alpha of 0.05.

The funder played no role in the design, conduct, or reporting of this study.

RESULTS

Between September 2017 and October 2018, we included 219 patients. Figure 1 shows the flow of patients throughout the study and reasons for exclusion. 87% of all patients who completed the MHQ also completed all mindset questionnaires while at three months, 60% of these patients also completed the MHQ. There were no significant differences patients who did and did not complete all necessary questionnaires.

Table 1 presents the baseline demographic characteristics and mindset variables of all patients that are included in the analysis. Patients had a mean age (SD) of 60 (7) years old and the majority were female (76%). On average, the MHQ pain subscale at three months improved by 10 points ($p < 0.001$, 95%CI [8 – 12]) and the MHQ hand function subscale increased by 5 points ($p < 0.001$, 95%CI [3 – 7]) compared to baseline.

In the univariable analyses, where we only adjusted for baseline MHQ values, being a smoker was associated with more pain at three months. Additionally, having more positive outcome expectations was associated with less pain and increased hand function at three months (see Table 2).

In the multivariable analyses a more positive outcome expectation and a better illness understanding (an illness perception subdomain) of patients' illness were associated with less pain (see Table 3). For hand function, a more positive outcome expectation was associated with better hand function at three months. Figures 2 and 3 illustrate the magnitude of the effect of pre-treatment expectations and illness perceptions on pain and hand function at three months, illustrating both the systematic effects and the relatively large variation between subjects.

Stepwise linear regression analysis revealed that 32% of the variance in pain at three months could be explained by patient characteristics and baseline pain. No additional

variance could be explained by psychological distress and pain catastrophizing, but outcome expectations and illness perceptions explained an additional 5% of the variance over and above all other variables (Supplementary Table 1). For hand function, 28% of the variance at three months could be explained by patient characteristics and baseline hand function. An additional 1% could be explained by psychological distress and pain catastrophizing and outcome expectations and illness perceptions explained an additional 4% over and above all other variables (Supplementary Table 2).

DISCUSSION

We investigated the relation between the psychological mindset of CMC-1 OA patients at the start of a non-operative treatment and patient-reported pain and hand function at three months after treatment. We found that two aspects of the patient mindset, positive outcome expectations and a better illness understanding, were independently associated with a better outcome of non-operative treatment for CMC-1 OA. After adjusting for patient characteristics, pain catastrophizing and psychological distress, we found that two aspects of the patient mindset, positive outcome expectations and a better illness understanding, were independently associated with a better outcome of non-operative treatment for CMC-1 OA

Our finding that more positive outcome expectations are associated with better outcomes is in line with several other studies [27, 28]. For example, Blanchard et al. [27] found that women with higher self-efficacy and outcome expectations were more likely to be active during cardiac rehabilitation and Lurie et al. [28] found that lumbar disc herniation patients with more positive outcome expectations had better outcomes and were more physically active during cardiac rehabilitation.

Several mechanisms may explain the relationship between outcome expectation and patient-reported outcomes. For example, studies indicate that having more positive treatment outcome expectations may trigger psychobiological mechanisms, such as anxiety reduction, positive affectivity, cognitive reinterpretation, treatment adherence, and conditioning [29-31]. This is confirmed in studies showing that interventions aimed to optimize patients` outcome expectations have improved outcomes [32-35]. Future studies could investigate how different expectation management strategies affect patients with CMC-1 OA and evaluate their effects on outcomes in daily clinical practice.

In addition to outcome expectations, our results show that patients that report a better illness understanding have less pain at three months. This is in line with findings from Hanusch et al.[11], who found that better illness understanding was associated with better early recovery after total knee arthroplasty. Moreover, Mosleh et al.[36] found that patients

with coronary heart disease who reported better illness understanding were more likely to adhere to exercise therapy.

While we studied associations, several studies have investigated strategies to change patients' illness perception. For instance, Lee et al. [37] educated trauma patients on the theory of illness perceptions and asked them to identify inadequate perceptions. In this study, they found that these patients obtained more positive illness perceptions, however the influence on outcome was unfortunately not studied. Future studies may investigate if incorporating these strategies in the treatment of CMC-1 OA also leads to a more positive illness perception and, most importantly, in better outcomes.

In contrast with previous literature, we did not find an association between outcomes of non-operative treatment for CMC-1 OA and pain catastrophizing and psychological distress. For example, the papers by Das De et al.[6] and Lozano-Calderon et al.[7] showed that these factors play an important role in patients with CMC-1 OA. However, in these studies post-treatment PROMs were not adjusted for pre-treatment PROMs, which might explain the difference with our own findings. As CMC-1 OA complaints are known to be associated with pain catastrophizing and psychological distress before treatment[38], it may be worthwhile to further investigate potential indirect pathways through which these mindset variables affect treatment outcomes in CMC-1 OA.

The strengths of our study are its longitudinal design and the fact that this is the first to study the influence of expectations alongside other psychosocial factors on outcome of non-operative treatment of CMC-1 OA. More specifically, most upper extremity studies do not take pre-treatment symptoms into account when investigating the role of the patients' mindset, even though pre-treatment symptoms play an important role in predicting outcomes in upper extremity conditions [3, 39].

Study limitations

A limitation of our study is the non-response rate during our study; 87% of all patients who completed the MHQ also completed all mindset questionnaires while at three months, 60% of these patients also completed the MHQ. However, there were no significant differences in baseline characteristics between the non-responders and the patients included in the analysis, suggesting no selection bias on the reported analyses. Furthermore, this percentage of missing data is representative of routine longitudinal data collection. For example, Crijns et al.[40] found similar rates of missing data in hand surgery patients.

CONCLUSION

Our finding that higher outcome expectation leads to better outcomes may challenge the common belief in orthopedic surgery, hand surgery and hand therapy that a patient should

not have too high expectations [41]. Discussing outcome expectations at the start of the non-operative treatment, in particular with patients who appear skeptic about the potential treatment benefits, might contribute to better outcomes. Our results also indicate that explaining the illness to a patient may also improve the outcome of treatment. Therefore, it might be worthwhile for clinicians to ensure that the patient understands the etiology and prognosis of their illness.

REFERENCES

1. Tsehaie J, Spekreijse KR, Wouters RM, Slijper HP, Feitz R, Hovius SER, et al. Outcome of a Hand Orthosis and Hand Therapy for Carpometacarpal Osteoarthritis in Daily Practice: A Prospective Cohort Study. J Hand Surg Am. 2018;43(11):1000-9 e1.
2. Chaisson CE, Zhang Y, McAlindon TE, Hannan MT, Aliabadi P, Naimark A, et al. Radiographic hand osteoarthritis: incidence, patterns, and influence of pre-existing disease in a population based sample. J Rheumatol. 1997;24(7):1337-43.
3. Tsehaie J, Spekreijse KR, Wouters RM, Feitz R, Hovius S, Slijper H, et al. Predicting outcome after hand orthosis and hand therapy for thumb carpometacarpal osteoarthritis; a prospective study. Arch Phys Med Rehabil. 2018.
4. Wouters RM, Tsehaie J, Slijper HP, Hovius SER, Feitz R, Hand-Wrist Study G, et al. Exercise Therapy in Addition to an Orthosis Reduces Pain More Than an Orthosis Alone in Patients With Thumb Base Osteoarthritis: A Propensity Score Matching Study. Arch Phys Med Rehabil. 2018.
5. Becker SJ, Bot AG, Curley SE, Jupiter JB, Ring D. A prospective randomized comparison of neoprene vs thermoplast hand-based thumb spica splinting for trapeziometacarpal arthrosis. Osteoarthritis Cartilage. 2013;21(5):668-75.
6. Das De S, Vranceanu AM, Ring DC. Contribution of kinesophobia and catastrophic thinking to upper-extremity-specific disability. J Bone Joint Surg Am. 2013;95(1):76-81.
7. Lozano-Calderon SA, Souer JS, Jupiter JB, Ring D. Psychological differences between patients that elect operative or nonoperative treatment for trapeziometacarpal joint arthrosis. Hand (N Y). 2008;3(3):271-5.
8. Makarawung DJ, Becker SJ, Bekkers S, Ring D. Disability and pain after cortisone versus placebo injection for trapeziometacarpal arthrosis and de Quervain syndrome. Hand (N Y). 2013;8(4):375-81.
9. Waljee J, McGlinn EP, Sears ED, Chung KC. Patient expectations and patient-reported outcomes in surgery: a systematic review. Surgery. 2014;155(5):799-808.
10. Smeets RJ, Beelen S, Goossens ME, Schouten EG, Knottnerus JA, Vlaeyen JW. Treatment expectancy and credibility are associated with the outcome of both physical and cognitive-behavioral treatment in chronic low back pain. Clin J Pain. 2008;24(4):305-15.
11. Hanusch BC, O'Connor DB, Ions P, Scott A, Gregg PJ. Effects of psychological distress and perceptions of illness on recovery from total knee replacement. Bone Joint J. 2014;96-B(2):210-6.
12. Dalili Z, Bayazi MH. The effectiveness of Mindfulness-Based Cognitive Therapy on the illness perception and Psychological Symptoms in patients with Rheumatoid Arthritis. Complement Ther Clin Pract. 2019;34:139-44.
13. Zyrianova Y, Kelly BD, Sheehan J, McCarthy C, Dinan TG. The psychological impact of arthritis: the effects of illness perception and coping. Ir J Med Sci. 2011;180(1):203-10.
14. Berner C, Erlacher L, Fenzl KH, Dorner TE. A cross-sectional study on self-reported physical and mental health-related quality of life in rheumatoid arthritis and the role of illness perception. Health Qual Life Outcomes. 2018;16(1):238.
15. Selles RW, Wouters RM, Poelstra R, van der Oest MJW, Porsius JT, Hovius SER, et al. Routine health outcome measurement: Development, design, and implementation of the Hand and Wrist Cohort. Plast Reconstr Surg. 2020.

16. Handchirurgie NVv. Conservatieve en Chirurgische Behandeling van Primaire Artrose van de Duimbasis. Nederlandse Vereniging voor Plastische Chirurgie; 2014.

17. Huijsmans R SH, Aufdemkampe G. . Michigan Hand Outcomes Questionnaire-DLV: een vragenlijst voor patienten met handfunctieproblemen. Fysiopraxis. 2001;9:38-41.

18. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg Am. 1998;23(4):575-87.

19. Zorgbedrijven EMaE. GemsTracker, copyright© [updated version 1.8.2. Available from: https://gemstracker.org.

20. Kroenke K, Spitzer RL, Williams JB, Lowe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics. 2009;50(6):613-21.

21. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychol Assessment. 1995;7(4):524-32.

22. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry. 2000;31(2):73-86.

23. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res. 2006;60(6):631-7.

24. de Raaij EJ, Schroder C, Maissan FJ, Pool JJ, Wittink H. Cross-cultural adaptation and measurement properties of the Brief Illness Perception Questionnaire-Dutch Language Version. Man Ther. 2012;17(4):330-5.

25. Timmermans I, Versteeg H, Meine M, Pedersen SS, Denollet J. Illness perceptions in patients with heart failure and an implantable cardioverter defibrillator: Dimensional structure, validity, and correlates of the brief illness perception questionnaire in Dutch, French and German patients. J Psychosom Res. 2017;97:1-8.

26. Mertens VC, Moser A, Verbunt J, Smeets R, Goossens M. Content Validity of the Credibility and Expectancy Questionnaire in a Pain Rehabilitation Setting. Pain Pract. 2017;17(7):902-13.

27. Blanchard C, Arthur HM, Gunn E. Self-efficacy and outcome expectations in cardiac rehabilitation: associations with women's physical activity. Rehabil Psychol. 2015;60(1):59-66.

28. Lurie JD, Henderson ER, McDonough CM, Berven SH, Scherer EA, Tosteson TD, et al. Effect of Expectations on Treatment Outcome for Lumbar Intervertebral Disc Herniation. Spine (Phila Pa 1976). 2016;41(9):803-9.

29. Morrison JD, Stuifbergen AK. Outcome expectations and physical activity in persons with longstanding multiple sclerosis. J Neurosci Nurs. 2014;46(3):171-9.

30. Stetler C. Adherence, expectations and the placebo response: why is good adherence to an inert treatment beneficial? Psychol Health. 2014;29(2):127-40.

31. Geers AL, Miller FG. Understanding and translating the knowledge about placebo effects: the contribution of psychology. Curr Opin Psychiatry. 2014;27(5):326-31.

32. Rief W, Shedden-Mora MC, Laferton JA, Auer C, Petrie KJ, Salzmann S, et al. Preoperative optimization of patient expectations improves long-term outcome in heart surgery patients: results of the randomized controlled PSY-HEART trial. BMC Med. 2017;15(1):4.

33. Lo SHS, Chang AM, Chau JPC. Stroke Self-Management Support Improves Survivors' Self-Efficacy and Outcome Expectation of Self-Management Behaviors. Stroke. 2018;49(3):758-60.

34. Kube T, Glombiewski JA, Rief W. Using Different Expectation Mechanisms to Optimize Treatment of Patients With Medical Conditions: A Systematic Review. Psychosomatic Medicine. 2018;80(6):535-43.

35. Kube T, Rief W, Gollwitzer M, Glombiewski JA. Introducing an EXperimental Paradigm to investigate Expectation Change (EXPEC). J Behav Ther Exp Psychiatry. 2018;59:92-9.

36. Mosleh SM, Almalik MM. Illness perception and adherence to healthy behaviour in Jordanian coronary heart disease patients. Eur J Cardiovasc Nurs. 2016;15(4):223-30.

37. Lee BO, Chien CS, Hung CC, Chou PL. Effects of an in-hospital nursing intervention on changing illness perceptions in patients with injury. J Adv Nurs. 2015;71(11):2540-50.

38. Hoogendam L, van der Oest MJW, Tsehaie J, Wouters RM, Vermeulen GM, Slijper HP, et al. Psychological factors are more strongly associated with pain than radiographic severity in non-invasively treated first carpometacarpal osteoarthritis. Disabil Rehabil. 2019:1-6.

39. MacDermid JC, Valdes K, Szekeres M, Naughton N, Algar L. The assessment of psychological factors on upper extremity disability: A scoping review. J Hand Ther. 2018;31(4):511-23.

40. Crijns TJ, Bernstein DN, Gonzalez R, Wilbur D, Ring D, Hammert WC. Operative Treatment is Not Associated with More Relief of Depression Symptoms than Nonoperative Treatment in Patients with Common Hand Illness. Clin Orthop Relat Res. 2020;478(6):1319-29.

41. Waljee JF, Chung KC. Commentary regarding "Evaluation of expectations and expectation fulfillment in patients treated for trapeziometacarpal osteoarthritis". J Hand Surg Am. 2015;40(3):491-2.

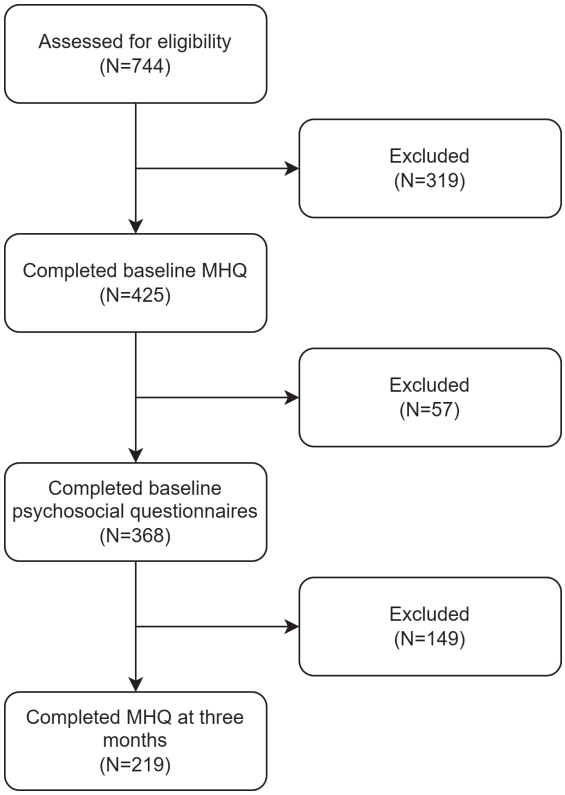


Figure 1. Flow diagram of the study.
MHQ = Michigan Hand outcome Questionnaire

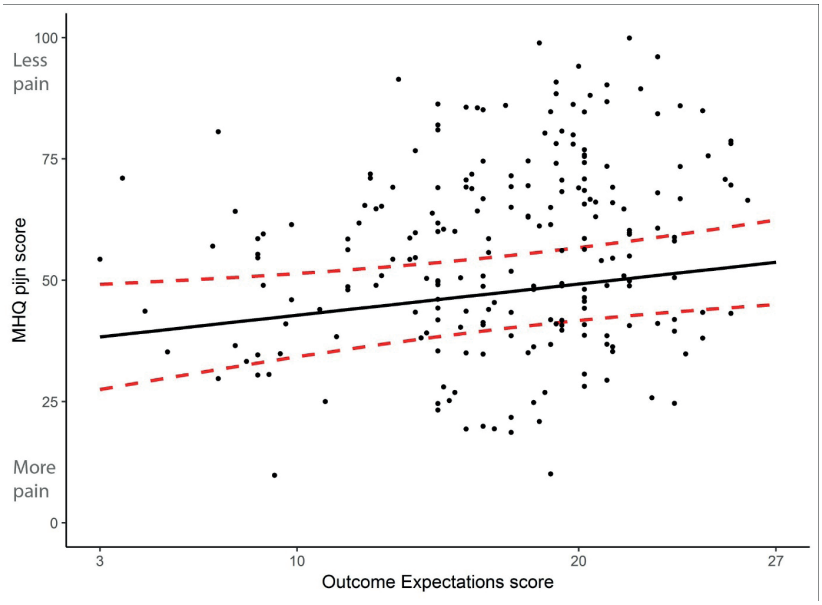


Figure 2A

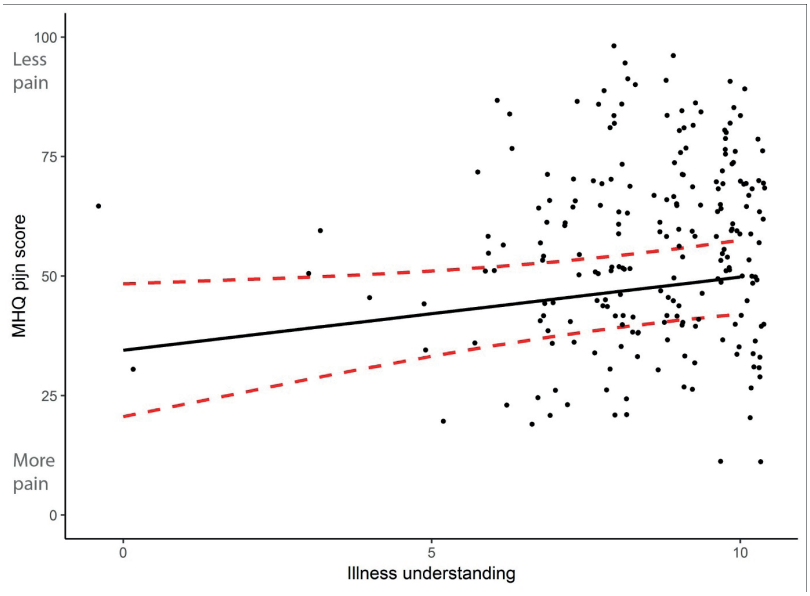


Figure 2B

Figure 2AB. Effect plots of association between outcome expectations (A) and illness understanding (B) and pain at three months. All points represent individual patients. Jitter, minimal random variance, has been added to display overlapping points. Higher Michigan Hand outcome Questionnaire (MHQ) pain score on the y-axis represents less pain. Higher scores on the x-axes represent more positive outcome expectations or better understanding.

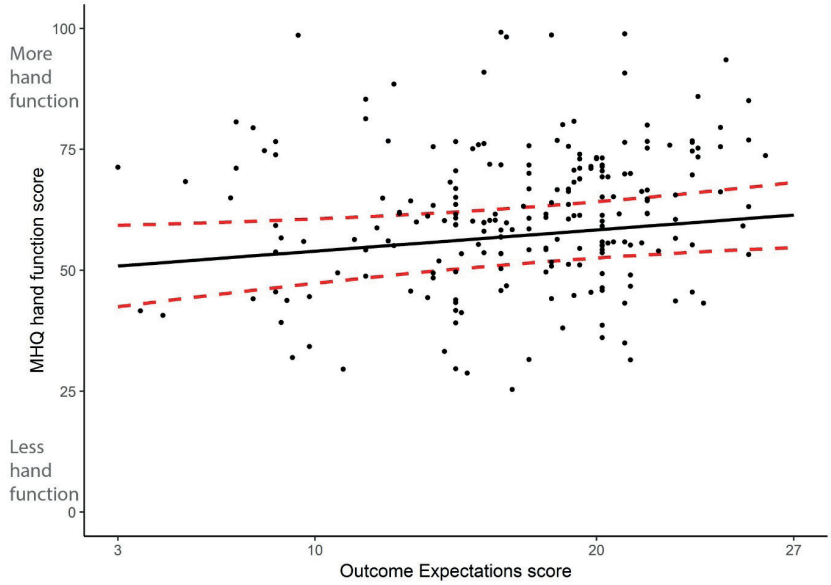


Figure 3. Effect plots of association between expectations and hand function at three months. All points represent the scores of individual patients. Jitter has been added to display overlapping points. Higher Michigan Hand outcome Questionnaire (MHQ) hand function score on the y-axis represents more hand function. Higher scores on the x-axes represent more positive outcome expectations.

Table 1. Baseline patient and psychosocial characteristics of the study population (N=219 patients receiving non-operative treatment for first carpometacarpal osteoarthritis)

	Value	Questionnaire Range
Age, year (SD)	60 (7)	
Sex, Male, (%)	52 (24)	
Duration of complaints, months, median (IQR)	10 (5-24)	
BMI, mean (SD)	27 (5)	
Current non-smoker, (%)	195 (89)	
Workload, (%)		
No paid labor	84 (38)	
Light physical labor	45 (21)	
Medium physical labor	64 (29)	
Heavy physical labor	26 (12)	
Baseline MHQ pain, mean (SD)	46 (16)	0 (worst) – 100 (best)
Baseline MHQ hand function, mean (SD)	56 (17)	0 (worst) – 100 (best)
PHQ Psychological Distress, median (IQR)	0 (0-2)	0 (best) – 12 (worst)
Pain Catastrophizing Score, median (IQR)	10 (5-18)	0 (best) – 52 (worst)
CEQ Expectations Score, median (IQR))	18 (15-20)	3 (worst) – 27 (best)
IPQ consequences, median (IQR)	7 (5-8)	0 (best) – 10 (worst)
IPQ timeline, median (IQR)	8 (6-10)	0 (best) – 10 (worst)
IPQ personal control, median (IQR)	5 (4-7)	0 (worst) – 10 (best)
IPQ identity, median (IQR)	7 (5-8)	0 (best) – 10 (worst)
IPQ concern, median (IQR)	7 (5-8)	0 (best) – 10 (worst)
IPQ illness comprehensibility, median (IQR)	9 (8-10)	0 (worst) – 10 (best)
IPQ emotional consequences, median (IQR)	4 (1.5-7)	0 (best) – 10 (worst)

Michigan Hand outcome Questionnaire (MHQ), Body Mass Index (BMI), Patient Health Questionnaire(PHQ), Credibility and Expectations Questionnaire (CEQ), Illness Perceptions Questionnaire (IPQ)

Table 2. Multivariable associations between MHQ at three months and individual predictors, adjusted for baseline MHQ, indicating the association of the different variables with the MHQ outcome

	MHQ Pain - three months			MHQ Hand function - three months		
	B	95%CI	β	B	95%CI	β
Age	0.26	[0-0.6]	0.10	0.00	[-0.2-0.2]	0.00
Sex. Male	-0.45	[-5.8-4.9]	-0.02	2.91	[-1-6.8]	0.20
Duration of complaints (months)	0.02	[-0.1-0.1]	0.03	-0.03	[-0.1-0]	-0.05
BMI	0.09	[-0.3-0.5]	0.02	-0.03	[-0.4-0.3]	-0.01
Current non-smoker	**9.53**	**[2.4-16.7]**	**0.49**	4.03	[-1.3-9.4]	0.28
Workload						
No paid labor (reference)	-	-			-	
Light physical labor	3.99	[-2-10]	0.21	3.50	[-1.1-8.1]	0.24
Medium physical labor	-2.21	[-7.6-3.2]	-0.11	0.04	[-4.1-4.2]	0.00
Heavy physical labor	-6.46	[-13.8-0.9]	-0.33	-1.80	[-7.4-3.8]	-0.12
PHQ Psychological Distress (0-12)	-0.56	[-1.6-0.4]	-0.07	**-0.78**	**[-1.5-0]**	**-0.12**
Pain catastrophizing scale (0-52)	-0.06	[-0.3-0.2]	-0.03	-0.18	[-0.4-0]	-0.11
CEQ Expectancy scale (3-27)	**0.66**	**[0.2-1.1]**	**0.16**	**0.53**	**[0.2-0.9]**	**0.17**
IPQ Consequences (0-10)	-0.28	[-1.4-0.8]	-0.03	-0.36	[-1.1-0.4]	-0.06
IPQ Timeline (0-10)	-0.86	[-1.9-0.1]	-0.10	-0.56	[-1.3-0.2]	-0.09
IPQ Personal control (0-10)	0.30	[-0.7-1.3]	0.03	**0.96**	**[0.2-1.7]**	**0.14**
IPQ Identity (0-10)	-0.46	[-1.4-0.5]	-0.06	-0.29	[-1-0.4]	-0.05
IPQ Concern (0-10)	-0.67	[-1.7-0.3]	-0.09	-0.50	[-1.2-0.2]	-0.09
IPQ Illness comprehensibility (0-10)	1.18	[-0.2-2.5]	0.10	0.40	[-0.6-1.4]	0.04
IPQ Emotional consequences (0-10)	0.14	[-0.7-1]	0.02	-0.55	[-1.2-0.1]	-0.11

Bold indicates statistically significant covariates
Michigan Hand outcome Questionnaire (MHQ). Body Mass Index (BMI). Patient Health Questionnaire (PHQ). Credibility and Expectations Questionnaire (CEQ). Illness Perceptions Questionnaire (IPQ)

Table 3. Multivariable linear regression model for pain and hand function at three months. adjusted for all covariates, indicating the adjusted contributions of all variables in a multivariable model. The bold numbers indicate the variables that are significantly associated with the outcome.

	MHQ Pain – three months			MHQ Hand function – three months		
	B	95%CI	β	B	95%CI	β
Age	0.21	[-0.2-0.6]	0.08	-0.10	[-0.4-0.2]	-0.05
Sex. Male	-1.64	[-7.4-4.1]	-0.08	3.20	[-1.2-7.6]	0.22
Duration of complaints (months)	0.03	[-0.1-0.1]	0.04	-0.01	[-0.1-0.1]	-0.01
BMI	0.12	[-0.3-0.6]	0.03	-0.03	[-0.4-0.3]	-0.01
Current non-smoker	**8.94**	**[1.6-16.3]**	**0.46**	3.44	[-2.3-9.2]	0.24
Workload						
No paid labor (reference)	1	-	0	1	-	0
Light physical labor	6.41	[0.1-12.8]	0.33	2.86	[-2.1-7.8]	0.20
Medium physical labor	0.43	[-5.8-6.7]	0.02	0.10	[-4.8-5]	0.01
Heavy physical labor	-3.65	[-11.7-4.4]	-0.19	-2.11	[-8.3-4.1]	-0.14
Baseline MHQ score (0-100)	**0.56**	**[0.4-0.7]**	**0.47**	**0.40**	**[0.3-0.5]**	**0.47**
PHQ Psychological Distress (0-12)	-0.39	[-1.5-0.8]	-0.05	-0.37	[-1.3-0.5]	-0.06
Pain catastrophizing scale (0-52)	0.11	[-0.2-0.5]	0.05	-0.01	[-0.3-0.2]	0.00
CEQ Expectancy scale (3-27)	**0.64**	**[0.1-1.2]**	**0.15**	**0.44**	**[0-0.8]**	**0.14**
IPQ Consequences (0-10)	0.13	[-1.2-1.5]	0.02	0.14	[-0.9-1.2]	0.02
IPQ Timeline (0-10)	-0.53	[-1.6-0.5]	-0.06	-0.12	[-0.9-0.7]	-0.02
IPQ Personal control (0-10)	-0.12	[-1.2-0.9]	-0.01	0.72	[-0.1-1.5]	0.11
IPQ Identity (0-10)	-0.13	[-1.2-1]	-0.02	-0.11	[-0.9-0.7]	-0.02
IPQ Concern (0-10)	-0.66	[-1.9-0.6]	-0.09	0.09	[-0.9-1.1]	0.02
IPQ Illness comprehensibility (0-10)	**1.53**	**[0.2-2.9]**	**0.13**	0.22	[-0.8-1.3]	0.02
IPQ Emotional consequences (0-10)	0.76	[-0.3-1.8]	0.11	-0.19	[-1-0.6]	-0.04

Bold indicates statistically significant covariates

Michigan Hand outcome Questionnaire (MHQ), Body Mass Index (BMI), Patient Health Questionnaire (PHQ), Credibility and Expectations Questionnaire (CEQ), Illness Perceptions Questionnaire (IPQ)

11

PATIENTS WITH HIGHER TREATMENT OUTCOME EXPECTATIONS ARE MORE SATISFIED WITH THE RESULTS OF NONOPERATIVE TREATMENT FOR THUMB BASE OSTEOARTHRITIS: A COHORT STUDY

Hoogendam, L., van der Oest, M. J. W., Wouters, R. M., Andrinopoulou, E. R., Vermeulen, G. M., Slijper, H. P., Porsius, J. T., Selles, R. W., The Hand-Wrist Study Group (2021). *Archives of physical medicine and rehabilitation*, *102*(8), 1533–1540.

ABSTRACT

Objective

To investigate how satisfaction with treatment outcome is associated with patient mindset and Michigan Hand Outcome Questionnaire (MHQ) scores at baseline and 3 months in patients receiving non-operative treatment for thumb base osteoarthritis (CMC-1 OA).

Design

Cohort study

Setting

20 outpatient locations of a clinic for hand surgery and hand therapy in The Netherlands.

Participants

Patients (n=308) receiving non-operative treatment for CMC-1 OA, including exercise therapy, an orthosis, or both, between September 2017 and February 2019.

Interventions

Non-operative treatment (i.e. exercise therapy, an orthosis, or both)

Main outcome measure

Satisfaction with treatment outcomes was measured after three months of treatment. We measured total MHQ score at baseline and at three months. As baseline mindset factors, patients completed questionnaires on treatment outcome expectations, illness perceptions, pain catastrophizing, and psychological distress. We used multivariable logistic regression analysis and mediation analysis to identify factors associated with satisfaction with treatment outcomes.

Results

More positive pre-treatment outcome expectations were associated with a higher probability of being satisfied with treatment outcomes at three months (odds ratio = 1.15, 95% CI 1.07-1.25). Only a relatively small part (33%) of this association was due to a higher total MHQ score at three months. None of the other mindset and hand function variables at baseline were associated with satisfaction with treatment outcomes.

Conclusions

This study demonstrates that patients with higher pre-treatment outcome expectations are more likely to be satisfied with treatment outcomes after three months of non-operative

treatment for CMC-1 OA. This association could only partially be explained by a better functional outcome at three months for patients who were satisfied. Health care providers treating patients non-operatively for CMC-1 OA should be aware of the importance of expectations and may take this into account in pre-treatment counseling.

INTRODUCTION

Osteoarthritis of the first carpometacarpal joint (CMC-1 OA) is a common disease, especially in postmenopausal women (1). Symptoms include pain, limitations in activities, and loss of hand function (2). Several non-operative and operative treatment options are available (3). Current practice is to treat patients non-operatively first. This often consists of exercise therapy, an orthosis, or both (4-6). Surgery can be considered if symptoms are not sufficiently relieved by non-operative treatment (4). Non-operative treatment is a successful treatment strategy for CMC-1 OA, which, on average, reduces pain and improves hand function (7, 8). Moreover, in a large cohort with a mean follow-up of 2.2 years, only 15% of the patients underwent further surgery (9).

Treatment outcomes such as pain relief and functional improvement have been frequently studied, however, in recent days there is increasing attention for the patients' interpretation of their treatment outcomes (10). Previous studies demonstrated that after hand therapy and an orthosis for CMC-1 OA , there is considerable variation in patients' satisfaction with treatment outcomes. However, it is still unknown which factors explain this variation in satisfaction with treatment outcomes for these patients (5, 9).

It has previously been reported that pain and hand function after treatment are associated with satisfaction with treatment outcomes for patients with hand and wrist disorders receiving surgical treatment or steroid injections (11, 12). Additionally, patient mindset has been shown to be associated with satisfaction with treatment outcomes, again following either surgical treatment or steroid injections (12, 13).

Patient mindset can be seen as particular associations and expectations that a patient has, which could affect a patients' attitude towards his treatment (14). Because patients generally have a particular mindset towards a treatment before starting treatment, communication between a clinician and patients could be an opportunity to modify this mindset, for example by changing expectations (15).

There have already been studies on the association between expectations and satisfaction in daily practice, but there is no consensus on this. Several authors suggested that patients with high expectations would be less satisfied, because these patients are less likely to have their expectations fulfilled (11, 13, 16). Many surgeons apply this principle in practice (17). However, other studies have suggested that patients should have positive expectations to improve treatment outcomes (18-22). Possibly, this suggests that there is an optimum for expectations. Because of the conflicting suggestions in literature, there currently is no consensus or best practice on how clinicians should deal with patients' expectations in order to optimize treatment outcomes and satisfaction.

While it has been reported that pain, hand function, and patient mindset are associated with satisfaction with treatment outcomes for patients receiving surgical treatment or

steroid injections for hand and wrist conditions, it remains unknown if these factors also explain satisfaction with treatment outcomes in patients receiving non-operative treatment for CMC-1 OA. In particular, the role of expectations is unclear. Therefore, the purpose of this study was to investigate which baseline characteristics, including total Michigan Hand Outcome Questionnaire (MHQ) score and patient mindset, are associated with the likelihood of being satisfied with treatment outcomes after three months of non-operative treatment for CMC-1 OA when accounting for total MHQ score at three months.

METHODS

Setting and study population

Between September 2017 and February 2019, this cohort study was performed with routine outcome measurement data from Xpert Clinic and Handtherapie Nederland,comprising 20 outpatient locations for hand surgery and hand therapy in The Netherlands. Over 150 hand therapists and 23 European Board certified (FESSH) hand surgeons are employed in our clinic. The cohort and data collection procedures have previously been described in more detail (23). All patients provided written informed consent and this study was approved by the Erasmus MC Medical Ethical Committee.

Patients were included when treated non-operatively for CMC-1 OA after being diagnosed with CMC-1 OA by a Federation of European Societies for Surgery of the Hand (FESSH) certified hand surgeon. The diagnosis was made based on clinical presentation and X-rays when required. Non-operative treatment consisted of immobilizing the CMC-1 joint using an orthosis and performing exercises to improve the active stability of the CMC-1 joint and strength of the thenar muscles. This treatment protocol has previously been described in more detail (5). Non-operative treatment was offered for at least three months before surgery was considered. All hand therapists received the same training on how to treat CMC-1 OA patients. However, treatment was not fully standardized as in randomized controlled trials and therapists could deviate from this protocol based on patient preferences and clinical considerations.

All patients were invited to complete the questionnaires as part of routine clinical care before and after treatment. The questionnaires were sent after the first consultation with the hand surgeon. In addition, baseline characteristics including age, gender, occupational intensity, duration of symptoms, hand dominance, and affected hand, were collected. Occupational intensity was categorized as: not employed, light occupational intensity (e.g., working in an office), moderate occupational intensity (e.g., working in a shop), or severe occupational intensity (e.g., construction work). Patients who did not complete all questionnaires of interest at baseline and three months were excluded from the study.

Outcome measurements

Satisfaction with treatment outcomes was the primary outcome measure. This was measured using a self-designed questionnaire administered three months after the start of non-operative treatment. In this questionnaire, we asked patients "To what extent are you satisfied with the treatment outcomes obtained so far?", which could be rated as poor, moderate, fair, good or excellent. We have dichotomized this, classifying patients rating their satisfaction with treatment outcomes as poor, moderate or fair as less satisfied, while classifying patients rating their satisfaction as good or excellent as satisfied. We dichotomized satisfaction, since the number of patients in some groups was not enough for analysis, and, from a clinical point of view, because we aimed to identify factors that predicted whether patients would be satisfied or dissatisfied with treatment outcomes.

At baseline and three months, patients were invited to complete the Michigan Hand Outcome Questionnaire (MHQ) (24). The MHQ is a patient-reported outcome measure with good reliability, validity, and responsiveness for CMC-1 OA patients (25). The MHQ consists of six domains (pain, hand function, aesthetics, work, activities of daily life, and satisfaction with hand function), each with a score ranging from 0-100 (0 = poorest function, 100 = ideal function). From these subscales, a total MHQ score is calculated for the affected hand, which is used in our analysis. We chose to use the total MHQ score at three months as functional improvement measure since it comprises a broad spectrum of domains relevant to patients with CMC-1 OA.

We used the Credibility and Expectancy Questionnaire (CEQ) (26) to measure outcome expectations regarding the treatment and the credibility of the treatment. This questionnaire has two domains (expectations and credibility) with three questions each. Scores per domain range from 3-27 (higher scores indicate higher expectations/credibility of the treatment).

In addition, we measured pain catastrophizing behavior, psychological distress, and illness perceptions. We measured this using the Pain Catastrophizing Scale (PCS)(27), the Patient Health Questionnaire-4 (PHQ-4)(28) and the Brief Illness Perception Questionnaire (B-IPQ) (29) respectively. We calculated a total score for each questionnaire. All questionnaires have been validated and good reliability has been reported (26, 30-35).

Statistical methods

We compared baseline characteristics for patients that were satisfied and less satisfied with treatment outcomes using T-tests for normally-distributed continuous data and Mann-Whitney-Wilcoxon tests for continuous data that were not normally distributed. Chi-square statistics were used for categorical data. Effect sizes (Cohen's d) were calculated for any statistically significant differences between continuous data. We performed a non-responder analysis to compare baseline characteristics of patients completing all questionnaires of

interest (responders) and patients who only completed the MHQ at baseline and three months.

After dichotomizing satisfaction with treatment outcomes, we used multivariable logistic regression analysis to determine which baseline variables were associated with the probability of being satisfied with treatment outcomes when adjusting for patient characteristics, patient-reported hand function at baseline, and patients' mindset. For the logistic regression model, odds ratios (OR) with 95% confidence intervals (95% CI) were calculated.

In the multivariable regression model, we included the aforementioned baseline characteristics, patient-reported hand function (total MHQ score), psychological factors (total PCS score, total PHQ-4 and total B-IPQ score), CEQ Expectancy Score, and CEQ Credibility Score. Using the rule of thumb of one variable per ten cases having the lowest-frequency outcome ('number of events') to fit a multivariable logistic regression model and with 13 variables of interest, we needed to include at least 130 patients with an event. Asssuming that 50% of all patients would be classified as satisfied (5), a minimum of 260 patients is needed. Since the number of patients treated in our clinic during the study period exceeded 260, we included all patients treated in our clinic during the study period.

Because both patients with overly high and very low expectations have been suggested to be less satisfied, we hypothesized that there might be an optimum for CEQ Expectancy Score (12, 13, 16, 18-22). We therefore tested whether our multivariable regression model would better fit the data when a non-linear effect of the CEQ Expectancy Score was included, using splines. We performed a likelihood ratio test to determine whether this model had a better fit than the model with CEQ Expectancy Score as a linear term.

We checked for multicollinearity in our multivariable logistic regression model using the variance inflation factor. We considered a variance inflation factor greater than 10 an indication for multicollinearity (36).

As a secondary analysis (in addition to our multivariable logistic regression), we performed a mediation analysis (37, 38). As previous studies in hand surgery and orthopaedics reported an association between treatment outcomes and satisfaction, the aim of the mediation analysis was to quantify how much of the association between a predictor and the dependent variable of interest (in this case: satisfaction with treatment outcomes) is the result of treatment outcomes (indirect effect) and how much is independent of that (direct effect) (12, 39). For this mediation model, our predictor and mediator variables were continuous, and our outcome was binary (40-43). For the mediation model, we used the linear terms of all variables.

We assessed whether mediation was present by bootstrapping the indirect effect, as proposed by Preacher and Hayes (44). We corrected for total MHQ score at intake in the

analysis. The outcome of the mediation analysis is the proportion mediated, which is the percentage of the effect of the significant independent variable(s) due to the total MHQ score at three months. This is calculated by multiplying the regression coefficient of the predictor on the mediator with the regression coefficient of the mediator on the outcome.

Mediation analysis was performed in Mplus version 8.1, using Mplus code based on Feingold et al. (45) for mediation with a continuous predictor and a non-rare binary outcome. All other analyses were performed using R statistical computing, version 3.5.2. For all tests, a p-value smaller than 0.05 was considered statistically significant.

RESULTS

In the study period, 656 patients were treated non-operatively for CMC-1 OA at our clinic and completed all relevant questionnaires at baseline. Of those patients, 308 patients also completed all questionnaires of interest three months after the start of treatment and were included in the analysis (Figure 1).

Of the included patients, 76% were female, the mean age was 61 years (SD 8), and the mean total MHQ score at baseline was 60 (SD 15). Table 1 shows all baseline characteristics of the included patients. We compared the baseline characteristics of the included patients (responders) to the patients who did not complete all questionnaires (non-responders) and only found that the included patients had a shorter duration of symptoms (responders: median 9, IQR (5-24), non-responders: median 12, IQR (6-24))(Supplemental table 1).

Figure 2 shows the distribution of satisfaction on the 5-point Likert scale and the division into the two categories; of the 308 patients, 46% were satisfied with treatment outcomes, while 54% were less satisfied with treatment outcomes.

Patients that were less satisfied with treatment outcomes reported worse MHQ score, lower expectations of the treatment, and less treatment credibility at baseline compared to patients that were satisfied with treatment outcomes (Table 1). Also, patients that were less satisfied scored worse on psychological distress, pain catastrophizing, and illness perceptions.

While several baseline variables were associated with satisfaction with outcome in the univariable analysis (Table 1), higher CEQ Expectancy Score was the only significant variable associated with higher probability of being satisfied with treatment outcomes in the multivariable analysis (OR = 1.15, 95% CI 1.07-1.25)(Table 2). This odds ratio indicates that patients with one point more on the CEQ Expectancy Score have a 15% increase in odds of being satisfied with treatment outcomes.

We visualized the effect of the CEQ Expectancy Score on satisfaction with treatment outcomes on the original scale in Figure 3, showing a linear trend. Additionally, the likelihood ratio test showed that the model with non-linear effects of the CEQ Expectancy

Score did not have significantly better fit (p=0.23) than the model with the CEQ Expectancy Score as a linear term. Therefore, we used the linear term of CEQ Expectancy Score in all analyses. We therefore were not able to find an optimum.

Based on the variance inflation factor, we did not find an indication for multicollinearity in the multivariable logistic regression model.

Mediation analysis

Based on the results from the multivariable logistic regression analysis, we hypothesized that patients with a higher CEQ Expectancy Score would also have a higher total MHQ score at three months and would, therefore, be more satisfied. To test this hypothesis, we performed a post-hoc mediation analysis. We found that only 33% of the effect of CEQ Expectancy Score on satisfaction with treatment outcomes was due to total MHQ score at three months (Figure 4). The remaining 67% of the effect of expectations can either be explained by a direct effect on satisfaction or by an indirect effect through another factor that was not measured. This indicates that a better MHQ score at three months can only partially explain the association between CEQ Expectancy Score and satisfaction with treatment outcomes.

DISCUSSION

In this study, we found that patients with more positive expectations of the treatment outcome are more likely to be satisfied with treatment outcomes after three months of non-operative treatment for CMC-1 OA. Additionally, we found that only one-third of this effect was due to better treatment outcomes at three months.

Previous studies on other types of osteoarthritis reported that patients with higher expectations have better treatment outcomes such as pain and function (46-52). This is in agreement with our finding that higher expectations are associated with a higher total MHQ score at three months. In addition, we found a positive association between higher expectations and satisfaction with treatment outcomes, which is not completely in line with previous studies. For example, Jain et al. (48), Mahomed et al. (52) and Neuprez et al. (53) also reported an association between higher expectations and satisfaction with treatment outcomes after orthopaedic surgery, while several authors reported a negative association between expectations and satisfaction (54) or suggested that it would be better to lower expectations (11, 16, 49). Possibly, these different findings may be explained by different treatments, different diseases (e.g., hip or knee OA), different questionnaires to assess expectations or different pre-treatment counselling. However, our results suggest that it would be better to optimize expectations to improve satisfaction and, to a lesser extent, improve treatment outcomes. A trial evaluating effects of different expectation

management strategies of clinicians might provide valuable insights on how to address patients' expectations of their treatment.

Perhaps, the time point where patients completed the CEQ may explain why we found a positive association between treatment outcome expectations and satisfaction with outcomes. Patients in our study completed the CEQ after the first consultation with the hand surgeon. It could be that surgeons know from experience which patients will respond well to non-operative treatment and will, therefore, provide individualized information on the expected results of the treatment to patients. This could, in turn, influence the CEQ Expectancy Score in individual patients. However, in our analysis, we controlled for patient characteristics, patient-rated hand function and patients' mindset and still found no other predictive baseline factors for satisfaction with treatment outcomes than CEQ Expectancy Score.

Strengths and limitations

A strength of this study is the large sample size in a population-based cohort. Second, to our knowledge, this is the first study investigating satisfaction with treatment outcomes after exercise therapy, an orthosis, or both for CMC-1 OA. Third, by performing a mediation analysis, we were able to provide more insight into the mechanism of the association we found.

However, our study also has several limitations. Satisfaction with treatment outcomes is a complex construct that is difficult to measure and difficult to fully comprehend, as, for example, pointed out in an editorial by Ring and Leopold (55). Ring and Leopold describe that there are many reasons why one patient can be satisfied with the treatment outcome while another patient may be dissatisfied, while having the same treatment outcome. However, while satisfaction with treatment result is a difficult construct and influenced by many factors, it is also a very important and relevant outcome measure in striving for patient-centered care. Therefore, future studies investigating the underlying mechanisms that determine a patient's satisfaction with treatment outcome are needed to provide patient-centered care that is tailored to the patient's needs.

Another limitation is that, in this study, we used a self-designed questionnaire to assess satisfaction with treatment outcomes, which has not been validated yet. However, to our knowledge, there are no validated patient-reported outcome measures available to measure satisfaction with treatment outcomes for patients with hand and wrist disorders. The questionnaire we used and the dichotomization of our outcome measure is very similar to questionnaires and analyses used in previous studies, which facilitates us to compare our results with other studies (11, 16). However, to avoid dichotomization and the loss of information as a result, a validated patient-reported outcome measure for satisfaction with treatment outcomes with a continuous scale is needed to further optimize personalized care for individual patients.

Previous studies have reported that the context of treatment (e.g., communication style of the health care provider) also affects satisfaction with treatment outcomes (19, 20, 56, 57), hence a limitation is that we did not include a measure for treatment context in our study. Future studies should therefore include such measures in their analysis when studying satisfaction with outcomes, as well as study how to incorporate this into clinical practice.

In our study we found that more positive expectations are associated with higher satisfaction with treatment outcomes after three months of non-operative treatment for CMC-1 OA, which can only partially be explained by better treatment outcomes. This suggests that optimizing expectations might improve satisfaction and, to a lesser extent, improve treatment outcomes. However, future experimental studies are needed to determine whether modifying expectations of patients receiving non-operative treatment for CMC-1 OA, for example by framing pre-treatment information in a positive manner, will positively affect satisfaction and treatment outcomes.

CONCLUSION

This study demonstrates that patients with higher treatment outcome expectations are more likely to be satisfied with treatment outcomes after three months of non-operative therapy for CMC-1 OA. Health care providers treating patients non-operatively for CMC-1 OA should be aware of the importance of expectations and should take this into account in pre-treatment counseling.

ACKNOWLEDGMENTS

The authors thank all patients who participated and allowed their data to be anonymously used for the present study.

GROUP COLLABORATION

The collaborators of the Hand-Wrist Study Group are: RAM Blomme, BJR Sluijter, DJJC van der Avoort, A Kroeze, J Smit, J Debeij, ET Walbeehm, GM van Couwelaar, GM Vermeulen, JP de Schipper, JFM Temming, JH van Uchelen, HL de Boer, KP de Haas, K Harmsen, OT Zöphel, R Feitz, JS Souer, R Koch, SER Hovius, TM Moojen, X Smit, R van Huis, PY Pennehouat, K Schoneveld, YE van Kooij, RM Wouters, P Zagt, FJ van Ewijk, JJ Veltkamp, A Fink, WA de Ridder, HP Slijper, RW Selles, JT Porsius, J Tsehaie, R Poelstra, MC Jansen, MJW van der Oest, PO Sun, L Hoogendam, JS Teunissen, Jak Dekker, M Jansen-Landheer, M ter Stege

REFERENCES

1. Dahaghin S, Bierma-Zeinstra SM, Ginai AZ, Pols HA, Hazes JM, Koes BW. Prevalence and pattern of radiographic hand osteoarthritis and association with pain and disability (the Rotterdam study). Ann Rheum Dis. 2005;64(5):682-7.
2. Anakwe RE, Middleton SD. Osteoarthritis at the base of the thumb. BMJ. 2011;343:d7122.
3. Pickrell BB, Eberlin KR. Thumb Basal Joint Arthritis. Clin Plast Surg. 2019;46(3):407-13.
4. van Uchelen J, Beumer A, Brink SM, Hoogvliet P, Moojen TM, Spaans AJ, et al. Richtlijn Conservatieve en Chirurgische Behandeling van Primaire Artrose van de Duimbasis [Dutch]: Nederlandse Vereniging voor Plastische Chirurgie (NVPC); 2014 [Available from: https://www.nvpc.nl/uploads/stand/150416DOC-MB-Definitieve_richtlijn_Conservatieve_en_Chirurgische_behandeling_duimbasisartrose_28-10-2014_aangenomen_ALV_14_april_2015149.pdf.
5. Wouters RM, Tsehaie J, Slijper HP, Hovius SER, Feitz R, Hand-Wrist Study G, et al. Exercise Therapy in Addition to an Orthosis Reduces Pain More Than an Orthosis Alone in Patients With Thumb Base Osteoarthritis: A Propensity Score Matching Study. Arch Phys Med Rehabil. 2018.
6. Kloppenburg M, Kroon FP, Blanco FJ, Doherty M, Dziedzic KS, Greibrokk E, et al. 2018 update of the EULAR recommendations for the management of hand osteoarthritis. Ann Rheum Dis. 2019;78(1):16-24.
7. Hamasaki T, Laprise S, Harris PG, Bureau NJ, Gaudreault N, Ziegler D, et al. Efficacy of non-surgical interventions for trapeziometacarpal (thumb base) osteoarthritis: A systematic review. Arthritis Care Res (Hoboken). 2019.
8. Kroon FPB, Carmona L, Schoones JW, Kloppenburg M. Efficacy and safety of non-pharmacological, pharmacological and surgical treatment for hand osteoarthritis: a systematic literature review informing the 2018 update of the EULAR recommendations for the management of hand osteoarthritis. RMD Open. 2018;4(2):e000734.
9. Tsehaie J, Spekreijse KR, Wouters RM, Slijper HP, Feitz R, Hovius SER, et al. Outcome of a Hand Orthosis and Hand Therapy for Carpometacarpal Osteoarthritis in Daily Practice: A Prospective Cohort Study. J Hand Surg Am. 2018.
10. Graham B. Defining and Measuring Patient Satisfaction. J Hand Surg Am. 2016;41(9):929-31.
11. Marks M, Audige L, Reissner L, Herren DB, Schindele S, Vliet Vlieland TP. Determinants of patient satisfaction after surgery or corticosteroid injection for trapeziometacarpal osteoarthritis: results of a prospective cohort study. Arch Orthop Trauma Surg. 2015;135(1):141-7.
12. Marks M, Herren DB, Vliet Vlieland TP, Simmen BR, Angst F, Goldhahn J. Determinants of patient satisfaction after orthopedic interventions to the hand: a review of the literature. J Hand Ther. 2011;24(4):303-12 e10; quiz 12.
13. Kadzielski J, Malhotra LR, Zurakowski D, Lee SG, Jupiter JB, Ring D. Evaluation of preoperative expectations and patient satisfaction after carpal tunnel release. J Hand Surg Am. 2008;33(10):1783-8.
14. Crum AJ, Leibowitz KA, Verghese A. Making mindset matter. BMJ. 2017;356:j674.
15. Crum A, Zuckerman B. Changing Mindsets to Enhance Treatment Effectiveness. JAMA. 2017;317(20):2063-4.
16. Frouzakis R, Herren DB, Marks M. Evaluation of expectations and expectation fulfillment in patients treated for trapeziometacarpal osteoarthritis. J Hand Surg Am. 2015;40(3):483-90.

17. Waljee JF, Chung KC. Commentary regarding "Evaluation of expectations and expectation fulfillment in patients treated for trapeziometacarpal osteoarthritis". J Hand Surg Am. 2015;40(3):491-2.

18. Brown WA. Expectation, the placebo effect and the response to treatment. R I Med J (2013). 2015;98(5):19-21.

19. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. Lancet. 2001;357(9258):757-62.

20. Testa M, Rossettini G. Enhance placebo, avoid nocebo: How contextual factors affect physiotherapy outcomes. Man Ther. 2016;24:65-74.

21. Cormier S, Lavigne GL, Choiniere M, Rainville P. Expectations predict chronic pain treatment outcomes. Pain. 2016;157(2):329-38.

22. Hayden JA, Wilson MN, Riley RD, Iles R, Pincus T, Ogilvie R. Individual recovery expectations and prognosis of outcomes in non-specific low back pain: prognostic factor review. Cochrane Database Syst Rev. 2019;2019(11).

23. Selles RW, Wouters RM, Poelstra R, van der Oest MJW, Porsius JT, Hovius SER, et al. Routine health outcome measurement: Development, design, and implementation of the Hand and Wrist Cohort. Plast Reconstr Surg. 2020.

24. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg Am. 1998;23(4):575-87.

25. Marks M, Audige L, Herren DB, Schindele S, Nelissen RG, Vliet Vlieland TP. Measurement properties of the German Michigan Hand Outcomes Questionnaire in patients with trapeziometacarpal osteoarthritis. Arthritis Care Res (Hoboken). 2014;66(2):245-52.

26. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry. 2000;31(2):73-86.

27. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychol Assessment. 1995;7(4):524-32.

28. Kroenke K, Spitzer RL, Williams JB, Lowe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics. 2009;50(6):613-21.

29. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res. 2006;60(6):631-7.

30. de Raaij EJ, Schroder C, Maissan FJ, Pool JJ, Wittink H. Cross-cultural adaptation and measurement properties of the Brief Illness Perception Questionnaire-Dutch Language Version. Man Ther. 2012;17(4):330-5.

31. Hallegraeff JM, van der Schans CP, Krijnen WP, de Greef MH. Measurement of acute nonspecific low back pain perception in primary care physical therapy: reliability and validity of the brief illness perception questionnaire. BMC Musculoskelet Disord. 2013;14:53.

32. Lowe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. J Affect Disord. 2010;122(1-2):86-95.

33. Mertens VC, Moser A, Verbunt J, Smeets R, Goossens M. Content Validity of the Credibility and Expectancy Questionnaire in a Pain Rehabilitation Setting. Pain Pract. 2017;17(7):902-13.

34. Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. The Pain Catastrophizing Scale: Further psychometric evaluation with adult samples. J Behav Med. 2000;23(4):351-65.

35. Osman A, Barrios FX, Kopper BA, Hauptmann W, Jones J, O'Neill E. Factor structure, reliability, and validity of the Pain Catastrophizing Scale. J Behav Med. 1997;20(6):589-605.

36. Kutner MH, Nachtsheim CJ, Neter J, Li W. Applied linear statistical models2005.

37. Baron RM, Kenny DA. The Moderator Mediator Variable Distinction in Social Psychological-Research - Conceptual, Strategic, and Statistical Considerations. J Pers Soc Psychol. 1986;51(6):1173-82.

38. Zhao XS, Lynch JG, Chen QM. Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. J Consum Res. 2010;37(2):197-206.

39. Kahlenberg CA, Nwachukwu BU, McLawhorn AS, Cross MB, Cornell CN, Padgett DE. Patient Satisfaction After Total Knee Replacement: A Systematic Review. HSS J. 2018;14(2):192-201.

40. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods. 2010;15(4):309-34.

41. Rijnhart JJM, Twisk JWR, Eekhout I, Heymans MW. Comparison of logistic-regression based methods for simple mediation analysis with a dichotomous outcome variable. BMC Med Res Methodol. 2019;19(1):19.

42. Vanderweele TJ. Explanation in causal inference: Methods for mediation and interaction. New York, NY, US: Oxford University Press; 2015. xvi, 706-xvi, p.

43. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. Am J Epidemiol. 2010;172(12):1339-48.

44. Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav Res Meth Ins C. 2004;36(4):717-31.

45. Feingold A, MacKinnon DP, Capaldi DM. Mediation analysis with binary outcomes: Direct and indirect effects of pro-alcohol influences on alcohol use disorders. Addict Behav. 2018.

46. Filbay SR, Judge A, Delmestri A, Arden NK, Group COS. Evaluating Patients' Expectations From a Novel Patient-Centered Perspective Predicts Knee Arthroplasty Outcome. J Arthroplasty. 2018;33(7):2146-52 e4.

47. Gandhi R, Davey JR, Mahomed N. Patient expectations predict greater pain relief with joint arthroplasty. J Arthroplasty. 2009;24(5):716-21.

48. Jain D, Bendich I, Nguyen LL, Nguyen LL, Lewis CG, Huddleston JI, et al. Do Patient Expectations Influence Patient-Reported Outcomes and Satisfaction in Total Hip Arthroplasty? A Prospective, Multicenter Study. J Arthroplasty. 2017;32(11):3322-7.

49. Jain D, Nguyen LL, Bendich I, Nguyen LL, Lewis CG, Huddleston JI, et al. Higher Patient Expectations Predict Higher Patient-Reported Outcomes, But Not Satisfaction, in Total Knee Arthroplasty Patients: A Prospective Multicenter Study. J Arthroplasty. 2017;32(9S):S166-S70.

50. Judge A, Cooper C, Arden NK, Williams S, Hobbs N, Dixon D, et al. Pre-operative expectation predicts 12-month post-operative outcome among patients undergoing primary total hip replacement in European orthopaedic centres. Osteoarthr Cartilage. 2011;19(6):659-67.

51. Lingard EA, Sledge CB, Learmonth ID, Kinemax Outcomes G. Patient expectations regarding total knee arthroplasty: differences among the United States, United Kingdom, and Australia. J Bone Joint Surg Am. 2006;88(6):1201-7.

52. Mahomed NN, Liang MH, Cook EF, Daltroy LH, Fortin PR, Fossel AH, et al. The importance of patient expectations in predicting functional outcomes after total joint arthroplasty. J Rheumatol. 2002;29(6):1273-9.

53. Neuprez A, Delcour JP, Fatemi F, Gillet P, Crielaard JM, Bruyere O, et al. Patients' Expectations Impact Their Satisfaction following Total Hip or Knee Arthroplasty. PLoS One. 2016;11(12):e0167911.

54. Soroceanu A, Ching A, Abdu W, McGuire K. Relationship between preoperative expectations, satisfaction, and functional outcomes in patients undergoing lumbar and cervical spine surgery: a multicenter study. Spine (Phila Pa 1976). 2012;37(2):E103-8.

55. Ring D, Leopold SS. Editorial-Measuring Satisfaction: Can It Be Done? Clin Orthop Relat Res. 2015;473(10):3071-3.

56. Rossettini G, Carlino E, Testa M. Clinical relevance of contextual factors as triggers of placebo and nocebo effects in musculoskeletal pain. BMC Musculoskelet Disord. 2018;19(1):27.

57. Suarez-Almazor ME, Looney C, Liu Y, Cox V, Pietz K, Marcus DM, et al. A randomized controlled trial of acupuncture for osteoarthritis of the knee: effects of patient-provider communication. Arthritis Care Res (Hoboken). 2010;62(9):1229-36.

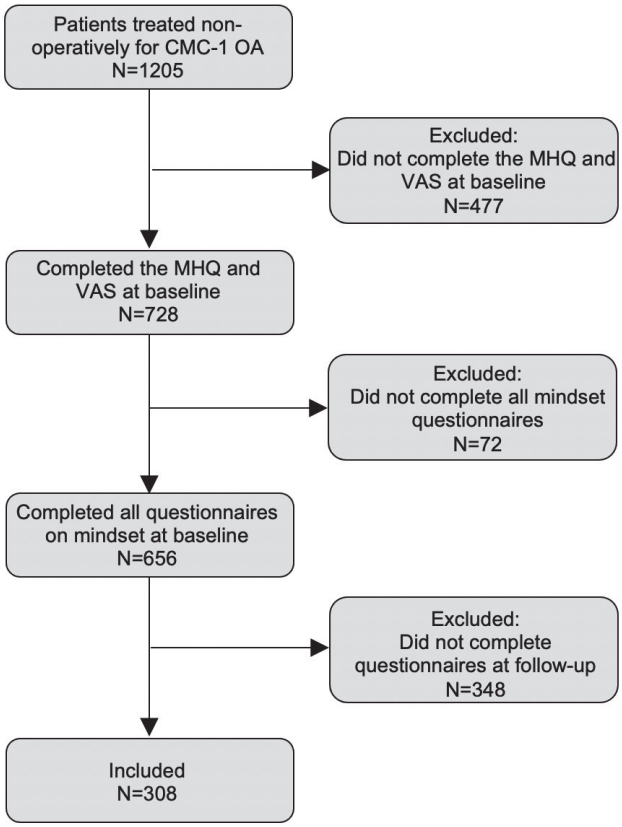


Figure 1. Flow chart of patient inclusion

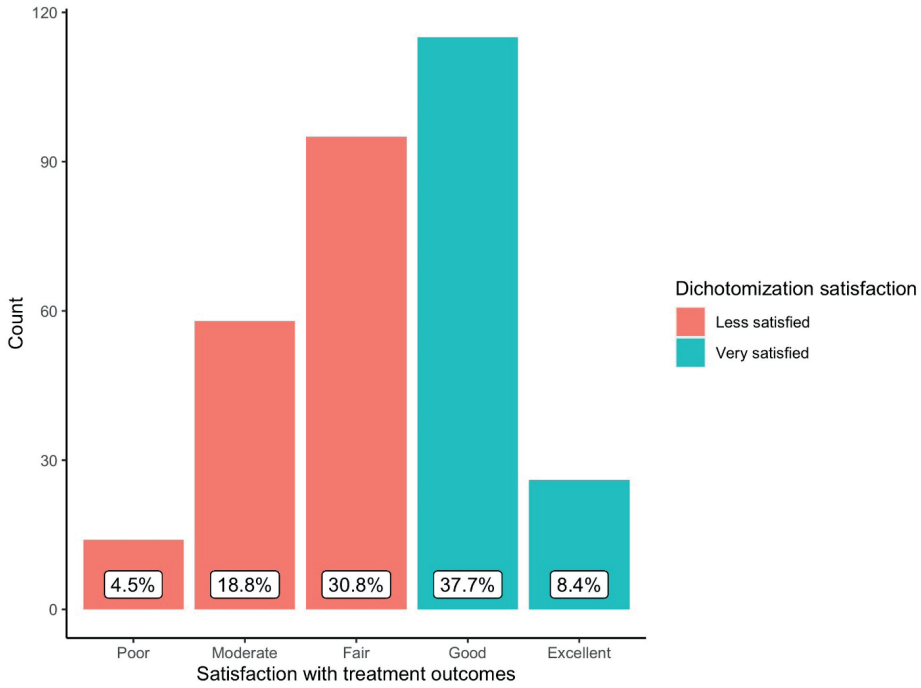


Figure 2. The distribution of satisfaction with treatment outcomes using the original 5-point scale and the distribution of dichotomized satisfaction with treatment outcomes.

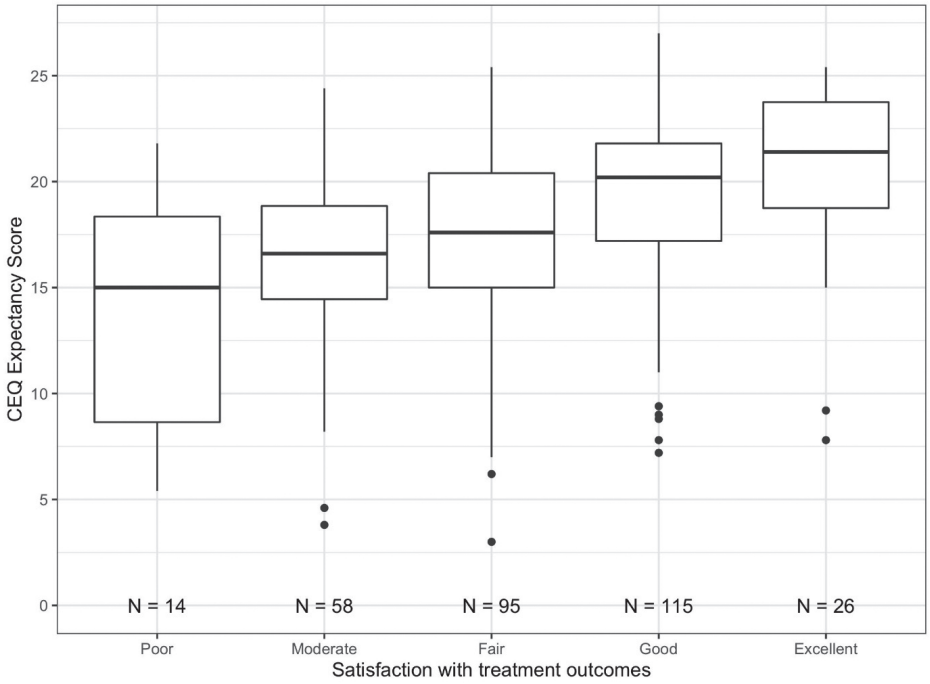


Figure 3. Box-and-whisker plot of CEQ Expectancy Score (range 3-27) per satisfaction category. The horizontal line represents the median and boxes represent the first and third quartile. The whiskers represent 1.5 times the interquartile range from the first and respectively third quartile.

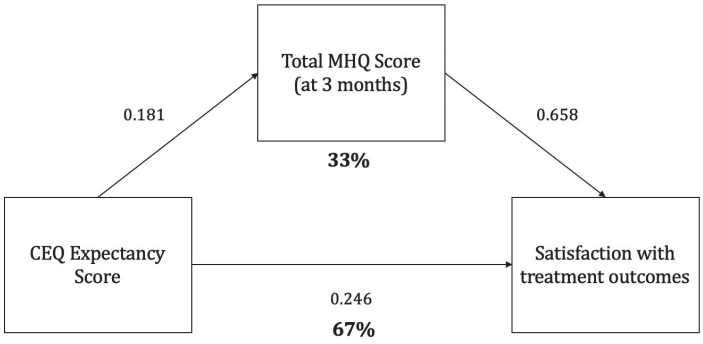


Figure 4. Mediation model. The relation between CEQ Expectancy Score and satisfaction with treatment outcomes was mediated by the total MHQ score at three months. The standardized regression coefficients are reported from the regression of CEQ Expectancy Score on total MHQ Score at three months and the regression of CEQ Expectancy Score and total MHQ Score at three months on satisfaction with treatment outcomes. All regression analyses were corrected for total MHQ score at baseline. The indirect effect (0.119) was divided by the total effect (0.365) of CEQ Score on satisfaction with treatment outcomes to obtain the proportion mediated (33%).

Table 1. Patient characteristics at baseline

	Questionnaire range (if applicable)	All included patients (n=308)	Satisfied with outcomes (n=141)	Less satisfied with outcomes (n=167)	Effect size	P-value
Age in years		61 (8)	61 (7)	61 (8)		0.781
Sex (%)						0.712
Female		76%	77%	75%		
Hand dominance (%)						0.683
Right		89%	88%	90%		
Left		6%	8%	5%		
Both		5%	4%	5%		
Affected hand (%)						0.525
Right		42%	42%	43%		
Left		46%	48%	44%		
Both		12%	10%	14%		
Dominant hand affected (%)		41%	42%	40%		0.849
Duration of symptoms in months, median (interquartile range)		9 (5-24)	8 (4-24)	12 (6-24)		0.454
Workload (%)						0.628
Not employed		39%	38%	40%		
Light		20%	22%	19%		
Moderate		30%	31%	29%		
Severe		11%	9%	13%		
MHQ score	0 – 100	60 (15)	63 (15)	57 (15)	0.37	0.001
PHQ score	0 – 12	1.2 (2.2)	0.9 (2.0)	1.4 (2.3)	-0.23	0.045
PCS score	0 – 52	11 (9)	10 (8)	13 (10)	-0.35	0.002
B-IPQ score	0 – 80	51 (9)	50 (10)	52 (8)	-0.23	0.044
CEQ Expectancy Score	3 – 27	18 (5)	20 (4)	17(5)	0.67	<0.001
CEQ Credibility Score	3 - 27	21 (4)	22 (4)	20 (4)	0.47	<0.001

* Values reported as mean (SD) unless otherwise stated.
MHQ = Michigan Hand outcomes Questionnaire, PHQ = Patient Health Questionnaire-4, PCS = Pain Catastrophizing Scale, B-IPQ = Brief Illness Perception Questionnaire, CEQ = Credibility and Expectancy Questionnaire

Table 2. Multivariable logistic regression analysis on satisfaction with treatment outcomes

	OR (95% CI)
Age	1.01 (0.97-1.05)
Sex, male	0.77 (0.41-1.44)
Dominant hand affected	1.15 (0.70-1.89)
Duration of symptoms	1.01 (1.00-1.02)
Workload	
Light	1.21 (0.59-2.48)
Moderate	1.15 (0.57-2.30)
Severe	1.02 (0.39-2.60)
MHQ score	1.02 (1.00-1.04)
PHQ score	1.00 (0.86-1.14)
PCS score	0.99 (0.95-1.02)
B-IPQ score	1.02 (1.00-1.04)
CEQ Expectancy Score	1.15 (1.07-1.25)***
CEQ Credibility Score	1.03 (0.95-1.12)

* $p \leq 0.05$
** $p \leq 0.01$
*** $p \leq 0.001$

12

WHICH FACTORS ARE ASSOCIATED WITH SATISFACTION WITH TREATMENT RESULTS IN PATIENTS WITH HAND AND WRIST CONDITIONS? A LARGE COHORT ANALYSIS

De Ridder, W. A., Wouters, R. M., Hoogendam, L., Vermeulen, G. M., Slijper, H. P., Selles, R. W., The Hand-Wrist Study Group (2022). *Clinical orthopaedics and related research*, *480*(7), 1287–1301.

My contribution to this publication was that I helped with the analyses and contributed to the manuscript's writing and revision. I have included this publication in my thesis, because it validates the findings from my previous study "Patients With Higher Treatment Outcome Expectations Are More Satisfied With the Results of Nonoperative Treatment for Thumb Base Osteoarthritis: A Cohort Study" in a broader population of patients with hand and wrist conditions.

ABSTRACT

Background

Satisfaction with treatment results is an important outcome domain in striving for patient-centered and value-based healthcare. Although numerous studies have investigated factors associated with satisfaction with treatment results, most studies used relatively small samples. Additionally, many studies have only investigated univariable associations instead of multivariable associations; to our awareness, none have investigated the independent association of baseline sociodemographics, quality of life, improvement in pain and function, experiences with healthcare delivery, and baseline measures of mental health with satisfaction with treatment results.

Questions/purposes

(1) What factors are independently associated with satisfaction with treatment results at 3 months post-treatment in patients treated for common hand and wrist conditions? (2) What factors are independently associated with the willingness to undergo the treatment again at 3 months post-treatment in patients treated for common hand and wrist conditions? Among the factors under study were baseline sociodemographics, quality of life, improvement in pain and function, experiences with healthcare delivery, and baseline measures of mental health.

Methods

Between August 2018 and May 2020, we included patients who underwent carpal tunnel release, nonsurgical or surgical treatment for thumb-base osteoarthritis, trigger finger release, limited fasciectomy for Dupuytren's contracture, or nonsurgical treatment for midcarpal laxity in one of the 28 centers of Xpert Clinics in the Netherlands. We screened 5859 patients with complete sociodemographics and data at baseline. Thirty-eight percent (2248 of 5859) of these patients had complete data at 3 months. Finally, participants were eligible for inclusion if they provided a relevant answer to the three patient-reported experience measures (PREM) items. A total of 424 patients did not do this because they answered "I don't know" or "not applicable" to a PREM item, leaving 31% (1824 of 5859) for inclusion in the final sample. A validated Satisfaction with Treatment Result Questionnaire was administered at 3 months, which identified the patients' level of satisfaction with treatment results so far on a 5-point Likert scale (research question 1, with answers of good, excellent, poor, moderate, or fair) and the patients' willingness to undergo the treatment again under similar circumstances (research question 2, with answers of yes or no). A hierarchical logistic regression model was used to identify whether baseline sociodemographic, change in outcome (patient-reported outcome measures for quality of life, hand function, and pain), baseline measures of mental health (including

treatment credibility [the extent to which a patient attributes credibility to a treatment] and expectations, illness perception, pain catastrophizing, anxiety and depression), and patient-reported experience measures (PREMs) were associated with each question of the Satisfaction with Treatment Result Questionnaire at 3 months post-treatment. We dichotomized our first question into good and excellent, which were considered more satisfied, and poor, moderate, and fair were considered less satisfied. After dichotomization, 57% (1042 of 1824) of patients were classified as more satisfied with the treatment results.

Results

The following variables were independently associated with satisfaction with treatment results, with an area under the curve of 0.82 (95% confidence interval 0.80 to 0.84) (arranged from the largest to the smallest standardized odds ratio): greater decrease in pain during physical load (SOR 2.52 [95% CI 2.18 to 2.92]; $p < 0.001$), patient's positive experience with the explanation of the pros and cons of the treatment (determined with the question: "Have you been explained the pros and cons of the treatment or surgery?") (SOR 1.83 [95% CI 1.41 to 2.38]; $p < 0.001$), greater improvement in hand function (SOR 1.76 [95% CI 1.54 to 2.01]; $p < 0.001$), patients' positive experience with the advice for at home (determined with the question: "Were you advised by the healthcare providers on how to deal with your illness or complaints in your home situation?") (SOR 1.57 [95% CI 1.21 to 2.04]; $p < 0.001$), patient's better personal control (determined with the question: "How much control do you feel you have over your illness?") (SOR 1.24 [95% CI 1.1 to 1.40]; $p < 0.001$), patient's more positive treatment expectations (SOR 1.23 [95% CI 1.04 to 1.46]; $p = 0.02$), longer expected illness duration by the patient (SOR 1.2 [95% CI 1.04 to 1.37]; $p = 0.01$), a smaller number of symptoms the patient saw as part of the illness (SOR 0.84 [95% CI 0.72 to 0.97]; $p = 0.02$), and less concern about the illness the patient experiences (SOR 0.84 [95% CI 0.72 to 0.99]; $p = 0.04$). For willingness to undergo the treatment again, the following variables were independently associated with an AUC of 0.81 (95% CI 0.78 to 0.83) (arranged from the largest to the smallest standardized OR): patient's positive experience with the information about the pros and cons (determined with the question: "Have you been explained the pros and cons of the treatment or surgery?") (SOR 2.05 [95% CI 1.50 to 2.8]; $p < 0.001$), greater improvement in hand function (SOR 1.80 [95% CI 1.54 to 2.11]; $p < 0.001$), greater decrease in pain during physical load (SOR 1.74 [95% CI 1.48 to 2.07]; $p < 0.001$), patient's positive experience with the advice for at home (determined with the question: "Were you advised by the healthcare providers on how to deal with your illness or complaints in your home situation?") (SOR 1.52 [95% CI 1.11 to 2.07]; $p = 0.01$), patient's positive experience with shared decision-making (determined with the question: "Did you decide together with the care providers which care or treatment you will receive?") (SOR 1.45 [95% CI 1.06 to 1.99]; $p = 0.02$), higher credibility the patient attributes to the treatment (SOR 1.44 [95% CI 1.20 to 1.73]; $p < 0.001$), longer symptom duration (SOR 1.27

[95% CI 1.09 to 1.52]; $p < 0.01$), and patient's better understanding of the condition (SOR 1.17 [95% CI 1.01 to 1.34]; $p = 0.03$).

Conclusion

Our findings suggest that to directly improve satisfaction with treatment results, clinicians might seek to: (1) improve the patient's experience with healthcare delivery, (2) try to influence illness perception, and (3) boost treatment expectations and credibility. Future research should confirm if these suggestions are valid and perhaps also investigate whether satisfaction with treatment results can be predicted (instead of explained, as was done in this study). Such prediction models, as well as other decision support tools that investigate patient-specific needs, may influence experience with healthcare delivery, expectations, or illness perceptions, which in turn may improve satisfaction with treatment results.

INTRODUCTION

Satisfaction with treatment results is an important outcome domain in striving for patient-centered and value-based healthcare. In these frameworks, the patient is central, and the aim is to achieve high value at low cost [1-6]. After all, is there value in a technically perfect surgical procedure, with no complications and excellent objective outcomes afterwards, if the patient is not satisfied with the treatment results? Although recognized as an important outcome domain [7], the interpretation of satisfaction with treatment results is difficult, and there are doubts about the face validity of questionnaires to measure satisfaction with treatment results [8]. However, the Satisfaction with Treatment Result Questionnaire has a good-to-excellent construct validity and a very high test-retest reliability [9], and we believe it is reasonable to use it in a study exploring this topic.

Several studies have investigated factors associated with satisfaction with treatment results [10-30]. Marks et al. [10] found associations between satisfaction and pain and symptoms, activities of daily living or function, aesthetics, embodiment, strength, ROM, fulfillment of expectations, deformity, workers compensation, and length of follow-up. Additionally, strong associations have been found between satisfaction and better patient-reported experience measures (PREMs), such as the provision of general and treatment information, and with physician communication and shared decision-making [11-16]. Furthermore, the relationship with the surgeon, particularly perceived empathy, is a driver of satisfaction with treatment results [17-20]. Associations with several measures of mental health have also been found. For instance, higher preoperative pain catastrophizing is associated with lower satisfaction after hand surgery [21,22], and more symptoms of depression are associated with greater dissatisfaction after carpal tunnel release [23]. There is no consensus on the association between treatment expectations and satisfaction with treatment results; several authors suggested that higher expectations may lead to lower satisfaction [24-26], whereas other studies found a reverse association [13,27-30].

Although the aforementioned studies investigated factors associated with patient satisfaction with treatment results, most studies used relatively small samples or used a univariable approach instead of a multivariable approach. Therefore, the independent association of baseline sociodemographics, quality of life, improvement in pain and function, experiences with healthcare delivery, and baseline measures of mental health with satisfaction with treatment results is still unclear. More knowledge on independent factors that are associated with satisfaction with treatment results may help clinicians to directly improve satisfaction with treatment results, as well as inform future studies aiming to develop interventions that improve satisfaction with treatment results.

Therefore, we asked: (1) What factors are independently associated with satisfaction with treatment results at 3 months post-treatment in patients treated for common hand and wrist conditions? (2) What factors are independently associated with the willingness to undergo

the treatment again at 3 months post-treatment in patients treated for common hand and wrist conditions? Among the studied factors were baseline sociodemographics, quality of life, improvement in pain and function, experiences with healthcare delivery, and baseline measures of mental health.

PATIENTS AND METHODS

Study Design

This was a cohort study using a longitudinally maintained, population-based sample of patients with hand and wrist conditions from the Hand Wrist Study Group cohort, reported following the Strengthening the Reporting of Observational Studies in Epidemiology guidelines [31].

Setting

Data collection using GemsTracker electronic data capture tools (GemsTracker 2020) was part of usual care and occurred between August 2018 and May 2020 at Xpert Clinics. The start date of the current PREM determined the start date of the study. All data were digitally collected using GemsTracker, a secure internet-based application for distributing questionnaires and forms during clinical research and quality registrations. Xpert Clinics comprises 28 clinics for hand surgery and therapy in The Netherlands. Twenty-three surgeons who have been certified by the Federation of European Societies for Surgery of the Hand and more than 150 hand therapists are employed at our treatment centers. At Xpert Clinics, treatment outcomes are evaluated in measurement tracks, each of which consists of treatments with similar relevant outcome domains and timepoints. After a diagnosis is registered during the first consultation, a measurement track is automatically activated, and patient-reported outcome measure forms are emailed to the patient. Details of this procedure have been published [32].

Participants

Participants were eligible for inclusion if they were adults who completed all relevant questionnaires. We included patients who underwent one of the following common treatments: trigger finger release (23% [423 of 1824]), limited fasciectomy (17% [307 of 1824]), trapeziectomy with or without ligament reconstruction tendon interposition for thumb base osteoarthritis (12% [213 of 1824]), carpal tunnel release (29% [521 of 1824]), hand therapy for midcarpal laxity (2% [35 of 1824]), and hand therapy for thumb base osteoarthritis (18% [325 of 1824]) (Table 1). Because the aim of this study was to investigate which factors explain satisfaction with treatment results in a general population of patients treated for hand and wrist disorders, we selected the largest pathology of each of the six

largest measurement tracks from our cohort [32]. Patients who underwent operative treatment were assessed at 3 months postoperatively, and patients who underwent nonoperative treatment were assessed 3 months after treatment was initiated.

We screened 5859 patients with complete sociodemographics and data at baseline. Thirty-eight percent (2248 of 5859) of patients had complete data at 3 months. Finally, participants were eligible for inclusion if they provided a relevant answer to the three PREM items. A total of 424 patients did not do this because they answered "I don't know" or "not applicable" to a PREM item, leaving 31% (1824 of 5859) for inclusion in the final sample (Fig. 1). There were no additional exclusion criteria.

To assess potential selection bias, we compared responder and nonresponder demographics and measures of mental health, using the standardized mean difference as an indication of imbalance [33]. Nonresponders were defined as patients who did not complete questionnaires at 3 months or did not provide a relevant answer to a PREM item. Responders were defined as patients who completed all relevant questionnaires at baseline and at 3 months. Responders and nonresponders both received treatment and remained in care. Besides difference in treatment type (Standardized Mean Difference 0.26), all standardized mean difference values were < 0.2, indicating that the magnitude of the standardized mean difference was even smaller than that defined as small by Cohen [34] (Supplementary Table 1; supplemental materials are available with the online version of *CORR*®). Additionally, we conducted a Little test (p = 0.27), which supported the idea that nonresponders could be considered missing at random [35-37].

Variables and Measurements

The primary outcomes in this study were the two questions of the Satisfaction with Treatment Result Questionnaire at 3 months after the start of treatment. The first question evaluates patients' satisfaction with treatment results on a 5-point Likert scale (answering options: poor, moderate, fair, good, and excellent). In the second question, patients indicated whether they would undergo the same procedure again under similar circumstances (yes or no). The Satisfaction with Treatment Result Questionnaire has a good-to-excellent construct validity and very high test-retest validity [9].

We classified variables we investigated as potentially associated with satisfaction into four categories: sociodemographic, clinical patient-reported outcome measures, measures of mental health, and PREMs.

Sociodemographic characteristics included sex (we report sex, not gender, as our data comes from the Dutch Citizen Service Administration, so we did not want to make any unsupported assumptions on gender), age, occupational status (unemployed or light, medium, or heavy physical labor), whether the patient visited the clinic for a second opinion, self-reported duration of symptoms (in months), whether the dominant hand was treated,

and whether the disease was recurrent (measured by the question: "Have been treated for the same disease before?"; the answer yes would be coded as recurrent).

Clinical patient-reported outcome measures included the change in patient-reported outcome measures for pain and hand function between baseline and 3 months, and health-related quality of life at 3 months. We used a VAS score (range 0 to 100) to measure pain during physical load (higher scores indicate more pain) and hand function (higher scores indicate better function). The VAS is a validated and widely used tool for measuring these constructs [38]. Although we also used more disease-specific questionnaires in our cohort (such as the Boston Carpal Tunnel Questionnaire, Patient-Rated Wrist/Hand Evaluation, and Michigan Hand outcomes Questionnaire), these differed among the treatments in our study sample and therefore are less well-suited to use for the current research question aiming at all patients with the most common hand and wrist conditions.

We measured health-related quality of life using the VAS of the EuroQol-5 Dimensions self-rated health questionnaire as an indication of the overall perceived health status (range 0 to 100; higher scores indicate better perceived health) [39].

To measure the patients' experience with healthcare delivery (which is different from satisfaction with treatment results [40]), we used the PREM questionnaire, based on the Consumer Quality Index, which is widely used in private practice clinics in the Netherlands [41]. The 11 items evaluate the patients' experience with healthcare delivery using a 5-point Likert scale (with answers ranging from no, not at all to yes, completely). Of this questionnaire, we only included three items because in the other items, ceiling effects were present that did not allow us to run our models. These items were experience with the explanation about the pros and cons of the treatment, experience with shared decision-making, and experience with the advice for at home.

Measures of mental health included anxiety and depression, pain catastrophizing, illness perceptions, and expectations. Anxiety and depression were measured with the Patient Health Questionnaire-4 (higher scores indicate more anxiety and depression) and pain catastrophizing was measured with the Pain Catastrophizing Scale (higher scores indicate a higher amount of catastrophizing). Illness perception was measured with the Brief Illness Perception Questionnaire [42,43]. The Brief Illness Perception Questionnaire measures how patients perceive their illness across eight domains (consequences, timeline, personal control, treatment control, identity, concern, coherence, and emotional response). Each domain is assessed with a single question (higher scores indicate more negative illness perceptions except for personal control, treatment control, and coherence) [44]. We excluded the domain of treatment control ("How much do you think your treatment can help your illness?") from our analyses because we considered that item to have a strong conceptual overlap with the expectancy subscale of the Credibility/Expectancy questionnaire. Treatment outcome expectations were measured with the Credibility/Expectancy

questionnaire [45]. The credibility subscale consists of three items measuring the credibility that the patient attributes to the treatment. A higher score reflects a higher attribution of credibility to a treatment. The expectancy subscale consists of three items measuring the expected magnitude of improvement because of the prescribed treatment. A higher score reflects a more positive treatment outcome expectation.

Ethical Approval

We obtained ethical approval for this study from Erasmus MC, Rotterdam, the Netherlands (MEC-2018-1088). Written informed consent was obtained from all patients.

Statistical Methods and Study Size

We dichotomized our outcome of satisfaction with treatment results into poor, moderate and fair as less satisfied, and good and excellent as more satisfied. After dichotomization, 57% (1042 of 1824) of participants were classified as more satisfied with the treatment results (19% [349 of 1824] answered excellent and 38% [693 of 1824] answered good), and 43% (782 of 1824) of patients were classified as less satisfied with the treatment results (26% [472 of 1824] reported their satisfaction was fair, 13% [231 of 1824] reported that it was moderate, and 4% [79 of 1824] reported that it was poor) (Fig. 2). This is comparable with other findings in our population [27,46-49]. Similarly, to further account for ceiling effects, we dichotomized the PREM items into negative experience (answering options: no, not at all, a little bit, partly, and mostly) and positive experience (answering option: yes, completely). The items used in the final analysis were: "Did you decide together with the care providers which care or treatment you will receive?" (hereinafter referred to as shared decision-making), "Have you been explained the pros and cons of the treatment or surgery?" (henceforth referred to as pros and cons), and "Were you advised by the healthcare providers on how to deal with your illness or complaints in your home situation?" (hereafter referred to as advice).

Because this study evaluated a diverse population of patients with common hand and wrist conditions, we adjusted for the type of treatment in the analyses. By adjusting for treatment in our analysis, we accounted for a potential influence of treatment on satisfaction with treatment results. To test the association of specific patient characteristics with satisfaction, we performed a hierarchical logistic regression analysis. In this hierarchical regression analysis, a set of variables is entered in a specific sequence to illustrate the added amount of explained variance of each set. In the first model, sociodemographic patient characteristics were entered, including age, sex, symptom duration, treatment side, dominance, type of work, and second-opinion visit. In the second step, we added clinical patient-reported outcome measures, including the EuroQol-5 Dimensions VAS self-rated health at 3 months, the change in VAS pain score during physical load, and VAS function score between baseline and 3 months. In the third step, we added the three items of the

PREM: shared decision-making, pros and cons, and advice. In the fourth step, we added measures of mental health, including the Brief Illness Perception Questionnaire items of consequences, timeline, personal control, identity, concern, coherence, emotional response, Patient Health Questionnaire anxiety and depression subscales, Pain Catastrophizing Scale, and Credibility/Expectancy Questionnaire subscales. An advantage of a hierarchical multivariable model is that by entering the next set of variables, certain variables might be pushed out of significance because variables may have shared variance. Therefore, in the most definitive multivariable model, only the variables that truly explain variance in the dependent variable remain. To account for potential strong correlations and multiple variables measuring the same construct, we evaluated multicollinearity using a correlation matrix (Supplementary Table 2; supplemental materials are available with the online version of *CORR®*) and variance inflation factor (Supplementary Table 3; supplemental materials are available with the online version of *CORR®*). A correlation coefficient of the Spearman rho greater than 0.7 was considered a strong correlation. A variance inflation factor greater than 3 was considered an indication of multicollinearity [50]. Based on the variance inflation factor (the highest variance inflation factor = 2.2) and the correlation matrix (highest Spearman rho = 6.8, which is only a moderate correlation), we did not find any indication for multicollinearity in the hierarchical logistic regression model. To illustrate the goodness of fit of the different models, we determined the area under the curve, the Nagelkerke r^2, and the receiver operating characteristic curves for each model.

With 1824 patients, 33 variables, an alpha of 0.05, and a conventional small effect size f^2 of 0.02, this study had a power of 95%. We additionally computed univariable associations between all variables. In addition to odds ratios, we reported standardized ORs by converting them to the same scale [51]. The nonstandardized odds ratios in our most definitive model indicate that with every unit increase in either a continuous, dichotomous, or categorical independent variable, the odds of being satisfied with the treatment results or being willing to undergo the treatment again increase or decrease by the value of the nonstandardized OR. Standardized ORs were converted to the same scale, which made it easier to make between-variable comparisons and determine the relative association of each explanatory variable.

All analyses were performed using R Statistical Programming, version 3.3.4 (R Project for Statistical Computing). For all tests, a p value < 0.05 was considered statistically significant.

RESULTS

Satisfaction with Treatment Results

In our most definitive model, we found an area under the curve of 0.82 (Table 2), indicating an excellent ability to distinguish more satisfied from less satisfied patients [52]. Satisfaction

with the treatment results was associated with the following variables (arranged from the largest to the smallest standardized OR): greater decrease in pain during physical load (SOR 2.52 [95% CI 2.18 to 2.92]; $p < 0.001$), patient's positive experience with the explanation of the pros and cons of the treatment (determined with the question: "Have you been explained the pros and cons of the treatment or surgery?") (SOR 1.83 [95% CI 1.41 to 2.38]; $p < 0.001$), greater improvement in hand function (SOR 1.76 [95% CI 1.54 to 2.01]; $p < 0.001$), patients' positive experience with the advice for at home (determined with the question: "Were you advised by the healthcare providers on how to deal with your illness or complaints in your home situation?") (SOR 1.57 [95% CI 1.21 to 2.04]; $p < 0.001$), patient's better personal control (determined with the question: "How much control do you feel you have over your illness?") (SOR 1.24 [95% CI 1.10 to 1.40]; $p < 0.001$), patient's more positive treatment expectations (SOR 1.23 [95% CI 1.04 to 1.46]; $p = 0.02$), longer expected illness duration by the patient (SOR 1.20 [95% CI 1.04 to 1.37]; $p = 0.01$), a smaller number of symptoms the patient saw as part of the illness (SOR 0.84 [95% CI 0.72 to 0.97]; $p = 0.02$), and less concern about the illness the patient experiences (SOR 0.84 [95% CI 0.72 to 0.99]; $p = 0.04$) (Fig. 3). When analyzing the separate steps of the different models, sociodemographics alone provided an area under the curve (AUC) of 0.60 (95% CI 0.57 to 0.62) for the level of satisfaction with treatment results. When adding clinical characteristics, the AUC was 0.79 (95% CI 0.77 to 0.81). This further increased to 0.81 (95% CI 0.79 to 0.81) when adding PREMs, and finally, the AUC increased to 0.82 (95% CI 0.80 to 0.84) for the level of satisfaction with treatment results (Fig. 4).

Analyzing differences in variables between the different steps of the model for satisfaction with treatment results, we found that there were two differences (Supplementary Table 4; supplemental materials are available with the online version of *CORR*®). First, in Model 1, recurrence (determined with the question: "Have you been treated for the same disease before?") was associated with a smaller probability of being satisfied with the treatment results (standardized OR 0.70 [95% CI 0.50 to 1.00]), but after adding the clinical patient-reported outcome measures in Model 2, there was no association. This implies that a different change in patient-reported outcome measure score has a shared variance with recurrence and pushes recurrence out of significance. This means that a different change in patient-reported outcome measure score is the stronger variable. Second, a higher EuroQol-5 Dimensions self-rated health score was associated with a larger probability of being satisfied with the treatment results in Model 2 (standardized OR 1.32 [95% CI 1.18 to 1.48) and Model 3 (standardized OR 1.29 [95% CI 1.15 to 1.45]). However, after adding measures of mental health and treatment expectations in Model 4, we found that the EuroQol-5 Dimensions self-rated health score was no longer associated, and several illness perception items and more positive expectations became associated with being satisfied with the treatment results. This finding suggests that EuroQol-5 Dimensions self-rated health has shared variance with specific measures of mental health, such as

illness perception. This means that the mental health measures are the stronger variables (Supplementary Table 5; supplemental materials are available with the online version of *CORR®*).

Willingness to Undergo the Treatment Again

In our most definitive model, we found an area under the curve of 0.81 (Table 3), indicating an excellent ability to distinguish patients that would be willing to undergo the treatment again from patients that would not [52]. Being willing to undergo the treatment again was associated with the following variables (arranged from the largest to the smallest standardized OR): patient's positive experience with the information about the pros and cons (determined with the question: "Have you been explained the pros and cons of the treatment or surgery?") (SOR 2.05 [95% CI 1.50 to 2.8]; $p < 0.001$), greater improvement in hand function (SOR 1.80 [95% CI 1.54 to 2.11]; $p < 0.001$), greater decrease in pain during physical load (SOR 1.74 [95% CI 1.48 to 2.07]; $p < 0.001$), patient's positive experience with the advice for at home (determined with the question: "Were you advised by the healthcare providers on how to deal with your illness or complaints in your home situation?") (SOR 1.52 [95% CI 1.11 to 2.07]; $p = 0.01$), patient's positive experience with shared decision-making (determined with the question: "Did you decide together with the care providers which care or treatment you will receive?") (SOR 1.45 [95% CI 1.06 to 1.99]; $p = 0.02$), higher credibility the patient attributes to the treatment (SOR 1.44 [95% CI 1.20 to 1.73]; $p < 0.001$), longer symptom duration (SOR 1.27 [95% CI 1.09 to 1.52]; $p < 0.01$), and patient's better understanding of the condition (SOR 1.17 [95% CI 1.01 to 1.34]; $p = 0.03$) (Fig. 5).

For the willingness to undergo treatment again, sociodemographics alone provided an AUC of 0.58 (95% CI 0.55 to 0.62). When adding clinical characteristics, the AUC was 0.75 (95% CI 0.72 to 0.78). This further increased to 0.79 (95% CI 0.77 to 0.82) when adding PREMs, and finally, the AUC was 0.81 (95% CI 0.78 to 0.83) for the willingness to undergo treatment again after adding measures of mental health (Fig. 6).

DISCUSSION

In the framework of patient-centered and value-based healthcare, satisfaction with treatment results is an important outcome domain. Before our study, it was unclear which factors were independently associated with satisfaction with treatment results and with a willingness to undergo the treatment again. We found a high explained variance in our models. The following variables were independently associated with satisfaction in either or both models: greater decrease in pain during physical load, patient's positive experience with the explanation of the pros and cons of the treatment, positive experience with the advice for at home (determined with the question: "Were you advised by the healthcare

providers on how to deal with your illness or complaints in your home situation?"), patient's positive experience with shared decision-making, higher credibility the patients attributes to the treatment, longer symptom duration, better personal control (determined with the question: "How much control do you feel you have over your illness"), patient's more positive treatment expectations, longer expected illness duration by the patient, patient's better understanding of the condition, a smaller number of symptoms the patient sees as part of the illness, and less concern about the illness the patient experiences. Many of these variables may be guided and can be used directly in daily clinic or in studies that develop interventions to improve satisfaction with treatment results.

Limitations

Whereas an advantage of our observational study design is its representation of daily practice, a limitation of the observational design is that a substantial proportion of patients did not respond. However, the nonresponder analysis did not show substantial differences, and the Little test strongly suggests that the data were missing at random. Therefore, we are confident that the high percentage of nonresponders did not influence our results.

A second limitation is the follow-up time in our study. We chose this timepoint because follow-up measurements for the PREM were only obtained at 3 months. As a result, the more extensive surgical treatments may not have reached their endpoint yet, and evaluating satisfaction with treatment results may be too soon at this timepoint. However, theoretically, this should not influence factors explaining variance in satisfaction with treatment results. In fact, there might be more variation in satisfaction with treatment results at 3 months, which may yield better results. Nevertheless, future studies might investigate different timepoints.

Another limitation is the variety of treatment types in our study. Combining different treatment types may have led to dilution of the results because certain variables might interact with the treatment type. However, we aimed to investigate which factors explain satisfaction with treatment results in a general population of patients treated for hand and wrist disorders. Therefore, we selected the most commonly used treatment type in each of the six largest measurement tracks from our cohort and adjusted for the treatment type in our models. By adjusting for the treatment type in our analysis, a potential influence of treatment type on satisfaction with treatment results is accounted for, and the remaining significant variables are independent of treatment type in the final hierarchical model. Therefore, these remaining variables can be generalized to a broader population of patients with hand and wrist conditions. The standardized mean difference between the treatment types was small. This further strengthens the generalizability of our study findings, perhaps even to patients with other musculoskeletal conditions such as hip osteoarthritis. However, future studies should validate our findings in other populations.

Additionally, because satisfaction with treatment results is a multidimensional construct, there are still doubts about the validity of instruments measuring this domain [8,53]. Although the Satisfaction with Treatment Result Questionnaire has a good-to-excellent construct validity and a very high test-retest reliability, future studies should further investigate its face validity.

Finally, we found a very high proportion of the finding explained by the variables in our model. An explanation for the little unexplained variance may be that we did not include all relevant variables in our models, such as additional aspects of experiences with healthcare delivery, coping strategies, goal attainment, the occurrence of complications, personal injury lawsuits, social health, or the specific course of rehabilitation. Additionally, our dichotomization may be a reason for unexplained variance, although this also has added value because our model thereby distinguishes between more satisfied and less satisfied patients. Moreover, although the Brief Illness Perception Questionnaire and Patient Health Questionnaire are valid tools, they might be interpreted differently by individuals, and they function as screening tools and lack the conceptual depth of more extensive questionnaires. Because satisfaction with treatment results is a complex domain, using more comprehensive measures of mental health may yield an even larger proportion of explained variance. Future studies might include these variables when investigating satisfaction with treatment results.

Discussion of Key Findings

Interestingly, all three included PREM items (positive experience with the explanation of the pros and cons; advice for how to deal with the complaints at home; and shared decision making) were associated with one or both of the Satisfaction with Treatment Result Questionnaire questions (which were: Are you satisfied with the treatment result so far? And, would you be willing to undergo the treatment again under similar circumstances?). These findings confirm that the patients' experience with healthcare delivery is associated with their satisfaction with the result. Based on these findings, healthcare providers may try to improve the experience with healthcare delivery, that is, by always explaining the pros and cons of a treatment and by providing adequate advice on how to deal with the complaints at home (such as by sending e-mails with treatment-specific information and educational movies). Also, healthcare providers may strive for better shared-decision making. Future research should inform if this will indeed improve satisfaction with treatment results.

In contrast to previous studies [8,20,23,53,54], depression was not associated with satisfaction in our most-definitive model. However, we did find a univariable association. This suggests that depression has a shared variance with other variables in our models; for example, other mental health items, such as the Illness Perception Questionnaire item of emotional response. Similarly, we did not find an association with pain catastrophizing, while

other studies did [20-22,55]. No other study on this topic that we know of has investigated the association of depression or pain catastrophizing in combination with illness perception, which may explain why our findings are different from those reported by others.

Another interesting finding here was that a higher score on the Credibility/Expectancy Questionnaire expectancy subscale (the more positive expectations a patient has of a treatment) was associated with better satisfaction with treatment results. This is especially noteworthy because several studies have suggested that clinicians ought to try to work to temper patients' expectations [24-26,56], and many surgeons believe that it is important to help patients to cultivate reasonable expectations before surgery. By contrast, several other studies have suggested that boosting expectations is associated with better outcomes [13,27-30,57]. Our findings support the latter suggestion. Related to this, the credibility subscale (the extent to which a patient attributes credibility to a treatment) was associated with the patient's willingness to undergo the treatment again. To our knowledge, no other studies have investigated factors explaining this willingness to undergo treatment again, but it seems sensible that someone who does not find a treatment credible may be less willing to undergo that treatment again. Hence, it might be helpful to investigate possible interventions to boost expectations and improve the credibility of specific treatments.

A possible intervention to influence the experience with healthcare delivery, expectations, and illness perception may, for example, be the creation of a decision-support tool to specifically investigate the patients' needs for the clinician to respond accordingly. Further, future studies should investigate whether satisfaction with treatment results can be predicted (instead of explained, such as in this study), so that a prediction model could be used as a decision tool and to show what outcomes the patient may expect. Another option is to provide more personalized information relevant to the patient, such as emailing treatment-specific pros and cons. Additionally, to influence illness perception, future studies might investigate the effect of discussing illness perceptions and expectations during the first consultation. However, these suggestions are all hypothetical and future research should investigate their added value.

CONCLUSIONS

We identified several influenceable factors independently associated with satisfaction with treatment results. To directly improve satisfaction with treatment results, clinicians might seek to: (1) improve the patient's experience with healthcare delivery, (2) try to influence illness perception, and (3) boost treatment expectations and credibility. However, these recommendations are all hypothetical, and future research should investigate their added value. Moreover, future studies should investigate whether satisfaction with treatment results can be predicted (instead of explained, as was done in this study), so that a prediction

model could be used as a decision-support tool that may inform shared-decision making and expectation management. Also, decision-support tools that investigate patient-specific needs may positively influence experience with healthcare delivery, expectations, and illness perception, which in turn may improve satisfaction with treatment results.

ACKNOWLEDGMENTS

We thank all the patients who completed questionnaires as part of their clinical care and who agreed that their data could be used anonymously for the present study. We thank the caregivers, and other personnel of Xpert Clinics and Equipe Zorgbedrijven for assisting in the routine outcome measurements that are the basis for this study.

GROUP AUTHORS

Members of the Hand-Wrist Study Group include: Dirk-Johannes Jacobus Cornelis van der Avoort MD, Richard Arjen Michiel Blomme MD, Herman Luitzen de Boer MD, Gijs Marijn van Couwelaar MD, Jan Debeij MD PhD, Jak Dekker MSc, Alexandra Fink PT, Klazina Paulina de Haas MD, Kennard Harmsen MD, Steven Eric Ruden Hovius MD PhD, Rob van Huis PT, Richard Koch MD, Alexander Kroeze MD, Thybout Matthias Moojen MD PhD, Mark Johannes Willem van der Oest PhD, Pierre-Yves Alain Adriaan Pennehouat PT, Willemijn Anna de Ridder PT MSc, Johannes Pieter de Schipper MD, Karin Schoneveld PT MSc, Berbel Jeannee Rinel Sluijter MD PhD, Jeronimus Maria Smit MD PhD, Xander Smit MD PhD, John Sebastiaan Souer MD PhD, Marloes Hendrina Paulina ter Stege MSc, Johannes Frederikes Maria Temming MD, Joris Sebastiaan Teunissen BSc, Jeroen Hein van Uchelen MD PhD, Joris Jan Veltkamp PT, Guus Maarten Vermeulen MD PhD, Erik Taco Walbeehm MD PhD, Oliver Theodor Zöphel MD PhD, Jelle Michiel Zuidam.

REFERENCES

1. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. *N Engl J Med.* 2012;366(9):780-781.

2. Basch E. Patient-Reported Outcomes - Harnessing Patients' Voices to Improve Clinical Care. *N Engl J Med.* 2017;376(2):105-108.

3. Porter ME. What is value in health care? *N Engl J Med.* 2010;363(26):2477-2481.

4. Goldhahn J, Angst F, Simmen BR. What counts: outcome assessment after distal radius fractures in aged patients. *J Orthop Trauma.* 2008;22(8 Suppl):S126-130.

5. Karnezis IA, Fragkiadakis EG. Association between objective clinical variables and patient-rated disability of the wrist. *J Bone Joint Surg Br.* 2002;84(7):967-970.

6. Rathert C, Wyrwich MD, Boren SA. Patient-centered care and outcomes: a systematic review of the literature. *Med Care Res Rev.* 2013;70(4):351-379.

7. Wouters RM, Jobi-Odeneye AO, de la Torre A, Joseph A, Hovius SER. A Standard Set for Outcome Measurement in Patients With Hand and Wrist Conditions: Consensus by the International Consortium for Health Outcomes Measurement Hand and Wrist Working Group. *J Hand Surg Am.* 2021;46(10):841-855.e847.

8. Ring D, Leopold SS. Editorial-Measuring Satisfaction: Can It Be Done? *Clin Orthop Relat Res.* 2015;473(10):3071-3073.

9. De Ridder WA, van Kooij YE, Vermeulen GM, Slijper HP, Selles RW, Wouters RM. Test-retest Reliability and Construct Validity of the Satisfaction with Treatment Result Questionnaire in Patients with Hand and Wrist Conditions: A Prospective Study. *Clin Orthop Relat Res.* 2021;479(9):2022-2032.

10. Marks M, Herren DB, Vliet Vlieland TPM, Simmen BR, Angst F, Goldhahn J. Determinants of patient satisfaction after orthopedic interventions to the hand: A review of the literature. *J Hand Ther.* 2011;24(4):303-312.

11. Tsehaie J, van der Oest MJW, Poelstra R, Selles RW, Feitz R, Slijper HP, Hovius SER, Porsius JT. Positive experience with treatment is associated with better surgical outcome in trapeziometacarpal osteoarthritis. *J Hand Surg Eur Vol.* 2019;44(7):714-721.

12. Mandl LA, Galvin DH, Bosch JP, George CC, Simmons BP, Axt TS, Fossel AH, Katz JN. Metacarpophalangeal arthroplasty in rheumatoid arthritis: What determines satisfaction with surgery? *J Rheumatol.* 2002;29(12):2488-2491.

13. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. *Lancet.* 2001;357(9258):757-762.

14. Parrish RC, Menendez ME, Mudgal CS, Jupiter JB, Chen NC, Ring D. Patient Satisfaction and its Relation to Perceived Visit Duration with a Hand Surgeon. *J Hand Surg (USA).* 2016;41(2):257-262e254.

15. Bot AG, Bossen JK, Herndon JH, Ruchelsman DE, Ring D, Vranceanu AM. Informed shared decision-making and patient satisfaction. *Psychosomatics.* 2014;55(6):586-594.

16. Bernstein J, Kupperman E, Kandel LA, Ahn J. Shared Decision Making, Fast and Slow: Implications for Informed Consent, Resource Utilization, and Patient Satisfaction in Orthopaedic Surgery. *J Am Acad Orthop Surg.* 2016;24(7):495-502.

17. Pollak KI, Alexander SC, Tulsky JA, Lyna P, Coffman CJ, Dolor RJ, Gulbrandsen P, Ostbye T. Physician empathy and listening: associations with patient satisfaction and autonomy. *J Am Board Fam Med*. 2011;24(6):665-672.

18. Kim SS, Kaplowitz S, Johnston MV. The effects of physician empathy on patient satisfaction and compliance. *Eval Health Prof.* 2004;27(3):237-251.

19. Menendez ME, Chen NC, Mudgal CS, Jupiter JB, Ring D. Physician Empathy as a Driver of Hand Surgery Patient Satisfaction. *J Hand Surg (USA).* 2015;40(9):1860-1865.

20. Vranceanu AM, Ring D. Factors associated with patient satisfaction. *J Hand Surg (USA).* 2011;36(9):1504-1508.

21. Mosegaard SB, Stilling M, Hansen TB. Higher preoperative pain catastrophizing increases the risk of low patient reported satisfaction after carpal tunnel release: a prospective study. *BMC Musculoskelet Disord.* 2020;21(1):42.

22. Mosegaard SB, Stilling M, Hansen TB. Pain Catastrophizing Scale as a predictor of low postoperative satisfaction after hand surgery. *J Orthop.* 2020;21:245-248.

23. Lozano Calderón SA, Paiva A, Ring D. Patient Satisfaction After Open Carpal Tunnel Release Correlates With Depression. *J Hand Surg (USA).* 2008;33(3):303-307.

24. Marks M, Audigé L, Reissner L, Herren DB, Schindele S, Vliet Vlieland TPM. Determinants of patient satisfaction after surgery or corticosteroid injection for trapeziometacarpal osteoarthritis: Results of a prospective cohort study. *Arch Orthop Trauma Surg.* 2015;135(1):141-147.

25. Kadzielski J, Malhotra LR, Zurakowski D, Lee SGP, Jupiter JB, Ring D. Evaluation of Preoperative Expectations and Patient Satisfaction After Carpal Tunnel Release. *J Hand Surg (USA).* 2008;33(10):1783-1788.

26. Frouzakis R, Herren DB, Marks M. Evaluation of expectations and expectation fulfillment in patients treated for trapeziometacarpal osteoarthritis. *J Hand Surg Am*. 2015;40(3):483-490.

27. Hoogendam L, van der Oest MJW, Wouters RM, Andrinopoulou ER, Vermeulen GM, Slijper HP, Porsius JT, Selles RW. Patients With Higher Treatment Outcome Expectations Are More Satisfied With the Results of Nonoperative Treatment for Thumb Base Osteoarthritis: A Cohort Study. *Arch Phys Med Rehabil.* 2021;102(8):1533-1540.

28. Mandl LA, Burke FD, Shaw Wilgis EF, Lyman S, Katz JN, Chung KC. Could preoperative preferences and expectations influence surgical decision making? Rheumatoid arthritis patients contemplating metacarpophalangeal joint arthroplasty. *Plast Reconstr Surg.* 2008;121(1):175-180.

29. Swarup I, Henn CM, Gulotta LV, Henn RF, 3rd. Patient expectations and satisfaction in orthopaedic surgery: A review of the literature. *J Clin Orthop Trauma*. 2019;10(4):755-760.

30. Poelstra R, Selles RW, Slijper HP, van der Oest MJW, Feitz R, Hovius SER, Porsius JT. Better patients' treatment experiences are associated with better postoperative results in Dupuytren's disease. *J Hand Surg Eur Vol.* 2018;43(8):848-854.

31. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453-1457.

32. Selles RW, Wouters RM, Poelstra R, van der Oest MJW, Porsius JT, Hovius SER, Moojen TM, van Kooij Y, Pennehouat PY, van Huis R, Vermeulen GM, Feitz R, Slijper HP. Routine Health Outcome Measurement: Development, Design, and Implementation of the Hand and Wrist Cohort. *Plast Reconstr Surg.* 2020;146(2):343-354.

33. Faraone SV. Interpreting estimates of treatment effects: implications for managed care. *P T.* 2008;33(12):700-711.

34. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers; 1988.

35. de Groot JA, Janssen KJ, Zwinderman AH, Bossuyt PM, Reitsma JB, Moons KG. Correcting for partial verification bias: a comparison of methods. *Ann Epidemiol.* 2011;21(2):139-148.

36. Little RJA. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association.* 1988;83(404):1198-1202.

37. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA, Neaton JD, Rotnitzky A, Scharfstein D, Shih WJ, Siegel JP, Stern H. The prevention and treatment of missing data in clinical trials. *N Engl J Med.* 2012;367(14):1355-1360.

38. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res (Hoboken).* 2011;63 Suppl 11:S240-252.

39. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L Better Than EQ-5D-3L? A Head-to-Head Comparison of Descriptive Systems and Value Sets from Seven Countries. *Pharmacoeconomics.* 2018;36(6):675-697.

40. York Health Economics Consortium. Patient-Reported Experience Measure (PREM) [online],. https://www.yhec.co.uk/glossary/patient-reported-experience-measure-prem/. Published 2016. Accessed November 30, 2021.

41. Delnoij DM, Rademakers JJ, Groenewegen PP. The Dutch consumer quality index: an example of stakeholder involvement in indicator development. *BMC Health Serv Res.* 2010;10:88.

42. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. *J Psychosom Res.* 2006;60(6):631-637.

43. de Raaij EJ, Schröder C, Maissan FJ, Pool JJ, Wittink H. Cross-cultural adaptation and measurement properties of the Brief Illness Perception Questionnaire-Dutch Language Version. *Man Ther.* 2012;17(4):330-335.

44. Kroenke K, Spitzer RL, Williams JB, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics.* 2009;50(6):613-621.

45. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry.* 2000;31(2):73-86.

46. Wouters RM, Slijper HP, Esteban Lopez L, Hovius SER, Selles RW. Beneficial Effects of Nonsurgical Treatment for Symptomatic Thumb Carpometacarpal Instability in Clinical Practice: A Cohort Study. *Arch Phys Med Rehabil.* 2020;101(3):434-441.

47. Wouters RM, Tsehaie J, Slijper HP, Hovius SER, Feitz R, Blomme A, Sluijter B, Schouten C, van der Avoort DJ, Walbeehm E, van Couwelaar G, Vermeulen G, Schipper H, Temming H, van Uchelen J, de Boer L, de Haas N, Zöphel O, Souer S, Hovius S, Moojen T, Smit X, van Huis R, Pennehouat PY, Schoneveld K, van Kooij Y, Wouters R, Zagt P, van Ewijk F, Moussault F, van Houwelingen R, Veltkamp J, Velde AT, Fink A, Slijper H, Selles R, Porsius J, Spekreijse K, Zhou C, Poelstra R, Janssen M, van der Oest M, Evers S, Dekker J, de Jong M, van Gestel J, Stege MT, Dekker M, Faber R, Santegoets F, Sieber-Rasch M, Gerritsen T, Selles RW. Exercise Therapy in Addition to an Orthosis Reduces Pain More Than an Orthosis Alone in Patients With Thumb Base Osteoarthritis: A Propensity Score Matching Study. *Arch Phys Med Rehabil.* 2019;100(6):1050-1060.

48. Tsehaie J, Spekreijse KR, Wouters RM, Slijper HP, Feitz R, Hovius SER, Selles RW. Outcome of a Hand Orthosis and Hand Therapy for Carpometacarpal Osteoarthritis in Daily Practice: A Prospective Cohort Study. *J Hand Surg Am.* 2018;43(11):1000-1009.e1001.

49. Zhou C, Hovius SER, Pieters AJ, Slijper HP, Feitz R, Selles RW. Comparative Effectiveness of Needle Aponeurotomy and Collagenase Injection for Dupuytren's Contracture: A Multicenter Study. *Plast Reconstr Surg Glob Open.* 2017;5(9):e1425.

50. Kutner MHN, C.J., Neter, J., Li, W. *Applied linear statistical models.* Boston: McGraw-Hill; 2005.

51. Sjölander A. Regression standardization with the R package stdReg. *Eur J Epidemiol.* 2016;31(6):563-574.

52. Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression.* Vol 398: John Wiley & Sons; 2013.

53. Sraj SA. CORR Insights®: Test-retest Reliability and Construct Validity of the Satisfaction with Treatment Result Questionnaire in Patients with Hand and Wrist Conditions: A Prospective Study. *Clin Orthop Relat Res.* 2021;479(9):2033-2035.

54. Ring D, Kadzielski J, Fabian L, Zurakowski D, Malhotra LR, Jupiter JB. Self-reported upper extremity health status correlates with depression. *J Bone Joint Surg Am.* 2006;88(9):1983-1988.

55. Ring D, Kadzielski J, Malhotra L, Lee SG, Jupiter JB. Psychological factors associated with idiopathic arm pain. *J Bone Joint Surg Am.* 2005;87(2):374-380.

56. Waljee JF, Chung KC. Commentary regarding "Evaluation of expectations and expectation fulfillment in patients treated for trapeziometacarpal osteoarthritis". *J Hand Surg Am.* 2015;40(3):491-492.

57. Peerdeman KJ, van Laarhoven AIM, Keij SM, Vase L, Rovers MM, Peters ML, Evers AWM. Relieving patients' pain with expectation interventions: a meta-analysis. *Pain.* 2016;157(6):1179-1191.

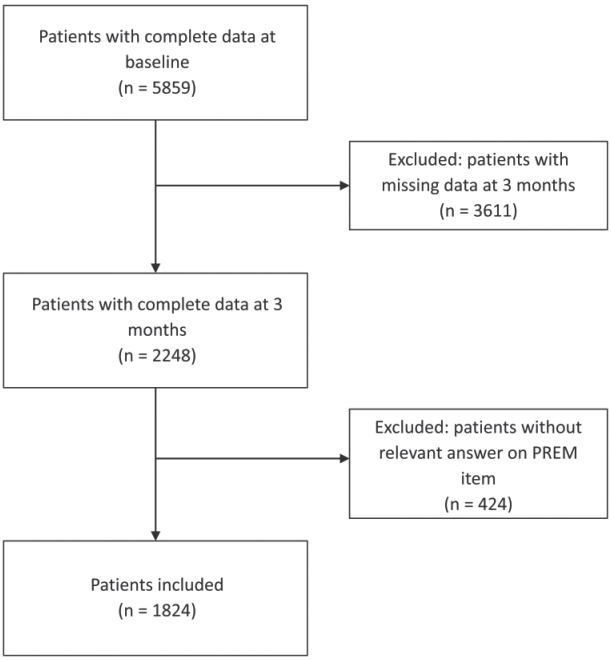


Figure 1. This flowchart illustrates the patient selection for this study.

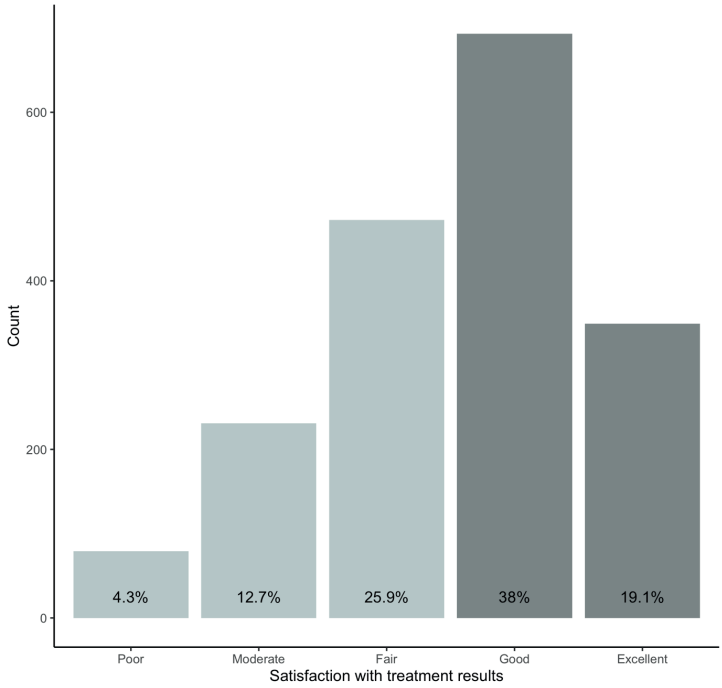


Figure 2. This graph shows the distribution of satisfaction with treatment results at 3 months, before and after dichotomization. Light grey indicates patients who are less satisfied; dark grey indicates those who are more satisfied.

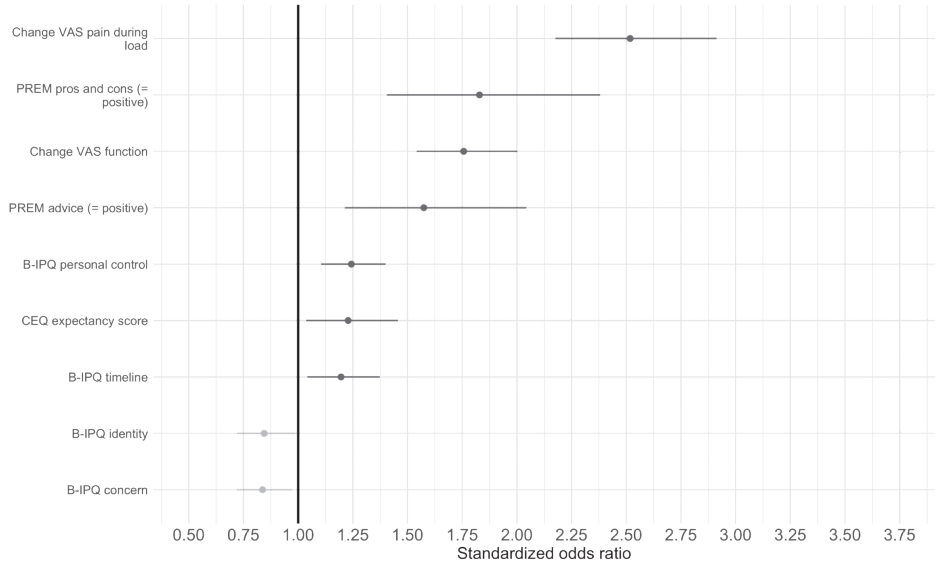


Figure 3. This figure shows the standardized ORs of the associated variables for patient satisfaction with treatment results. Positive associations are shown in dark grey; negative associations are shown in light grey; PREM = patient-reported experience measures; CEQ = Credibility/Expectancy Questionnaire; B-IPQ = Brief Illness Perception Questionnaire.

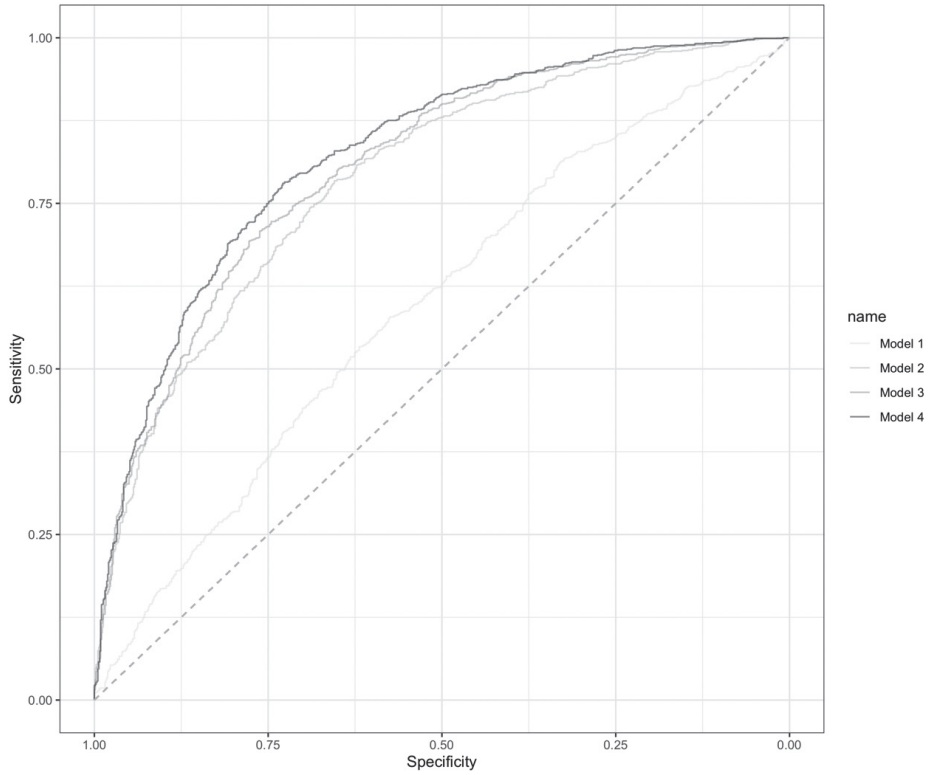


Figure 4. This graph shows the area under the receiver operating characteristic curve for all models explaining the level of satisfaction with treatment results, using the 5-point Likert scale (question 1). The dashed line indicates a discriminative ability of 0.50. Model 1, including sociodemographics, had an AUC of 0.60, and Model 2, after adding clinical patient-reported outcome measures, had an AUC of 0.79. Model 3, after adding PREMs, had an AUC of 0.81, and after adding measures of mental health, the most definitive model had an AUC of 0.82.

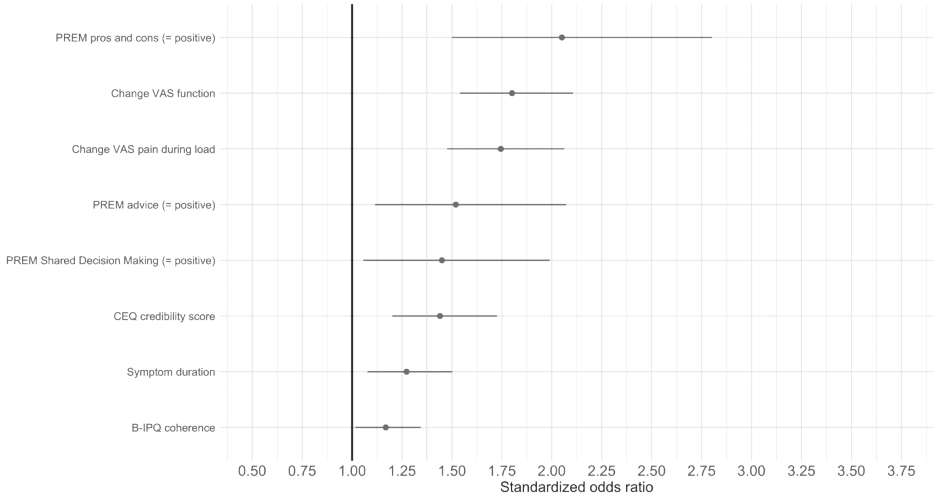


Figure 5. This figure shows standardized ORs of the associated variables for the patient's willingness to undergo the treatment again; PREM = patient-reported experience measures; CEQ = Credibility/Expectancy Questionnaire; B-IPQ = Brief Illness Perception Questionnaire.

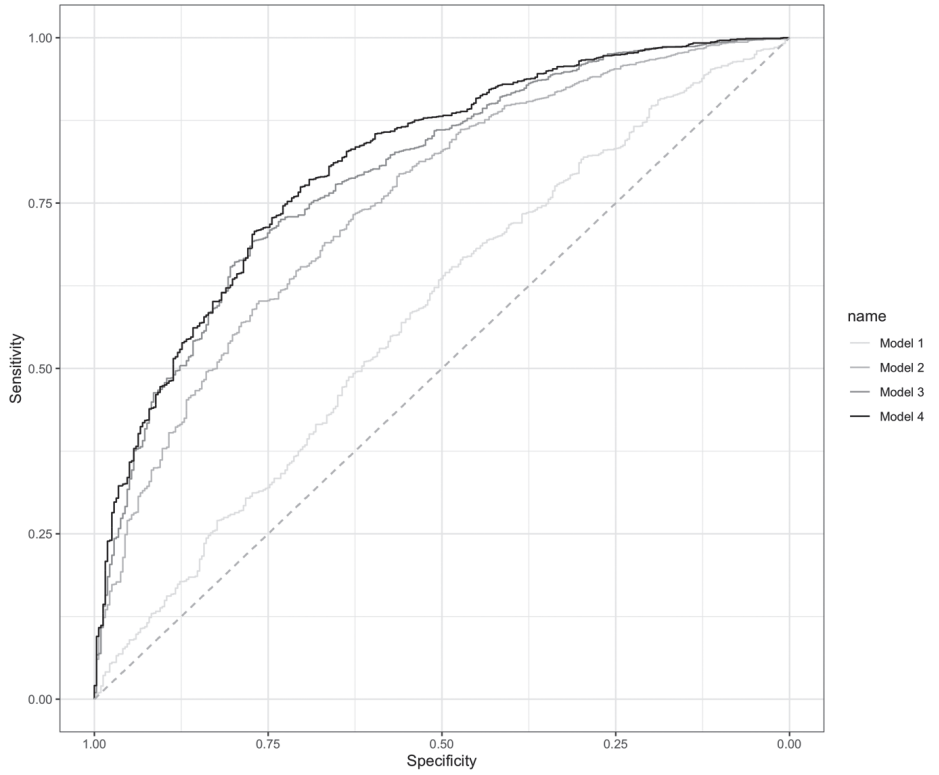


Figure 6. This graph shows the area under the receiver operating characteristic curve for all models explaining the patient's willingness to undergo the treatment again (yes or no; question 2). The dashed line indicates a discriminative ability of 0.50. Model 1, including sociodemographics, had an AUC of 0.58, and after adding clinical patient-reported outcome measures, Model 2 had an AUC of 0.75. After adding PREMs, Model 3 had an AUC of 0.79. After adding measures of mental health, the most definitive model had an AUC of 0.81.

Table 1. Characteristics at baseline of all included patients (n = 1824)

	Value
Age in years	59 ± 11
Sex (Male)	39 (704)
Second opinion	2 (42)
Recurrence (Yes)	8 (146)
Hand dominance	
Right	88 (1607)
Left	8 (153)
Both	4 (64)
Dominant hand treated	49 (902)
Symptom duration in months median (interquartile range)	12 (6-24)
Workload	
Not employed	40 (734)
Light	27 (492)
Moderate	23 (427)
Severe	9 (171)
BMI in kg/m^2	27 ± 5
Smoking (No)	86 (1571)

Data presented as mean ± SD or % (n), unless otherwise noted.

Table 2. Most definitive model after the hierarchical logistic regression analyses (n = 1824) using sociodemographic, clinical characteristics, experience, and mental health characteristics explaining satisfaction with treatment results

Variables	Range (when applicable)	Nonstandardized OR (95% CI)	Standardized OR (95% CI)	p value
Age in years		0.99 (0.98-1.01)	0.92 (0.79-1.07)	0.27
Sex (male)		1.22 (0.95-1.59)	1.22 (0.95-1.59)	0.13
BMI		0.98 (0.96-1.00)	0.91 (0.81-1.02)	0.11
Dominant side treated (yes)		0.87 (0.69-1.10)	0.87 (0.69-1.10)	0.24
Workload (reference = unemployed)				
Light		1.04 (0.76-1.42)	1.04 (0.76-1.42)	0.81
Moderate		1.07 (0.77-1.48)	1.07 (0.77-1.48)	0.70
Severe		0.79 (0.50-1.24)	0.79 (0.50-1.24)	0.30
Symptom duration in months		1.00 (1.00-1.00)	1.03 (0.91-1.16)	0.66
Second opinion (no)		1.02 (0.48-2.18)	1.02 (0.48-2.18)	0.96
Recurrence (yes)		0.95 (0.63-1.45)	0.95 (0.63-1.45)	0.81
Smoking (no)		0.94 (0.67-1.32)	0.94 (0.67-1.32)	0.73
EQ-5D VAS self-rated health	0-100	1.01 (1.00-1.01)	1.13 (1.00-1.28)	0.05
Change in VAS pain during load	0-100	1.03 (1.02-1.03)	2.52 (2.18-2.92)	< 0.001
Change in VAS function	0-100	1.02 (1.01-1.02)	1.76 (1.54-2.01)	< 0.001
PREM shared decision-making positive (yes)		1.04 (0.80-1.36)	1.04 (0.80-1.36)	0.77
PREM pros/cons positive (yes)		1.83 (1.41-2.38)	1.83 (1.41-2.38)	< 0.001
PREM advice positive (yes)		1.57 (1.21-2.04)	1.57 (1.21-2.04)	< 0.001
B-IPQ consequences	0-10	0.95 (0.89-1.01)	0.88 (0.75-1.04)	0.12
B-IPQ timeline	0-10	1.06 (1.01-1.12)	1.20 (1.04-1.37)	0.01
B-IPQ personal control	0-10	1.09 (1.04-1.14)	1.24 (1.10-1.40)	< 0.001
B-IPQ identity	0-10	0.93 (0.88-0.99)	0.84 (0.72-0.97)	0.02
B-IPQ concern	0-10	0.94 (0.89-1.00)	0.84 (0.72-0.99)	0.04

Variables	Range (when applicable)	Nonstandardized OR (95% CI)	Standardized OR (95% CI)	p value
B-IPQ coherence	0-10	0.98 (0.92-1.04)	0.95 (0.84-1.08)	0.43
B-IPQ emotional response	0-10	1.00 (0.95-1.06)	1.01 (0.86-1.18)	0.94
CEQ credibility score	3-27	1.03 (0.98-1.08)	1.11 (0.95-1.30)	0.19
CEQ expectancy score	3-27	1.05 (1.01-1.09)	1.23 (1.04-1.46)	0.02
PCS total score	0-52	0.99 (0.97-1.00)	0.90 (0.78-1.04)	0.17
PHQ-4 total score	0-12	1.01 (0.95-1.08)	1.03 (0.89-1.18)	0.70

Nonstandardized and standardized odds ratios, 95% CIs, and p values are displayed, along with the AUC and the Nagelkerke r^2 for the model; the nonstandardized odds ratios in our most definitive model indicate that with every unit increase in either a continuous, dichotomous, or categorical independent variable, the odds of being satisfied with the treatment results increase or decrease by the value of the nonstandardized OR; standardized odds ratio are converted to the same scale, which makes it easier to make between-variable comparisons and determine the relative association of each explanatory variable; interpretation AUC (ability of the model to discriminate between more satisfied and less satisfied patients) = 0.82; interpretation Nagelkerke r^2 (goodness of fit of the model) = 0.39; EQ5D = EuroQol-5 Dimensions; PREM = Patient-Reported Experience Measures; B-IPQ = Brief Illness Perception Questionnaire; CEQ = Credibility/Expectancy Questionnaire; PCS = Pain Catastrophizing Scale; PHQ = Patient Health Questionnaire; OR = Odds Ratio; SOR = Standardized Odds Ratio.

Table 3. Most-definitive model after the hierarchical logistic regression analyses (n = 1824) using sociodemographic, clinical characteristics, experience, and mental health characteristics explaining undergo treatment again

Variables	Range (when applicable)	Nonstandardized OR (95% CI)	Standardized OR (95% CI)	p value
Age in years		0.99 (0.97-1.01)	0.90 (0.75-1.09)	0.28
Sex (male)		1.11 (0.80-1.54)	1.11 (0.80-1.54)	0.53
BMI		0.99 (0.96-1.02)	0.93 (0.81-1.07)	0.33
Dominant side treated (yes)		0.84 (0.63-1.11)	0.84 (0.63-1.11)	0.23
Workload (reference = unemployed)				
Light		1.30 (0.87-1.93)	1.30 (0.87-1.93)	0.20
Moderate		0.85 (0.56-1.27)	0.85 (0.56-1.27)	0.42
Severe		0.77 (0.44-1.35)	0.77 (0.44-1.35)	0.35
Symptom duration in months		1.01 (1.00-1.01)	1.27 (1.09-1.52)	< 0.01
Second opinion (no)		1.30 (0.52-3.00)	1.30 (0.52-3.00)	0.55
Recurrence (yes)		1.00 (0.62-1.64)	1.00 (0.62-1.64)	0.99
Smoking (no)		0.87 (0.56-1.32)	0.87 (0.56-1.32)	0.51
EQ-5D VAS self-rated health	0-100	1.00 (0.99-1.01)	0.96 (0.82-1.12)	0.65
Change in VAS pain during load	0-100	1.02 (1.01-1.02)	1.74 (1.48-2.07)	< 0.001
Change in VAS function	0-100	1.02 (1.01-1.02)	1.80 (1.54-2.11)	<0.001
PREM shared decision-making positive (yes)		1.45 (1.06-1.99)	1.45 (1.06-1.99)	0.02
PREM pros cons positive (yes)		2.05 (1.50-2.80)	2.05 (1.50-2.80)	< 0.001
PREM advice positive (yes)		1.52 (1.11-2.07)	1.52 (1.11-2.07)	0.01
B-IPQ consequences	0-10	0.95 (0.87-1.02)	0.87 (0.71-1.06)	0.17
B-IPQ timeline	0-10	1.01 (0.95-1.07)	1.02 (0.86-1.21)	0.82
B-IPQ personal control	0-10	1.02 (0.97-1.08)	1.06 (0.92-1.23)	0.41
B-IPQ identity	0-10	1.00 (0.93-1.07)	1.00 (0.82-1.20)	0.96
B-IPQ concern	0-10	0.99 (0.92-1.06)	0.96 (0.79-1.17)	0.71
B-IPQ coherence	0-10	1.08 (1.01-1.16)	1.17 (1.01-1.34)	0.03

Variables	Range (when applicable)	Nonstandardized OR (95% CI)	Standardized OR (95% CI)	p value
B-IPQ emotional response	0-10	1.00 (0.94-1.07)	1.01 (0.83-1.23)	0.93
CEQ credibility score	3-27	1.11 (1.06-1.18)	1.44 (1.20-1.73)	< 0.001
CEQ expectancy score	3-27	0.99 (0.94-1.04)	0.96 (0.78-1.18)	0.71
PCS total score	0-52	1.00 (0.98-1.02)	0.97 (0.82-1.15)	0.73
PHQ-4 total score	0-12	0.98 (0.90-1.06)	0.96 (0.81-1.13)	0.59

Nonstandardized and standardized odds ratios, 95% CIs, and p values are displayed, along with the AUC and Nagelkerke's r^2 for the model; the non-standardized odds ratios in our most-definitive model indicate that with every unit increase in either a continuous, dichotomous, or categorical independent variable, the odds of being willing to undergo the treatment again increase or decrease by the value of the nonstandardized OR; standardized odds ratios are converted to the same scale, which makes it easier to make between-variable comparisons and determine the relative association of each explanatory variable; interpretation AUC (ability of the model to discriminate between willing or not willing to undergo again) = 0.81; interpretation of the Nagelkerke r^2 (goodness of fit of the model) = 0.29;
EQ-5D = EuroQol-5 Dimensions; PREM = Patient-Reported Experience Measures; B-IPQ = Brief Illness Perception Questionnaire; CEQ = Credibility/Expectancy Questionnaire; PCS = Pain Catastrophizing Scale; PHQ = Patient Health Questionnaire; OR = Odds Ratio; SOR = Standardized Odds Ratio.

13

TREATMENT INVASIVENESS AND ILLNESS PERCEPTIONS ARE STRONGLY ASSOCIATED WITH OUTCOME EXPECTATIONS IN PATIENTS TREATED FOR HAND OR WRIST CONDITIONS: A CROSS-SECTIONAL STUDY

de Ridder, W. A., Hoogendam, L., Zeroual, F., Slijper, H. P., Wouters, R. M., Vermeulen, G. M., Selles, R. W., van der Oest, M. J. W., The Hand-Wrist Study Group (2023). *Clinical orthopaedics and related research*, *481*(5), 994–1005.

ABSTRACT

Background

Multiple studies have shown that more-positive outcome expectations are associated with better treatment outcomes. Although this has not been shown to represent a causal relationship, there nonetheless is an interest in positively modifying outcome expectations to improve treatment outcomes. However, little is known about what is independently associated with outcome expectations in clinical practice. For example, it is unknown to what extent expectations are associated with contextual factors such as treatment or patient characteristics such as sociodemographics, or patient-reported outcome measures (PROMs) on patient perceptions of physical or mental health or illness. Studying factors associated with outcome expectations may provide relevant information for clinicians and researchers aiming to improve outcome expectations. Improving expectations might, in turn, improve treatment outcomes.

Question/purpose

Which factors (that is, sociodemographics, PROMs, illness perceptions, treatment, surgeon, and location) are independently associated with outcome expectations in patients with hand or wrist conditions?

Methods

This was a cross-sectional study. Between July 2018 and December 2021, we screened 21,327 patients with a diagnosed hand or wrist condition with complete baseline sociodemographic data such as age and workload. Sixty percent (12,765 of 21,327) of patients completed all relevant PROMs. We excluded patients receiving rare treatments, leaving 58% (12,345 of 21,327) for inclusion in the final sample. Those who participated were more often scheduled for surgical treatment and had higher expectations. We performed a multilevel analysis involving two steps. First, we evaluated whether patients receiving the same treatment, being counseled by the same surgeon, or being treated at the same location have more similar outcome expectations. We found that only patients receiving the same treatment had more similar outcome expectations. Therefore, we used a multilevel regression model to account for this correlation within treatments, and added treatment characteristics (such as nonsurgical versus minor or major surgery, which explained the effectiveness of each treatment) to potential explanatory factors. Second, in the multilevel hierarchical regression analysis, we added sociodemographics (Model 1), PROMs for physical and mental health (Model 2), illness perceptions (Model 3), and treatment characteristics (most-definitive model) to assess the explained variance in outcome expectations per step and the relative association with outcome expectations.

Results

Sociodemographic factors such as age and workload explained 1% of the variance in outcome expectations. An additional 2% was explained by baseline PROMs for physical and mental health, 9% by illness perceptions, and 18% by treatment characteristics, resulting in an explained variance of 29% of the most-definitive model. A large number of patient and treatment characteristics were associated with outcome expectations. We used standardized betas to compare the magnitude of the effect of the different continuous and categorical variables. Among the associated variables, minor surgery (standardized beta [β] = 0.56 [95% confidence interval 0.44 to 0.68]; $p < 0.001$) and major surgery (β = 0.61 [95% CI 0.49 to 0.73]; $p < 0.001$) had the strongest positive association with outcome expectations (receiving surgery is associated with higher outcome expectations than nonsurgical treatment). A longer illness duration expected by the patient (-0.23 [95% CI -0.24 to -0.21]; $p < 0.001$) and being treated for the same condition as before (-0.08 [95% CI -0.14 to -0.03]; $p = 0.003$) had the strongest negative association with outcome expectations.

Conclusions

Outcome expectations are mainly associated with the invasiveness of the treatment and by patients' illness perceptions; patients before surgical treatment have more positive expectations of the treatment outcome than patients before nonsurgical treatment, even after accounting for differences in clinical and psychosocial profiles. In addition, patients with a more-positive perception of their illness had more-positive expectations of their treatment. Our findings suggest expectation management should be tailored to the specific treatment (such as surgical versus nonsurgical) and the specific patient (including their perception of their illness). It may be more beneficial to test and implement expectation management strategies for nonsurgical treatments such as physical therapy than for surgical treatments, given that our findings indicate a greater need to do so. An additional advantage of such a strategy is that successful interventions may prevent converting to surgical interventions, which is a goal of the stepped-care principles of standard care. Future studies might investigate the causality of the association between pretreatment expectations and outcomes by performing an experimental study such as a randomized controlled trial, in which boosting expectations is compared with usual care in nonsurgical and surgical groups.

INTRODUCTION

Patients have expectations at the beginning of their treatments regarding potential outcomes. Several studies have shown these expectations play an important role in treatment outcomes [1-4]. Although some studies suggested expectations of medical treatments are already too high and should be tempered by the clinician to cultivate realistic expectations for the patient [5-8], several meta-analyses have found that patients with more-positive pretreatment expectations achieve better outcomes [1-4]. Additionally, in patients treated for hand or wrist conditions, more-positive expectations have been reported to be associated with better outcomes [9-11]. In addition, positive expectations of the treatment outcomes are considered a key mechanism of placebo effects [12,13]. The placebo effect, or contextual nonspecific effect, is a psychobiological effect that is attributed to the overall therapeutic context [14,15]. This context can consist of patient-specific and clinician-specific factors, and the interaction of patient, clinician, treatment location, and treatment factors [16]. Clinical trials have shown considerable improvement in patients in placebo groups compared with an active or no treatment group [17,18]. Although positive expectations increase the contextual, nonspecific effects of a treatment, expectations may vary across patients and may depend on the type of treatment the patient is about to undergo. For example, previous studies showed that patients with hand or wrist disorders scheduled for surgery have higher expectations than similar patients scheduled for nonsurgical treatment [19,20].

Rationale

Using the contextual effects of a treatment may improve healthcare. Because the contextual nonspecific effect is believed to work through positive expectations of the outcome of a treatment, boosting expectations might be an important part of delivering high-quality care. However, little is known about factors independently associated with patient outcome expectations in clinical practice. Knowing the independent factors associated with outcome expectations may help clinicians to improve expectations. Improving expectations might, in turn, improve treatment outcomes. Moreover, it may inform future studies in the development of interventions that boost expectations.

Therefore, we asked: Which factors (such as, sociodemographics, patient-reported outcome measures [PROMs], illness perceptions, treatment, surgeon, and location) are independently associated with outcome expectations in patients with hand or wrist conditions?

PATIENTS AND METHODS

Study Design

This was a cross-sectional study using a population-based sample of patients with hand or wrist conditions treated at our institution, and was reported following the STrengthening the Reporting of Observational studies in Epidemiology statement [21].

Setting

Data collection was part of usual care and occurred between July 2018 and December 2021 at Xpert Clinics. Xpert Clinics currently comprises 25 clinics for hand surgery and hand therapy in the Netherlands. Twenty-three surgeons are certified by the Federation of European Societies for Surgery of the Hand, and more than 150 hand therapists are employed at our treatment centers. Xpert Clinics offers insured care for hand and wrist conditions with no access restrictions because it is covered by public health insurance. At Xpert Clinics, outcomes are routinely evaluated [22]. After a diagnosis is registered during the first consultation, a measurement track is activated, and PROM forms are emailed to the patient. All data are digitally collected using GemsTracker electronic data capture tools (GemsTracker 2020, Erasmus MC and Equipe Zorgbedrijven), a secure internet-based application for distributing questionnaires and forms during clinical research and quality registrations. More details of the procedure at Xpert Clinics have been published [22].

Participants

Participants were eligible for inclusion if they were adults treated for a hand or wrist condition during the study period. We included patients from all measurement tracks, but excluded rare treatments with fewer than 20 patients for generalizability. Treatments can be divided into nonsurgical treatments (such as orthotics, exercise therapy, or injections), minor surgery (including trigger finger release or De Quervain release), and major surgery (such as trapeziectomy with or without ligament reconstruction tendon interposition for osteoarthritis of the thumb base, or corrective osteotomy for radius malunions). Additionally, we excluded patients who did not complete all relevant questionnaires. The number of patients treated during the study period determined the sample size.

We screened 21,327 patients with complete baseline sociodemographic data such as age and workload. Sixty percent (12,765 of 21,327) of patients completed all relevant PROMs. Finally, we excluded patients receiving rare treatments, leaving 58% (12,345 of 21,327) for inclusion in the final sample (Fig. 1). To assess potential selection bias, we performed two nonresponder analyses. For this, we used the standardized mean difference as a measure of imbalance (standardized mean difference > 0.2 is considered to be imbalanced [23]). First, we compared the sociodemographic characteristics of patients who completed the Credibility/Expectancy Questionnaire (CEQ) (defined as responders) with patients who did

not (defined as nonresponders). Second, we compared sociodemographic characteristics and the CEQ expectancy score of patients who additionally completed the other questionnaires of interest (responders) with patients who did not (nonresponders). In the first analysis, we found a small difference between responders and nonresponders (standardized mean difference = 0.43) (Supplemental Table 1; supplemental materials are available with the online version of *CORR*®). In the second analysis, we found a small difference in treatment group (standardized mean difference = 0.28) and CEQ expectancy score (standardized mean difference = 0.21) (Supplemental Table 2; supplemental materials are available with the online version of *CORR*®). Those who participated were more likely to be in the surgical treatment group and to have higher expectations.

To assess the association between different degrees of surgical invasiveness, we distinguished nonsurgical treatment (such as hand therapy for thumb-base osteoarthritis), minor surgery (such as trigger finger release), and major surgery (such as Triangular Fibrocartilage Complex reinsertion). Twenty-nine percent (3544 of 12,345) of the final sample were scheduled for nonsurgical treatment, 49% (6022 of 12,345) for a minor surgical intervention, and 23% (2779 of 12,345) for a major surgical intervention (Table 1). The number of surgical patients in the present study does not reflect the actual distribution of surgical versus nonsurgical patients at Xpert Clinics, because the inclusion of patients in the present study depends on whether a measurement track is assigned. At the time of this study, no measurement tracks were started in our cohort in patients with, for example, a "wait and see" policy or patients receiving steroid injections. Therefore, the proportion of surgical patients is overestimated in this study. Patients in the major surgery group had a longer duration of symptoms and were more often treated for the same disease previously. Patients in the minor surgery group had the most positive expectations (Supplemental Table 3; supplemental materials are available with the online version of *CORR*®). Furthermore, to assess potential differences between patients scheduled for nonsurgical treatment and patients scheduled for surgical treatment, we stratified patients into two treatment groups: nonsurgical and surgical. Seventy-one percent (8801 of 12,345) were scheduled for either minor or major surgery.

Variables and Measurements

The primary outcome in this study was patients' outcome expectations of the treatment. We measured outcome expectations with the expectancy subscale of the CEQ [24]. This subscale consists of three items measuring the expected magnitude of improvement because of the prescribed treatment. Summed scores range from 3 to 27, where a higher score reflects a more positive treatment outcome expectation.

Independent Variables

We believed patients receiving the same treatment, counseled by the same surgeon, or treated at the same location might have more similar outcome expectations than other patients. To evaluate this, we used multilevel regression modeling with a random intercept and no fixed factors and intraclass correlation coefficients (ICC). Only for treatment, we found that patients were more similar in outcome expectations (Supplemental Digital Content 1; supplemental materials are available with the online version of *CORR*®). Therefore, we included the treatment level in all subsequent analyses.

Patient Characteristics

We divided patient characteristics into three subcategories: sociodemographics, PROMs for physical and mental health, and illness perception. Sociodemographic characteristics included age, sex (not gender, because we collect sex at the Dutch Citizen Service Administration, and we did not want to make unsupported assumptions), therapist-reported duration of symptoms (in months), hand dominance, therapist-reported occupational intensity (unemployed or light, moderate, or heavy physical labor), whether the patient visited the clinic for a second opinion, and whether the disease was recurrent (measured by the question: "Have been treated for the same disease before?"; the answer yes would be coded as recurrent. This means that a patient answering "yes" had the same or a different treatment for the same disease previously).

PROMs for physical and mental health included pain, hand function, health-related quality of life, psychologic distress, and pain catastrophizing at baseline. We used a VAS score (range 0 to 100) to measure the mean pain as experienced in the preceding week (higher scores indicate more pain) and hand function (higher scores indicate better function). The VAS is a validated and widely used tool for measuring these constructs [25]. We measured health-related quality of life using the VAS of the EuroQol-5 Dimensions self-rated health questionnaire as an indication of the overall perceived health status (range 0 to 100; higher scores indicate better perceived health) [26,27]. Psychologic distress was measured with the Patient Health Questionnaire-4 (range 0 to 12; higher scores indicate more distress [28]), and pain catastrophizing was measured with the Pain Catastrophizing Scale (range 0 to 52; higher scores indicate a higher amount of catastrophizing [29]).

The last set of patient characteristics concerned illness perception as measured with the Brief Illness Perception Questionnaire [30,31]. The Brief Illness Perception Questionnaire measures patients' perception of their illness across eight domains (consequences, timeline, personal control, treatment control, identity, concern, coherence, and emotional response). Each domain is assessed with a single question (range 0 to 10; higher scores indicate more negative illness perceptions except for personal control, treatment control, and coherence, where the reverse is true) [30]. We excluded the domain of treatment control ("How much

do you think your treatment can help your illness?") because of conceptual overlap with outcome expectations.

Treatment Characteristics

The treatment characteristics concerned the invasiveness and past effectiveness of the treatment. As an indicator of invasiveness, we coded a treatment as nonsurgical, minor surgery, or major surgery. In addition, as a proxy for the influence of the clinician's explanation of treatment effectiveness, for each treatment, we calculated the mean improvement in function achieved in patients treated previously, using VAS function scores (-100 = maximum deterioration in function; 100 = maximum improvement in function) administered at baseline and at 3 months. We did the same for pain (-100 = maximum deterioration in pain; 100 = maximum improvement in pain).

Finally, we used the Patient-Reported Experience Measure to measure the patient's experience with healthcare delivery, directly after the first consultation. This questionnaire is based on the Consumer Quality Index [32]. The Patient-Reported Experience Measure comprises 16 questions rated on a 4-point Likert scale, including questions about accessibility, reception in the clinic, and communication of the physician.

Ethical Approval

Ethical approval for this study was obtained from the medical ethics committee of the Erasmus MC Medical Centre, Rotterdam (MEC-2018-1088). Informed consent was obtained from patients before data collection started.

Statistical Methods

We used multilevel hierarchical regression analyses to test the relative association of specific patient and treatment characteristics with outcome expectations. In a hierarchical regression analysis, a set of variables is entered into a specific sequence to illustrate each set's added amount of explained variance. This means that variables that add no or little to the explained variance remain in the model. In the first model, we entered all sociodemographic patient characteristics (such as sex, age, and occupational intensity). We added PROMs for physical and mental health (such as quality of life, pain, function, and psychologic distress) in the second model, illness perceptions in the third model, and treatment characteristics in the most-definitive model (the fourth model). An advantage of hierarchical regression is that because of shared variance, some variables might be pushed out of significance when entering the next step. Consequently, only variables that are truly associated with outcome expectations remain significant in the final model. For each model, the explained variance using multilevel partitioning was calculated.

Finally, we performed a stratified analysis to compare differences between factors associated with outcome expectations between patients scheduled for nonsurgical treatment and those scheduled for surgical treatment. Stratification is a useful strategy to identify interactions between subgroups such as treatment type.

A variance inflation factor greater than 3 was considered to indicate multicollinearity [33]. Based on the variance inflation factors (the highest-variance inflation factor in the multilevel hierarchical regression model equaled 2.05, in the stratified nonsurgical model, it equaled 2.12, and in the stratified surgical model, it equaled 2.03), we did not find any indication for multicollinearity in our models.

For all analyses, a p value < 0.05 was considered statistically significant. We used R statistical software version 4.1.1 for the analyses.

RESULTS

Factors Independently Associated With Outcome Expectations

In our most-definitive model, we found an explained variance of 29%. When analyzing the separate steps of the different models, sociodemographics alone provided an explained variance of 1% in outcome expectations. PROMs for physical and mental health added 2% to the explained variance. Illness perceptions (9%) and treatment characteristics (18%) explained the largest amount of variance in outcome expectations.

We used standardized betas to compare the magnitude of the effect of the different continuous and categorical variables. Higher outcome expectations were associated with the following sociodemographic variables (Fig. 2) (arranged from the largest to the smallest standardized beta coefficients): higher age (0.07; $p < 0.001$), occupational intensity (heavy: 0.06; $p = 0.02$; light: 0.06; $p = 0.002$; moderate: 0.06; $p = 0.008$), shorter duration of symptoms (0.03; $p < 0.001$); female sex (0.05; $p = 0.002$), and not having been treated for the same condition before (0.08; $p = 0.003$) (Table 2). Higher outcome expectations were associated with the following baseline PROMs for physical and mental health (largest to smallest standardized beta coefficients): a higher EQ-5D self-rated health score (0.07; $p < 0.001$), better hand function (0.05; $p < 0.001$), and more pain catastrophizing (0.02; $p = 0.048$). Six of seven illness perception items were associated with greater outcome expectations (from largest to smallest): shorter illness duration expected by the patient (-0.23; $p < 0.001$), better understanding of the condition by the patient (0.12; $p < 0.001$), the more the illness affects the patient's life (0.09; $p < 0.001$), less concern about the illness the patient experiences (-0.08; $p < 0.001$), a larger number of symptoms the patient views as being part of their illness (0.05; $p < 0.001$), and the less the illness affects the patient emotionally (-0.04; $p < 0.001$). The largest standardized beta coefficients were for

treatment characteristics: major surgical treatment (0.61; $p < 0.001$) and minor surgical treatment (0.56; $p < 0.001$). This means that being at the start of a major surgical treatment increases the outcome expectations by 2.75 points (95% confidence interval 2.21 to 3.29; $p < 0.001$) compared with being at the start of a nonsurgical treatment (Supplemental Table 4; supplemental materials are available with the online version of *CORR*®). The mean functional improvement of the treatment was also associated with outcome expectations (0.17; $p < 0.001$) (Table 2).

Analyzing differences in variables between the different steps of the model, we found only one difference (Supplemental Table 5; supplemental materials are available with the online version of *CORR*®). In Model 1, visiting the clinic for a second opinion was associated with lower expectations, but after adding PROMs for physical and mental health, there was no association. This implies that one (or more) of the PROMs, such as pain catastrophizing, have a shared variance with a second opinion and pushes the variable second opinion out of significance.

Differences Between Patients Scheduled for Nonsurgical Treatment and Those Scheduled for Surgical Treatment

In the most-definitive model, including sociodemographics, PROMs for physical and mental health, illness perception, and treatment characteristics, we found an explained variance of 25% for outcome expectations of patients scheduled for nonsurgical treatment. Sociodemographics explained 2%, PROMs for physical and mental health explained 2%, illness perception explained 16%, and treatment characteristics explained 5%. For the outcome expectations of patients scheduled for surgical treatment, the most-definitive model explained 14% of the variance. Sociodemographics explained 2%, PROMs explained 2%, illness perception explained 8%, and treatment characteristics explained 2%.

When comparing the factors associated with outcome expectations between patients scheduled for nonsurgical treatment and those scheduled for surgical treatment, we found greater personal control was associated with more-positive expectations in nonsurgical patients (0.13; $p < 0.001$), whereas higher personal control was associated with more-negative expectations in surgical patients (-0.05; $p < 0.001$) (Fig. 3). Psychologic distress was associated with expectations only in nonsurgical patients (depression: -0.04; $p = 0.04$; anxiety: 0.08; $p < 0.001$). Pain catastrophizing (0.03; $p = 0.03$), whether the patient has been treated for the same disease before (-0.11; $p = 0.001$), and a larger number of symptoms the patient views as being part of their Illness (0.08; $p < 0.001$) were associated with expectations only in surgical patients (Table 2).

DISCUSSION

Multiple studies have shown that more-positive outcome expectations are associated with better treatment outcomes [1-4,9-11], and there is an interest in positively modifying outcome expectations to improve treatment outcomes. However, little was known about factors independently associated with outcome expectations. Studying factors associated with outcome expectations may provide relevant information for clinicians and researchers aiming to improve outcome expectations. Improving expectations might, in turn, improve treatment outcomes. We found patients' outcome expectations for a hand or wrist condition were higher when patients had more-positive perceptions of their illness. Furthermore, patients scheduled for surgical treatment had higher outcome expectations than patients scheduled for nonsurgical treatment, even after adjusting for differences in clinical profile and mindset between patents. Our findings can be used directly in daily clinic by improving expectations and illness perceptions, especially for nonsurgical patients, or in studies that develop interventions to improve expectations.

Limitations

Our study has several limitations. First, because this was an observational study, no causal conclusions can be drawn. Although we theorized the variables in our model drive outcome expectations, the reverse could be just as true for several variables (outcome expectations may be causing illness perceptions), or instead, the relationship may be bidirectional. Experimental studies are necessary to test whether outcome expectations might be strengthened by influencing illness perceptions. Second, we found small differences between patients who responded to the survey (responders) and those who did not (nonresponders). Nonresponders were more often scheduled for nonsurgical treatment and had lower expectations. This is in line with other studies that showed nonsurgical patients are more likely to be lost to follow-up than surgical patients [9,20,34]. Furthermore, our study and others showed that patients scheduled for nonsurgical treatment have lower expectations [7,20,35], so we may assume the difference in expectations between responders and nonresponders is caused by the difference in treatment type we found in the nonresponder analysis. Still, we may have overestimated the expectations of patients undergoing nonsurgical treatment in our study. Third, our study examined pretreatment expectations, but several studies suggested outcome expectations may change during treatment and this change may influence treatment outcomes [35,36]. Nevertheless, a robust association between pretreatment outcome expectations and treatment outcomes has been found in several medical areas [1, 5, 6, 27], indicating the importance of addressing pretreatment expectations. Future research could investigate whether the extent to which outcome expectations change during treatment depends on the type of treatment and how this change affects outcome.

Association of Location, Surgeon, and Treatment Variables With Outcome Expectations

Nineteen percent of the variance in outcome expectations was attributable to differences between treatments rather than differences within treatments. Considering the surgeon and location level, we found the variance in outcome expectations was because of differences in surgeon or location, and almost none was because of differences between surgeons or locations. Theoretically, a surgeon adjusts his or her behavior to the patient, treatment, or other factors, such as workload. This might explain why we mainly saw within-surgeon differences.

Patient and Treatment Factors Independently Associated With Outcome Expectations

Our study showed illness perception is an important factor strongly associated with outcome expectations. The more positively patients perceived their illnesses, the more positive their expectations were of the treatment outcome. Perceived chronicity of the disease and the perceived understanding of the disease displayed the strongest independent association. Given studies usually investigate variables associated with outcome expectations of a single treatment, previous researchers may have missed an important overarching factor driving expectations: the type of treatment a patient is about to undergo. In our study, approximately 18% of the total variance across patients was explained by the treatment invasiveness (nonsurgical, minor, or major surgical treatments) and the past effectiveness of the treatment. These results might indicate that patients believe treatment invasiveness is positively associated with better outcomes, resulting in higher pretreatment expectations by patients scheduled for surgical treatment. This finding is in line with those of other studies [19,20,34]. Our study indicates that expectation management should be tailored to the specific treatment (surgical or nonsurgical) and to the specific patient (including their perception of illness). For example, an intervention aimed to increase the understanding of a specific illness and accompanying treatment (such as offering an illness-specific or patient-specific elearning module with psychoeducation to provide information and support so a patient will better understand their illness and treatment) might effectively correct false (negative) beliefs regarding treatment invasiveness in nonsurgical patients and thus improve their pretreatment expectations.

We found an association with the treatment effectiveness based on the mean improvement in function in historical patients, but not for the mean improvement in pain. Hypothetically, in their explanation of treatment effectiveness, clinicians might avoid strong statements about pain, because the amount of improvement in pain differs greatly between patients and between treatments. However, statements on hand function, including a statement such as: “you will be able to return to work within 12 weeks,” might be safer because this outcome may be more predictable. Additionally, we did not find an association between the

amount of pain at baseline, whereas for function, we found patients with better pretreatment function had higher expectations. This suggests pain might be less important for outcome expectations than pretreatment level of function is.

Differences Between Patients Scheduled for Nonsurgical Treatment and Those Scheduled for Surgical Treatment

The degree of control patients feel they have over their illness was the only illness perception domain not associated with outcome expectations in our hierarchical regression model. However, our stratified analysis shows that the more personal control a nonsurgical patient experienced, the more positive the outcome expectations were, whereas the reverse was true for surgical patients. Because of this opposite effect, they may likely have cancelled each other out in the overall regression analysis. This opposite effect might guide intervention for improving outcome expectations. Patients with an internal locus of control perceive themselves as having a great deal of personal control over their outcomes, whereas patients with an external locus of control believe their outcomes result from external influences. Considering the locus of control, improving outcome expectations in nonsurgical patients should entail an increase in personal control (such as a greater understanding of illness and self-efficacy). In contrast, the outcome expectations of surgical patients might be improved by discussing important external influences (including physician experience and the likelihood of success with treatment).

CONCLUSION

So far, there is some promising evidence for expectancy-focused interventions to improve treatment outcomes [37]. Expectation management appears to be an important element of delivering high-quality healthcare. Our findings suggest expectation management should be tailored to the specific treatment (such as surgical versus nonsurgical) and the specific patient (including their perception of their illness). It may be more beneficial to test and implement expectation management strategies such as physical therapy for nonsurgical treatments than for surgical treatments, given our findings indicate a greater need to do so. An additional advantage of such a strategy is that successful interventions may be able to prevent converting to surgical interventions, which is a goal of the stepped-care principles of standard care. Future studies might investigate the association between pretreatment expectations and outcomes by performing an experimental study, such as a randomized controlled trial, in which boosting expectations is compared with usual care (with no special attention to expectations) in nonsurgical and surgical groups.

ACKNOWLEDGMENTS

We thank all patients who participated and allowed their data to be anonymously used for the present study.

GROUP AUTHORS

Members of the Hand-Wrist Study Group include: Richard Arjen Michiel Blomme, Dirk-Johannes Jacobus Cornelis van der Avoort, Alexander Kroeze, Jeronimus Maria Smit, Jan Debeij, Erik Taco Walbeehm, Gijs Marijn van Couwelaar, Johannes Pieter de Schipper, Johannes Frederikes Maria Temming, Jeroen Hein van Uchelen, Herman Luitzen de Boer, Kennard Harmsen, Oliver Theodor Zöphel, Thybout Matthias Moojen, Xander Smit, Gert-Jan Halbesma, Rob van Huis, Pierre Yves Alain Adriaan Pennehouat, Karin Schoneveld, Yara Eline van Kooij, Joris Jan Veltkamp, Alexandra Fink, Joris Sebastiaan Teunissen, Jaimy Emerentiana Koopman, Nina Louisa Loos, Jak Dekker, Ward Rogier Bijlsma, and Marloes Hendrina Paulina ter Stege

REFERENCES

1. Auer CJ, Glombiewski JA, Doering BK, et al. Patients' expectations predict surgery outcomes: a meta-analysis. *Int J Behav Med.* 2016;23:49-62.
2. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. *J Psychosom Res.* 2006;60:631-637.
3. Charlesworth JEG, Petkovic G, Kelley JM, et al. Effects of placebos without deception compared with no treatment: A systematic review and meta-analysis. *J Evid Based Med.* 2017;10:97-107.
4. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers; 1988
5. Constantino MJ, Glass CR, Arnkoff DB, Ametrano RM, Smith JZ. Expectations. *Psychotherapy Relationships That Work: Evidence-Based Responsiveness*: Oxford University Press; 2011.
6. Crow R, Gage H, Hampson S, et al. The role of expectancies in the placebo effect and their use in the delivery of health care: a systematic review. *Health Technol Assess.* 1999;3:1-96.
7. de Raaij EJ, Schröder C, Maissan FJ, Pool JJ, Wittink H. Cross-cultural adaptation and measurement properties of the Brief Illness Perception Questionnaire-Dutch Language Version. *Man Ther.* 2012;17:330-335.
8. De Ridder WA, Wouters RM, Hoogendam L, et al. Which factors are associated with satisfaction with treatment results in patients with hand and wrist conditions? A large cohort analysis. *Clin Orthop Relat Res.* 2022.
9. Delnoij DM, Rademakers JJ, Groenewegen PP. The Dutch consumer quality index: an example of stakeholder involvement in indicator development. *BMC Health Serv Res.* 2010;10:88.
10. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry.* 2000;31:73-86.
11. Duivenvoorden T, Vissers MM, Verhaar JA, et al. Anxiety and depressive symptoms before and after total hip and knee arthroplasty: a prospective multicentre study. *Osteoarthritis Cartilage.* 2013;21:1834-1840.
12. Finniss DG, Kaptchuk TJ, Miller F, Benedetti F. Biological, clinical, and ethical advances of placebo effects. *Lancet.* 2010;375:686-695.
13. Frouzakis R, Herren DB, Marks M. Evaluation of expectations and expectation fulfillment in patients treated for trapeziometacarpal osteoarthritis. *J Hand Surg Am.* 2015;40:483-490.
14. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res (Hoboken).* 2011;63 Suppl 11:S240-252.
15. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.* 2011;20:1727-1736.
16. Hoogendam L, Koopman JE, van Kooij YE, et al. What are the minimally important changes of four commonly used patient-reported outcome measures for 36 hand and wrist condition-treatment combinations? *Clin Orthop Relat Res.* 2022;480:1152-1166.
17. Hoogendam L, van der Oest MJW, Wouters RM, et al. Patients with higher treatment outcome expectations are more satisfied with the results of nonoperative treatment for thumb base osteoarthritis: a cohort study. *Arch Phys Med Rehabil.* 2021;102:1533-1540.

18. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *Pharmacoeconomics.* 2018;36:675-697.

19. Kadzielski J, Malhotra LR, Zurakowski D, et al. Evaluation of preoperative expectations and patient satisfaction after carpal tunnel release. *J Hand Surg Am.* 2008;33:1783-1788.

20. Kong J, Benedetti F. Placebo and nocebo effects: an introduction to psychological and biological mechanisms. *Handb Exp Pharmacol.* 2014;225:3-15.

21. Kroenke K, Spitzer RL, Williams JB, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics.* 2009;50:613-621.

22. Kutner MHN, C.J., Neter, J., Li, W. *Applied linear statistical models.* Boston: McGraw-Hill; 2005

23. Mandl LA, Burke FD, Shaw Wilgis EF, et al. Could preoperative preferences and expectations influence surgical decision making? Rheumatoid arthritis patients contemplating metacarpophalangeal joint arthroplasty. *Plast Reconstr Surg.* 2008;121:175-180.

24. Marks M, Audigé L, Reissner L, et al. Determinants of patient satisfaction after surgery or corticosteroid injection for trapeziometacarpal osteoarthritis: Results of a prospective cohort study. *Arch Orthop Trauma Surg.* 2015;135:141-147.

25. Miller FG, Kaptchuk TJ. The power of context: reconceptualizing the placebo effect. *J R Soc Med.* 2008;101:222-225.

26. Myers H, Keefe F, George SZ, et al. The influence of a cognitive behavioural approach on changing patient expectations for conservative care in shoulder pain treatment: a protocol for a pragmatic randomized controlled trial. *BMC Musculoskelet Disord.* 2021;22:727.

27. Peerdeman KJ, van Laarhoven AIM, Keij SM, et al. Relieving patients' pain with expectation interventions: a meta-analysis. *Pain.* 2016;157:1179-1191.

28. Price DD, Finniss DG, Benedetti F. A comprehensive review of the placebo effect: recent advances and current thought. *Annu Rev Psychol.* 2008;59:565-590.

29. Selles RW, Wouters RM, Poelstra R, et al. Routine health outcome measurement: development, design, and implementation of the Hand and Wrist Cohort. *Plast Reconstr Surg.* 2020;146:343-354.

30. Stewart-Williams S, Podd J. The placebo effect: dissolving the expectancy versus conditioning debate. *Psychol Bull.* 2004;130:324-340.

31. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. *Psychological Assessment.* 1995;7:524-532.

32. Tolk JJ, Janssen RPA, Haanstra TM, et al. The influence of expectation modification in knee arthroplasty on satisfaction of patients: a randomized controlled trial. *Bone Joint J.* 2021;103-B:619-626.

33. van der Oest MJW, Hoogendam L, Wouters RM, et al. Associations between positive treatment outcome expectations, illness understanding, and outcomes: a cohort study on non-operative treatment of first carpometacarpal osteoarthritis. *Disabil Rehabil.* 2021:1-8.

34. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol.* 2008;61:344-349.

35. Waljee JF, Chung KC. Commentary regarding "Evaluation of expectations and expectation fulfillment in patients treated for trapeziometacarpal osteoarthritis". *J Hand Surg Am.* 2015;40:491-492.

36. Wouters RM, Vranceanu AM, Slijper HP, et al. Patients with thumb-base osteoarthritis scheduled for surgery have more symptoms, worse psychological profile, and higher expectations than nonsurgical counterparts: a large cohort analysis. *Clin Orthop Relat Res.* 2019;477:2735-2746.

37. Zou K, Wong J, Abdullah N, et al. Examination of overall treatment effect and the proportion attributable to contextual effect in osteoarthritis: meta-analysis of randomised controlled trials. *Ann Rheum Dis.* 2016;75:1964-1970.

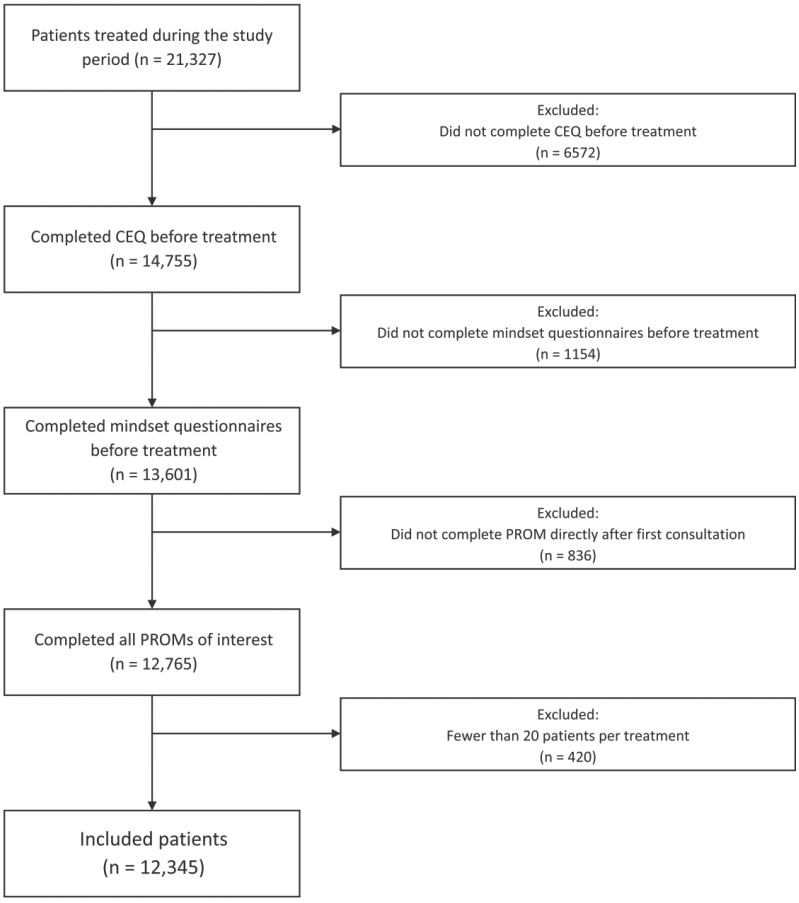


Figure 1. This flowchart represents the patients who were included in this study. CEQ = Credibility and Expectancy Questionnaire.

Figure 2. Standardized regression coefficients of the hierarchical multilevel regression model explain outcome expectations. Only significant variables are shown. EQ-5D = EuroQol-5 Dimensions; B-IPQ = Brief Illness Perception Questionnaire; PCS = Pain Catastrophizing Scale; PHQ = Patient Health Questionnaire.


Explaining CEQ expectation score
Mean functional improvement of treatment
B-IPQ personal control
B-IPQ coherence
B-IPQ consequences
EQ-5D index score
Moderate occupational intensity
Light occupational intensity
PHQ anxiety score
Age
VAS hand function
Duration of symptoms
PHQ depression score
B-IPQ concern
Sex (male)
B-IPQ timeline
-1
-0.5
0
0.5
1
Standardized estimates
A

Figure 3A

Figure 3B

Figure 3. Standardized regression coefficients of the stratified hierarchical multilevel regression models explain outcome expectations for (A) nonsurgical treatment and (B) surgical treatment. Only significant variables are shown. EQ-5D = EuroQol-5 Dimensions; B-IPQ = Brief Illness Perception Questionnaire; PCS = Pain Catastrophizing Scale; PHQ = Patient Health Questionnaire.

Table 1. Characteristics of the included patients (n = 12,345)

Characteristics	Total
Age in years	55 ± 15
Sex (female)	65 (7986)
Duration of symptoms in months	8 (4 -18)
Hand dominance	
Right	89 (10,960)
Left	8 (1013)
Both	3 (372)
Occupational intensity	
Not employed	37 (4553)
Light (working in an office)	28 (3506)
Moderate (working in a shop)	25 (3110)
Severe (working in construction)	10 (1176)
Second opinion	2 (301)
Recurrent disease	8 (1028)
Treatment group	
Nonsurgical treatment	29 (3544)
Minor surgery	49 (6022)
Major surgery	23 (2779)

Data presented as mean ± SD, median (IQR) or % (n).
Nonsurgical treatments includes e.g., orthotics, exercise therapy, injections; minor surgery includes minor surgical interventions e.g., trigger finger release, De Quervain release; major surgery includes more invasive interventions, e.g., trapeziectomy with or without ligament reconstruction tendon interposition for thumb base osteoarthritis, corrective osteotomy for radius malunions.

Table 2. Most-definitive model (standardized beta coefficients) after the hierarchical linear regression analyses (n = 12,345) using sociodemographics, PROMs for physical and mental health, illness perception, and treatment characteristics explaining outcome expectations

Variables	Expectations for all treatments		Expectations for nonsurgical treatment		Expectations for surgical treatment	
	Standardized coefficients (95% CI)	p value	Standardized coefficients (95% CI)	p value	Standardized coefficients (95% CI)	p value
Sociodemographics						
Age in years	0.07 (0.05 to 0.09)	< 0.001	0.07 (0.04 to 0.11)	< 0.001	0.08 (0.06 to 0.11)	< 0.001
Sex (female)	0.05 (0.02 to 0.09)	0.002	0.09 (0.03 to 0.16)	0.01	0.05 (0.0 to 0.09)	0.04
Light occupational intensity (reference: not employed)	0.06 (0.02 to 0.10)	0.002	0.08 (0.00 to 0.16)	0.04	0.06 (0.00 to 0.11)	0.04
Moderate occupational intensity (reference not employed)	0.06 (0.02 to 0.10)	0.01	0.09 (0.01 to 0.17)	0.03	0.04 (-0.02 to 0.09)	0.18
Heavy occupational intensity (reference not employed)	0.06 (0.01 to 0.12)	0.02	0.05 (-0.07 to 0.16)	0.43	0.09 (0.01 to 0.16)	0.02
Second opinion: No	0.06 (-0.04 to 0.16)	0.21	0.06 (-0.14 to 0.26)	0.57	0.04 (-0.08 to 0.17)	0.49
Duration of symptoms in months	-0.03 (-0.05 to -0.02)	< 0.001	-0.04 (-0.07 to -0.01)	0.004	-0.03 (-0.05 to -0.01)	0.01
Right hand dominance (reference: left)	-0.01 (-0.06 to 0.04)	0.7	-0.04 (-0.15 to 0.07)	0.49	-0.01 (-0.08 to 0.06)	0.78
Both hand dominance (reference: left)	-0.03 (-0.13 to 0.07)	0.56	-0.07 (-0.26 to 0.13)	0.5	-0.02 (-0.15 to 0.11)	0.77
Recurrent: yes	-0.08 (-0.14 to -0.03)	0.003	-0.12 (-0.28 to 0.05)	0.17	-0.11 (-0.18 to -0.05)	0.001
PROMs for physical and mental health						
VAS function	0.05 (0.03 to 0.06)	< 0.001	0.03 (0.00 to 0.06)	0.05	0.06 (0.04 to 0.09)	< 0.001
VAS pain	0.01 (-0.01 to 0.03)	0.21	0.01 (-0.02 to 0.05)	0.41	0.01 (-0.01 to 0.04)	0.29
EQ-5D self-rated health	0.07 (0.05 to 0.09)	< 0.001	0.09 (0.05 to 0.13)	< 0.001	0.07 (0.04 to 0.10)	< 0.001
Pain Catastrophizing Score	0.02 (0.00 to 0.04)	0.048	0.01 (-0.02 to 0.05)	0.46	0.03 (0.00 to 0.05)	0.03
PHQ Depression Score	-0.01 (-0.04 to 0.01)	0.18	-0.04 (-0.09 to -0.00)	0.04	0.00 (-0.02 to 0.03)	0.74
PHQ Anxiety Score	0.01 (-0.01 to 0.03)	0.3	0.08 (0.04 to 0.12)	< 0.001	-0.03 (-0.05 to 0.00)	0.06
Illness perception						
B-IPQ Consequences	0.09 (0.07 to 0.11)	< 0.001	0.11 (0.07 to 0.15)	< 0.001	0.10 (0.07 to 0.13)	< 0.001

Variables	Expectations for all treatments		Expectations for nonsurgical treatment		Expectations for surgical treatment	
	Standardized coefficients (95% CI)	p value	Standardized coefficients (95% CI)	p value	Standardized coefficients (95% CI)	p value
B-IPQ Timeline	-0.23 (-0.24 to -0.21)	< 0.001	-0.36 (-0.39 to -0.32)	< 0.001	-0.21 (-0.24 to -0.19)	< 0.001
B-IPQ Personal control	0.01 (-0.01 to 0.02)	0.45	0.13 (0.10 to 0.16)	< 0.001	-0.05 (-0.07 to -0.03)	< 0.001
B-IPQ Identity	0.05 (0.03 to 0.06)	< 0.001	0.03 (-0.01 to 0.06)	0.15	0.06 (0.04 to 0.09)	<0.001
B-IPQ Concern	-0.08 (-0.10 to 0.06)	< 0.001	-0.06 (-0.10 to -0.02)	0.002	-0.10 (-0.13 to -0.07)	< 0.001
B-IPQ Coherence	0.12 (0.10 to 0.13)	< 0.001	0.12 (0.09 to 0.15)	< 0.001	0.13 (0.11 to 0.15)	< 0.001
B-IPQ Emotional response	-0.04 (-0.06 to -0.02)	< 0.001	-0.02 (-0.06 to 0.02)	0.23	-0.07 (-0.10 to -0.04)	< 0.001
Treatment characteristics						
Type treatment (minor surgery)	0.56 (0.44 to 0.68)	< 0.001	NA	NA	NA	NA
Type treatment (major surgery)	0.61 (0.49 to 0.73)	< 0.001	NA	NA	0.01 (-0.11 to 0.12)	0.92
Mean improvement pain	-0.05 (-0.11 to 0.02)	0.167	-0.02 (-0.08 to 0.05)	0.62	-0.02 (-0.09 to 0.05)	0.59
Mean improvement function	0.17 (0.11 to 0.24)	< 0.001	0.15 (0.09 to 0.22)	< 0.001	0.11 (0.04 to 0.18)	0.002
Random effects						
σ^2	0.67		0.75		0.85	
$\tau_{00\ treatment}$	0.02		0.00		0.02	
ICC	0.03		0.01		0.02	
$N_{treatment}$	56		17		39	
Observations	12,345		3544		8801	
Marginal or conditional r^2	0.293/0.314		0.252/0.256		0.137/0.154	

Standardized beta coefficients, 95% CIs, and p values are displayed, along with the random effects and explained variance expressed in the marginal r^2 for the most-definitive model with all treatments, nonsurgical treatments, and surgical treatments. Standardized β coefficients, converted to the same scale, are reported to allow easier between-variable comparisons and determine the relative association of each explanatory variable. EQ-5D = EuroQol-5 Dimensions; PROM = patient-reported outcome measures; B-IPQ = Brief Illness Perception Questionnaire; PCS = Pain Catastrophizing Scale; PHQ = Patient Health Questionnaire.

14

INVOLVEMENT IN A PERSONAL INJURY CLAIM IS ASSOCIATED WITH MORE PAIN AND DELAYED RETURN TO WORK AFTER ELECTIVE NONSURGICAL OR SURGICAL TREATMENT FOR HAND OR WRIST DISORDERS: A PROPENSITY SCORE-MATCHED COMPARATIVE STUDY

Hoogendam, L., van der Oest, M. J. W., Souer, J. S., Selles, R. W., Hovius, S. E. R., Feitz, R., The Hand-Wrist Study Group (2023). *Clinical orthopaedics and related research*, *481*(4), 751–762.

ABSTRACT

Background

A small proportion of patients treated for a hand or wrist condition are also involved in a personal injury claim that may or may not be related to the reason for seeking treatment. There are already indications that patients involved in a personal injury claim have more severe symptoms preoperatively and worse surgical outcomes. However, for nonsurgical treatment, it is unknown whether involvement in a personal injury claim affects treatment outcomes. Similarly, it is unknown whether treatment invasiveness affects the association between involvement in a personal injury claim and the outcomes of nonsurgical treatment. Finally, most studies did not take preoperative differences into account.

Questions/purposes

(1) Do patients with a claim have more pain during loading, less function, and longer time to return to work after nonsurgical treatment than matched patients without a personal injury claim? (2) Do patients with a personal injury claim have more pain, less function, and longer time to return to work after minor surgery than matched patients without a personal injury claim? (3) Do patients with a personal injury claim have more pain, less function, and longer time to return to work after major surgery than matched patients without a personal injury claim?

Methods

We used data from a longitudinally maintained database of patients treated for hand or wrist disorders in the Netherlands between December 2012 and May 2020. During the study period, 35,749 patients for whom involvement in a personal injury claim was known were treated nonsurgically or surgically for hand or wrist disorders. All patients were invited to complete the VAS (scores range from 0 to 100) for pain and hand function before treatment and at follow-up. We excluded patients who did not complete the VAS on pain and hand function before treatment and those who received a rare treatment, which we defined as fewer than 20 occurrences in our dataset, resulting in 29,101 patients who were eligible for evaluation in this study. Employed patients (66% [19,134 of 29,101]) were also asked to complete a questionnaire regarding return to work. We distinguished among nonsurgical treatment (follow-up at 3 months), minor surgery (such as trigger finger release, with follow-up of 3 months), and major surgery (such as trapeziectomy, with follow-up at 12 months). The mean age was 53 ± 15 years, 64% (18,695 of 29,101) were women, and 2% (651 of 29,101) of all patients were involved in a personal injury claim. For each outcome and treatment type, patients with a personal injury claim were matched to similar patients without a personal injury claim using 1:2 propensity score matching to account for differences in patient characteristics and baseline pain and hand function. For nonsurgical

treatment VAS analysis, there were 115 personal injury claim patients and 230 matched control patients, and for return to work analysis, there were 83 claim and 166 control patients. For minor surgery VAS analysis, there were 172 personal injury claim patients and 344 matched control patients, and for return to work analysis, there were 108 claim and 216 control patients. For major surgery VAS analysis, there were 129 personal injury claim patients and 258 matched control patients, and for return to work analysis, there were 117 claim and 234 control patients.

Results

For patients treated nonsurgically, those with a claim had more pain during load at 3 months than matched patients without a personal injury claim (49 ± 30 versus 39 ± 30, adjusted mean difference 9 [95% confidence interval (CI) 2 to 15]; p = 0.008), but there was no difference in hand function (61 ± 27 versus 66 ± 28, -5 [95% CI -11 to 1]; p = 0.11). Each week, patients with a personal injury claim had a 39% lower probability of returning to work than patients without a claim (HR 0.61 [95% CI 0.45 to 0.84]; p = 0.002). For patients with an injury claim at 3 months after minor surgery, there was more pain (44 ± 30 versus 34 ± 29, adjusted mean difference 10 [95% CI 5 to 15]; p < 0.001), lower function (60 ± 28 versus 69 ± 28, adjusted mean difference -9 [95% CI -14 to -4]; p = 0.001), and 32% lower probability of returning to work each week (HR 0.68 [95% CI 0.52 to 0.89]; p = 0.005). For patients with an injury claim at 1 year after major surgery, there was more pain (36 ± 29 versus 27 ± 27, adjusted mean difference 9 [95% CI 4 to 15]; p = 0.002), worse hand function (66 ± 28 versus 76 ± 26, adjusted mean difference -9 [95% CI -15 to -4]; p = 0.001) and a 45% lower probability of returning to work each week (HR 0.55 [95% CI 0.42 to 0.73]; p < 0.001).

Conclusion

Personal injury claim involvement was associated with more posttreatment pain and a longer time to return to work for patients treated for hand or wrist disorders, regardless of treatment invasiveness. Patients with a personal injury claim who underwent surgery also rated their postoperative hand function as worse than similar patients who did not have a claim. Depending on treatment invasiveness, only 42% to 55% of the personal injury claim patients experienced a clinically relevant improvement in pain. We recommend that clinicians extensively discuss the expected treatment outcomes and the low probability of a clinically relevant improvement in pain with their personal injury claim patients, and that they broach the possibility of postponing treatment.

INTRODUCTION

Patients involved in a personal injury claim seem to have worse symptoms before treatment and poorer outcomes after treatment for hand or wrist conditions than patients without a personal injury claim. Personal injury claim involvement may comprise work-related injury [11], traffic accidents [30], and medical malpractice [29]. Specifically, work-related injuries have been studied frequently [11, 15]. A meta-analysis of 62 studies on the surgical treatment outcomes of patients with work-related upper extremity injuries who were involved in a personal injury claim showed that these patients reported more pain and less function after surgery than patients without such a claim [11].

Although multiple studies have shown an association between personal injury claims and worse treatment outcomes in general, there still remain questions regarding the association between involvement in a personal injury claim and the treatment outcomes of hand and wrist disorders. Because previous studies [11, 15] mainly focused on the outcomes of surgical treatment, it is currently unknown whether this association is also present for patients receiving nonsurgical treatment for hand and wrist conditions. Additionally, Fujihara et al. [11] demonstrated that a considerable number of studies assessing personal injury involvement only reported postoperative treatment outcomes compared with a pre- and posttreatment change. It has been hypothesized that patients are more likely to start a personal injury claim when treatment outcomes are poor, thus explaining the association of personal injury claim involvement and poorer treatment outcomes [15]. Similarly, there are indications that a difference in preoperative symptom severity explains this association [4]. Therefore, it would be relevant to compare treatment outcomes of patients with personal injury claim involvement to patients that were similar before treatment.

Therefore, we asked: (1) Do patients with a claim have more pain during loading, less function, and longer time to return to work after nonsurgical treatment than matched patients without a personal injury claim? (2) Do patients with a personal injury claim have more pain, less function, and longer time to return to work after minor surgery than matched patients without a personal injury claim? (3) Do patients with a personal injury claim have more pain, less function, and longer time to return to work after major surgery than matched patients without a personal injury claim?

PATIENTS AND METHODS

In this matched comparative study using data from a longitudinally maintained database, we investigated whether involvement in a personal injury claim is associated with pain, hand function, and time to return to work after nonsurgical treatment, minor surgery, or major surgical treatment for hand or wrist disorders.

Setting

Data were collected from December 2012 to May 2020 at Xpert Clinics. Xpert Clinics has more than 25 locations and employs 23 European board-certified hand surgeons and more than 100 hand therapists.

As part of routine outcome measurements, all patients were invited to complete the VAS for pain and hand function before treatment, and both the VAS and a questionnaire about return to work after treatment. This cohort and data collection have recently been described in more detail [31].

In the Netherlands, a personal injury claim procedure generally takes 1 to 3 years. Patients involved in a personal injury claim can be compensated for the loss of (future) income, (future) medical costs, and general compensatory damages. Most medical costs are reimbursed through basic health insurance, which covers general practitioner visits, medical specialty care, hospital stay, most medicine, and a limited number of occupational therapy sessions. With the exception of general practitioner visits, there is a compulsory deductible of USD 392 before health care costs are reimbursed by health care insurance. Physical therapy, including hand therapy, is not reimbursed under basic health care insurance.

Participants

We included patients whose involvement in a personal injury claim was known and who completed the VAS questionnaire on pain and hand function before treatment. We then divided patients into three subcohorts: patients who were treated nonsurgically (follow-up at 3 months); patients receiving minor surgery, such as trigger finger release (follow-up at 3 months); and patients receiving major surgery, such as triangular fibrocartilage complex (TFCC) reinsertion (follow-up at 12 months) (Supplementary Table 1). We excluded patients if they did not provide data for at least one of the outcome measures (VAS or return to work) at the last measurement. Additionally, we excluded patients who received a rare treatment (which we defined as fewer than 20 observations in our cohort).

During the study period, 29,101 patients were eligible for participation in this study (Fig. 1). The mean age was 53 ± 15 years, and 64% (18,695 of 29,101) were women. Forty-five percent (12,957 of 29,101) of patients underwent minor surgical treatment, and 2% (651 of 29,101) of patients were involved in a personal injury claim. We found differences in patient characteristics and VAS scores before treatment between patients with a personal injury claim and those without one (Table 1). Patients involved in a personal injury claim were typically younger, had a shorter duration of symptoms, and performed moderate or severe physical labor more often than those who were not involved in a claim. Moreover, patients with a personal injury claim received major surgical treatment more often and scored worse regarding pain and hand function before treatment.

Two percent (115 of 5633) of the nonsurgically treated patients that completed the VAS before and after treatment were involved in a personal injury claim and were matched to 230 patients without a personal injury claim for VAS analyses. After matching, there was a difference (defined as a SMD ≥ 0.10) in gender, dominant side, and treatment between the matched patients with a claim and those without (Supplementary Figure 1), indicating that additional adjustment was required. Additionally, 2% (83 of 4169) of the nonsurgically treated patients who completed the VAS questionnaire before treatment and the questionnaire about return to work and who were employed were involved in a personal injury claim; we matched them to 166 patients without a claim for return-to-work analyses. After matching, there was a difference in treatment between matched patients with and without a claim (Supplementary Figure 2).

For the minor surgery subcohort, 2% (172 of 9590) of patients were involved in a personal injury claim and were matched to 344 patients without a claim for VAS analyses. After matching, there was a difference in gender, dominant side, and treatment (Supplementary Figure 3). For the return-to-work analyses, 2% (108 of 6155) of patients were involved in a claim and were matched to 216 patients without a claim. After matching, there was a difference in the type of occupation and treatment (Supplementary Figure 4).

For the major surgery subcohort, 3% (129 of 4389) of patients were involved in a personal injury claim and were matched to 258 patients without involvement in a claim for the VAS analyses. After matching, there was a difference in treatment (Supplementary Figure 5). For the return-to-work analyses, 4% (117 of 3339) of patients were involved in a claim and were matched to 234 patients without a claim. After matching, there was a difference in symptom duration, VAS function, and treatment (Supplementary Figure 6).

Data Collection

Patients were invited to complete the VAS questionnaire before and after treatment [31]. Patients were asked to score five outcome domains (average pain, pain during load, pain at rest, hand function, and satisfaction with the hand) on a VAS (scores range from 0 to 100). For pain, 0 indicates no pain and 100 indicates extreme pain; for function and satisfaction, 0 indicates poor function and 100 indicates optimal function [20]. For this study, we used the subscales of VAS pain during load and VAS hand function, which we consider the most relevant for patients with hand or wrist disorders. After treatment (at 3 months for nonsurgical treatment and minor surgery or at 12 months for major surgery), patients were invited to complete the VAS questionnaire again. The minimum important change (MIC) of the VAS pain during loading was reported as 13 for nonsurgical treatment and 23 for surgical treatment. For hand function, it was reported as 10 for nonsurgical treatment and 19 for surgical treatment [19].

In addition, patients who were employed before treatment were asked to complete a questionnaire on return to work after treatment [34]. Return to work was defined as working at least 50% of a patient's contractual hours and performing normal working activities.

Sample Size

The study sample was determined by the number of patients treated during the study period. For each subcohort and outcome, we calculated effect sizes to assess whether we had a sufficient number of patients for each analysis to detect a medium-sized effect (Cohen d 0.50) with 80% power [6]. The significance level was set at 0.017, to account for multiple testing of three outcomes for each subcohort. With the available number of patients, we were able to detect effect sizes ranging from 0.35 to 0.37, indicating that we had sufficient power to detect a medium-size effect in the analyses.

Matching

We matched patients based on patient characteristics (age, gender, occupation, and dominant side), baseline VAS scores, and treatment to compare patient-reported outcome measure scores at the last measurement for patients with and without personal injury claims. We used 2:1 propensity score matching to compare treatment outcomes, where we matched two patients without personal injury claims to one patient with a personal injury claim [3]. The number of patients differed for each analysis because not all patients completed all questionnaires. Therefore, we performed matching for each analysis separately. After matching, we calculated the standardized mean difference (SMD) between matched patients with and without personal injury claims to assess whether the propensity score matching was successful. The SMD should be below 0.10 for all variables for matching to be considered successful [25].

Ethical Approval

Institutional review board approval was obtained from the ethics committee of the Erasmus Medical center (MEC-2018-1088). All patients provided written informed consent.

Statistical Analyses

We used t-tests to compare normally distributed continuous variables. For nonnormally distributed variables, we used Wilcoxon tests. We compared categorical outcomes using a chi-square test.

When the SMD was greater than 0.10 for one or more variables after matching, linear regression (for VAS scores) or a Cox proportional hazards model (for return to work) was applied to adjust for variables with an SMD greater than 0.10 after matching [25].

All analyses were performed in R, version 3.6.3. For all analyses, $p < 0.017$ was considered statistically significant because of multiple testing corrections. The R code used for the analyses is available online at https://doi.org/10.5281/zenodo.5607378 [17].

RESULTS

Nonsurgical Treatment

After propensity score matching and adjustment for gender, hand dominance, and treatment, at 3 months after treatment, patients with personal injury claims had more pain with loading than those without such a claim (49 ± 30 versus 39 ± 30, adjusted mean difference 9 [95% confidence interval (CI) 2 to 15]; $p = 0.008$) (Fig. 2A). However, there was no difference in hand function (61 ± 27 versus 66 ± 28, -5 [95% CI -11 to 1]; $p = 0.11$) (Fig. 2B). Forty-five percent (52 of 115) and 50% (57 of 115) of patients with personal injury claims reached the minimum important change (MIC) of VAS pain and function, respectively (Table 2). Among employed patients, after propensity score matching and adjusting for treatment, we found that patients with personal injury claims had a 39% lower probability of returning to work each week than those without a claim (HR 0.61 [95% CI 0.45 to 0.84]; $p = 0.002$) (Fig. 3).

Minor Surgery

After propensity score matching and adjustment for gender, hand dominance, and treatment, at 3 months after treatment, patients with personal injury claims had more pain with loading than those without such a claim (44 ± 30 versus 34 ± 29, adjusted mean difference 10 [95% CI 5 to 15]; $p < 0.001$) (Fig. 4A). Additionally, patients with personal injury claims had a worse hand function than those without such a claim (60 ± 28 versus 69 ± 28, adjusted mean difference -9 [95% CI -14 to -4]; $p = 0.001$) (Fig. 4B). Forty-two percent (72 of 172) and 53% (91 of 172) of patients with personal injury claims reached the MIC of VAS pain and function, respectively (Table 3). Among employed patients, the median time to return to work was 5 weeks for patients with an injury claim and 3 weeks for matched patients without a claim (Fig. 5). After propensity score matching and adjusting for type of occupation and treatment, we found that patients with personal injury claims had a 32% lower probability of returning to work each week (HR 0.68 [95% CI 0.52 to 0.89]; $p = 0.005$).

Major Surgery

After propensity score matching and adjustment for treatment, at 12 months after treatment, patients with personal injury claims had more pain with loading than those without such a claim (36 ± 29 versus 27 ± 27, adjusted mean difference 9 [95% CI 4 to 15]; $p = 0.002$)

(Fig. 6A). Additionally, patients with personal injury claims had a worse hand function than those without a claim (66 ± 28 versus 76 ± 26, adjusted mean difference -9 [95% CI -15 to -4]; p = 0.001) (Fig. 6B). Fifty-five percent (71 of 129) and 59% (76 of 129) of personal injury claim patients reached the MIC of VAS pain and function, respectively (Table 4). Among employed patients, the median time to return to work was 12 weeks for patients with a claim and 8 weeks for matched patients (Fig. 7). After propensity score matching and adjusting for symptom duration, VAS function, and treatment, we found that patients with personal injury claims had a 45% lower probability of returning to work each week than those without a claim (HR 0.55 [95% CI 0.42 to 0.73]; p < 0.001) (Fig. 7).

DISCUSSION

Although a small proportion of patients with hand or wrist disorders is involved in a personal injury claim, there is evidence that personal injury claim involvement is associated with worse surgical treatment outcomes. However, it was unknown whether this association is also present in nonsurgical treatment. Additionally, many studies did not take preoperative differences into account, possibly explaining this association. We found a consistent negative association between personal injury claim involvement and pain and time to return to work after nonsurgical treatment, minor surgery, and major surgery when comparing personal injury claim patients to similar patients without personal injury claim. Patients with a personal injury claim who were treated surgically also reported a worse hand function. Additionally, only 42% to 55% experienced a clinically relevant improvement in pain. We suggest that clinicians discuss the expected treatment outcomes extensively with their patients who have filed personal injury claims before deciding on treatment, and that clinicians should introduce the idea of postponing treatment.

Limitations

This study has some limitations. We used the VAS to measure pain during load and hand function. Although the VAS has been validated to measure pain [20], it has not been validated to measure hand function. Using validated patient-reported outcome measures such as the Patient Rated Wrist/Hand Evaluation or the Michigan Hand outcomes Questionnaire would have better suited. Still, previous studies using the Disabilities of the Hand, Arm, and Shoulder questionnaire and the American Shoulder and Elbow surgeons score found similar results [2, 4, 5, 8], supporting our findings.

We were unable to determine whether the reason for treatment was related to the personal injury claim. For example, there may be a difference in symptom severity, psychological characteristics, and potential treatment outcomes between a patient involved in a personal injury claim for a broken foot who now happens to seek treatment for Dupuytren

contracture, compared with a patient with a personal injury claim who is undergoing a TFCC reinsertion for wrist pain after an accident. Future studies are needed to determine whether unrelated elective care can be postponed when patients have an ongoing personal injury claim.

This observational comparative study did not account for psychological factors, such as depression, pain catastrophizing, or illness perceptions. Several studies have suggested that patients with a personal injury claim have a more negative psychological profile [10, 24]. For example, patients with prolonged sick leave might have more negative illness perceptions than realistically can be expected [12]. Moreover, patients involved in a workers compensation claim have been shown to be at risk for mental illness, including depression [21, 26]. A negative psychological profile has been shown to be associated with treatment outcomes in multiple hand or wrist disorders; including carpal tunnel release [32], upper extremity fractures [1], and TFCC reinsertion [33]. Therefore, we recommend further research to identify pretreatment differences in psychological profiles and how psychological profiles of patients with personal injury claims change during treatment. This could provide insights to guide possible interventions, such as additional psychological, social worker support, or motivational interviewing to improve treatment outcomes [14].

Additionally, there are likely differences in legal systems and claim compensation between countries in terms of the amount of compensation received or required and the duration of a personal injury claim procedure [22]. These differences might influence treatment outcomes [13]. When we compare patients in this study to patients in the United States, for example, a notable difference will be the need for medical costs compensation, which will be highly relevant for US patients, but less so for Dutch patients, as these costs are mostly reimbursed under basic health insurance. Still, loss of (future) income and general compensatory damages will be relevant issues to patients in both settings. Additionally, feelings of psychological distress [7] and perceived injustice [23, 28] will affect personal injury claim patients irrespective countries, suggesting that the similarities in personal injury claims between countries may outweigh the differences. This is in line with a meta-analysis of surgical treatment outcomes of patients with workers compensation or involved in litigation [15], where it was shown that the association between compensation and worse treatment outcomes was consistently present in the United States, Canada, Europe, and Australia.

Nonsurgical Treatment

For patients treated nonsurgically, we found that patients involved in a personal injury claim experienced more pain during load and were more likely to return to work later than patients not involved in a claim. Although it has been reported that personal injury claim

involvement is associated with poorer outcomes in terms of pain and hand function for surgically treated patients [11], we find a similar association for nonsurgical treatment.

Minor Surgery

For patients receiving minor surgery (such as trigger finger release), patients with personal injury claims had more pain, worse function, and a longer time to return to work, with a median time to return to work of 5 weeks compared with 3 weeks for controls. A difference in return to work was also reported by Dunn et al. [9], who found that patients involved in a workers compensation claim are absent from work on average 5 weeks longer. We found a smaller difference of 2 weeks' absence for patients involved in a personal injury claim than matched patients without such a claim. Possibly, work-related injuries lead to a disturbed working relationship which negatively affects the time to return to work, which may explain the 3-week difference in the return-to-work time. We recommend a qualitative study to investigate barriers and facilitators of return to work for patients involved in a personal injury claim to guide potential interventions.

Major Surgery

For patients receiving major surgery (such as TFCC reinsertion), patients with personal injury claims had more pain, worse function, and a longer time to return to work. In the meta-analysis by Fujihara et al. [11], it was unclear whether personal injury claims were associated with time to return to work after surgery, whereas this association was present in our study. Because the treatment invasiveness (minor or major surgery) likely affects the time to return to work, studying minor and major procedures separately might have provided a clearer overview of this association.

Other Considerations

The differences in pain and hand function we found in all treatment groups were relatively small compared with those reported in previous studies [8, 16]. For example, a 15-point difference in VAS pain was reported after rotator cuff repair for patients with and without involvement in a personal injury claim [16], whereas we found a 9-point difference in VAS pain scores during loading after major surgery. Because it has been repeatedly shown that patients involved in a personal injury claim already report more symptoms before treatment [11, 16], correcting for baseline differences may result in a smaller difference after treatment [4], possibly explaining the difference in effect magnitude. Additionally, only 42% to 55% of patients with personal injury claims experienced a clinically relevant improvement in pain after minor and major surgery, respectively. For example, when we compare this to surgery for thumb base osteoarthritis, we find that a clinically relevant improvement in pain has been reported for 71% to 84% of patients [18, 27]. This raises the question of whether patients actually benefit from treatment while they have an ongoing personal injury claim.

Future studies may evaluate whether postponing treatment is an option and how this may affect outcomes. Finally, there is an increasing interest in the development of prediction models for individualized predictions of expected treatment outcomes. Considering that personal injury claim involvement is associated with worse treatment outcomes, regardless of treatment invasiveness, we strongly recommend researchers working on prediction models for individual treatment outcomes to include personal injury claim involvement as a candidate predictor for their models.

Conclusion

This matched comparative study found that involvement in a personal injury claim was associated with worse treatment outcomes for patients treated nonsurgically or surgically for hand or wrist disorders. Regardless of treatment invasiveness (nonsurgical treatment, minor surgery, or major surgery), the time to return to work was longer and posttreatment pain was higher for personal injury claim patients. Surgical personal injury claim patients also rated their postoperative hand function as worse than similar patients. Depending on treatment invasiveness, only 42% to 55% of the personal injury claim patients experienced a clinically relevant improvement in pain. We recommend that clinicians do the following with their patients who have personal injury claims: thoroughly discuss the expected treatment outcomes and the low probability of a clinically relevant improvement in pain as well as introduce the possibility of postponing treatment.

ACKNOWLEDGMENTS

We thank all patients who participated and allowed their data to be anonymously used for the present study.

GROUP AUTHORSHIP

The Hand Wrist Study Group includes: Richard Arjen Michiel Blomme, Dirk-Johannes Jacobus Cornelis van der Avoort, Alexander Kroeze, Jeronimus Maria Smit, Jan Debeij, Erik Taco Walbeehm, Gijs Marijn van Couwelaar, Guus Maarten Vermeulen, Johannes Pieter de Schipper, Johannes Frederikes Maria Temming, Jeroen Hein van Uchelen, Herman Luitzen de Boer, Kennard Harmsen, Oliver Theodor Zöphel, Thybout Matthias Moojen, Xander Smit, Gert-Jan Halbesma, Rob van Huis, Pierre Yves Alain Adriaan Pennehouat, Karin Schoneveld, Yara Eline van Kooij, Robbert Maarten Wouters, Joris Jan Veltkamp, Alexandra Fink, Willemijn Anna de Ridder, Joris Sebastiaan Teunissen, Jaimy Emerentiana Koopman, Nina Louisa Loos, Jak Dekker, Ward Rogier Bijlsma, Marloes Hendrina Paulina ter Stege, and Harm Pieter Slijper.

REFERENCES

1. Al Salman A, Shah R, Thomas JE, et al. Symptoms of depression and catastrophic thinking attenuate the relationship of pain intensity and magnitude of incapability with fracture severity. *J Psychosom Res.* 2022;158:110915.
2. Atanda A Jr, O'Brien DF, Kraeutler MJ, et al. Outcomes after distal biceps repair in patients with workers' compensation claims. *J Shoulder Elbow Surg.* 2013;22:299-304.
3. Austin PC. Statistical criteria for selecting the optimal number of untreated subjects matched to each treated subject when using many-to-one matching on the propensity score. *Am J Epidemiol.* 2010;172:1092-1097.
4. Balyk R, Luciak-Corea C, Otto D, et al. Do outcomes differ after rotator cuff repair for patients receiving workers' compensation? *Clin Orthop Relat Res.* 2008;466:3025-3033.
5. Banerjee M, Balke M, Bouillon B, et al. Soft tissue injury of the shoulder after single non-dislocating trauma: prevalence and spectrum of intraoperative findings during shoulder arthroscopy and treatment results. *Arch Orthop Trauma Surg.* 2015;135:103-109.
6. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates; 1988.
7. Collie A, Sheehan L, Lane TJ, et al. Psychological distress in workers' compensation claimants: prevalence, predictors and mental health service use. *J Occup Rehabil.* 2020;30:194-202.
8. Doherty C, Gan BS, Grewal R. Ulnar shortening osteotomy for ulnar impaction syndrome. *J Wrist Surg.* 2014;3:85-90.
9. Dunn JC, Kusnezov NA, Koehler LR, et al. Outcomes following carpal tunnel release in patients receiving workers' compensation: a systematic review. *Hand (N Y).* 2018;13:137-142.
10. Elbers NA, Hulst L, Cuijpers P, et al. Do compensation processes impair mental health? A meta-analysis. *Injury.* 2013;44:674-683.
11. Fujihara Y, Shauver MJ, Lark ME, et al. The effect of workers' compensation on outcome measurement methods after upper extremity surgery: a systematic review and meta-analysis. *Plast Reconstr Surg.* 2017;139:923-933.
12. Giri P, Poole J, Nightingale P, et al. Perceptions of illness and their impact on sickness absence. *Occup Med (Lond).* 2009;59:550-555.
13. Grant GM, O'Donnell ML, Spittal MJ, et al. Relationship between stressfulness of claiming for injury compensation and long-term recovery: a prospective cohort study. *JAMA Psychiatry.* 2014;71:446-453.
14. Gross DP, Park J, Rayani F, et al. Motivational interviewing improves sustainable return to work in injured workers after rehabilitation: a cluster randomized controlled trial. *Arch Phys Med Rehabil.* 2017;98:2355-2363.
15. Harris I, Mulford J, Solomon M, et al. Association between compensation status and outcome after surgery: a meta-analysis. *JAMA.* 2005;293:1644-1652.
16. Henn RF, 3rd, Tashjian RZ, Kang L, et al. Patients with workers' compensation claims have worse outcomes after rotator cuff repair. *J Bone Joint Surg Am.* 2008;90:2105-2113.
17. Hoogendam L. Personal injury claim. Available at: https://doi.org/10.5281/zenodo.5607378. Accessed October 28, 2021.

18. Hoogendam L, Bink T, de Lange J, et al. Which tendon plasty has the best outcome? A comparison of four tendon plasty techniques in a large cohort of patients with symptomatic trapeziometacarpal osteoarthritis. *Plast Reconstr Surg.* 2022;150:364e-374e.

19. Hoogendam L, Koopman JE, Van Kooij Y, et al. The minimal important change of four common patient-reported outcome measures for 36 different hand and wrist conditions. *Clin Orthop Relat Res*. 2022;480:1152-1166

20. Kahl C, Cleland JA. Visual analogue scale, numeric pain rating scale and the McGill pain Questionnaire: an overview of psychometric properties. *Physical Therapy Reviews.* 2005;10:123-128.

21. Kim J. Depression as a psychosocial consequence of occupational injury in the US working population: findings from the medical expenditure panel survey. *Bmc Public Health*. 2013;13.

22. Lippel K. Preserving workers' dignity in workers' compensation systems: an international perspective. *Am J Ind Med.* 2012;55:519-536.

23. Murgatroyd DF, Cameron ID, Harris IA. Understanding the effect of compensation on recovery from severe motor vehicle crash injuries: a qualitative study. *Inj Prev.* 2011;17:222-227.

24. Murgatroyd DF, Casey PP, Cameron ID, et al. The effect of financial compensation on health outcomes following musculoskeletal injury: systematic review. *PLoS One.* 2015;10:e0117597.

25. Nguyen TL, Collins GS, Spence J, et al. Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. *BMC Med Res Methodol.* 2017;17:78.

26. Orchard C, Carnide N, Mustard C, et al. Prevalence of serious mental illness and mental health service use after a workplace injury: a longitudinal study of workers' compensation claimants in Victoria, Australia. *Occup Environ Med.* 2020;77:185-187.

27. Ottenhoff JSE, Spaans AJ, Braakenburg A, et al. Joint distraction for thumb carpometacarpal osteoarthritis: 2-year follow-up results of 20 patients. *J Wrist Surg.* 2021;10:502-510.

28. Quintner J, Galbraith M. The great trade-off in workers' compensation: perceptions of injustice by those experiencing persistent pain. *Pain Med.* 2022;23:456-465.

29. Schaffer AC, Jena AB, Seabury SA, et al. Rates and characteristics of paid malpractice claims among US physicians by specialty, 1992-2014. *JAMA Intern Med.* 2017;177:710-718.

30. Scollay CE, Berecki-Gisolf J, Grant GM. Trends in lawyer use in road traffic injury compensation claims. *PLoS One.* 2020;15:e0231025.

31. Selles RW, Wouters RM, Poelstra R, et al. Routine health outcome measurement: development, design, and implementation of the Hand and Wrist Cohort. *Plast Reconstr Surg.* 2020.

32. Sun PO, Walbeehm ET, Selles RW, et al. Patient mindset and the success of carpal tunnel release. *Plast Reconstr Surg.* 2021;147:66e-75e.

33. Teunissen JS, van der Oest MJW, van Groeninghen DE, et al. The impact of psychosocial variables on initial presentation and surgical outcome for ulnar-sided wrist pathology: a cohort study with 1-year follow-up. *BMC Musculoskelet Disord.* 2022;23:109.

34. van der Oest MJW, Teunissen JS, Poelstra R, et al. Factors affecting return to work after surgical treatment of trapeziometacarpal joint osteoarthritis. *J Hand Surg Eur Vol.* 2020:1753193420978631.

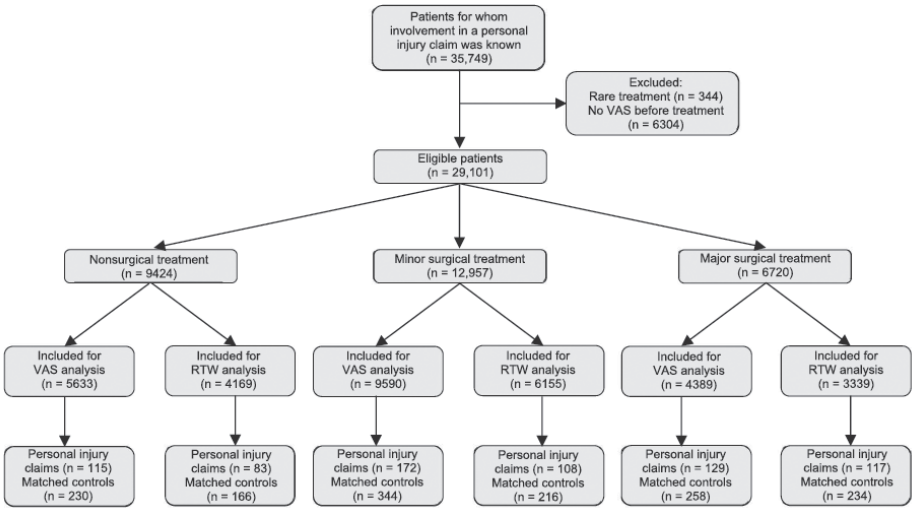


Figure 1. This flowchart shows the patients who were included in this study; RTW = return to work. The inclusion criteria differed for the VAS analyses (patients who completed the VAS prior to treatment and at final follow-up) and the RTW analyses (employed patients who completed VAS prior to treatment), leading to some overlap in patients for these analyses. Therefore, the number of patients per treatment type can be lower than the sum of patients included in the VAS analyses and RTW analyses.

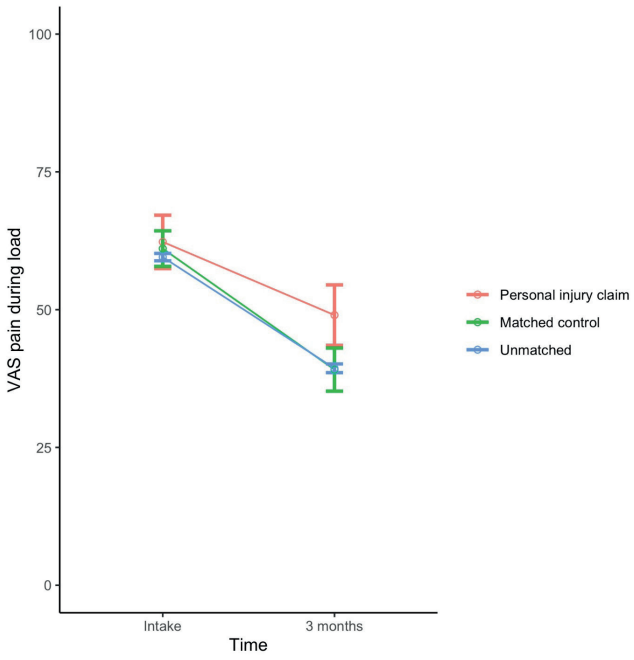


Figure 2A

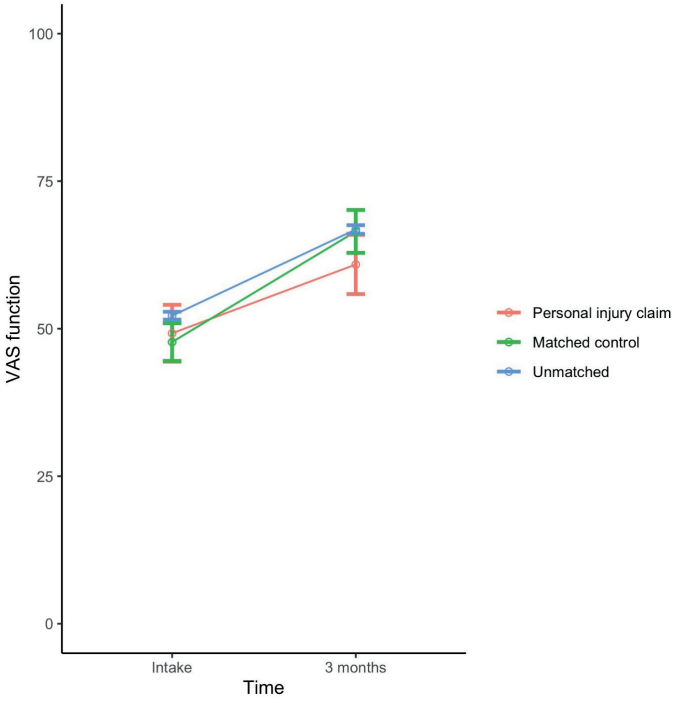


Figure 2B

Figure 2A-B. These graphs show the mean VAS scores for (A) pain and (B) hand function during load over time after nonsurgical treatment. The error bars represent the 95% CI. Values are shown for patients involved in a personal injury claim, matched control patients, and unmatched patients not involved in a personal injury claim.

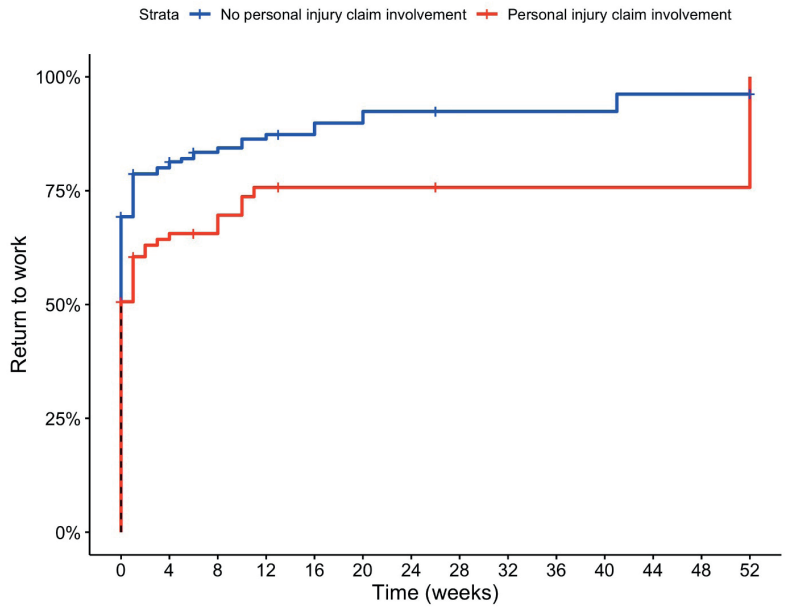


Figure 3. This Kaplan Meier curve represents the time to return to work after nonsurgical treatment for patients involved in a personal injury claim and matched controls.

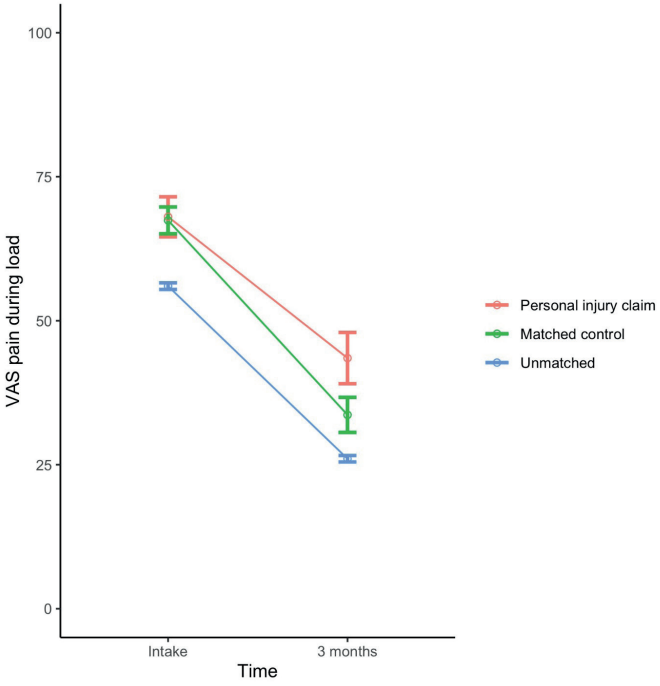

100
75
50
25
0
VAS pain during load
Intake
3 months
Time
Personal injury claim
Matched control
Unmatched

Figure 4A

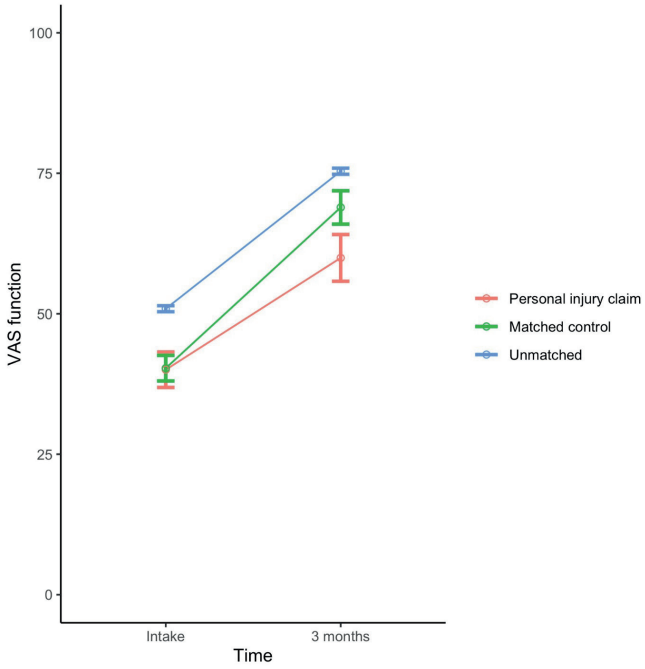


Figure 4B

Figure 4A-B. These graphs show the mean VAS scores for (A) pain and (B) hand function during load over time after minor surgery. The error bars represent the 95% CI. Values are shown for patients involved in a personal injury claim, matched control patients, and unmatched patients not involved in a personal injury claim.

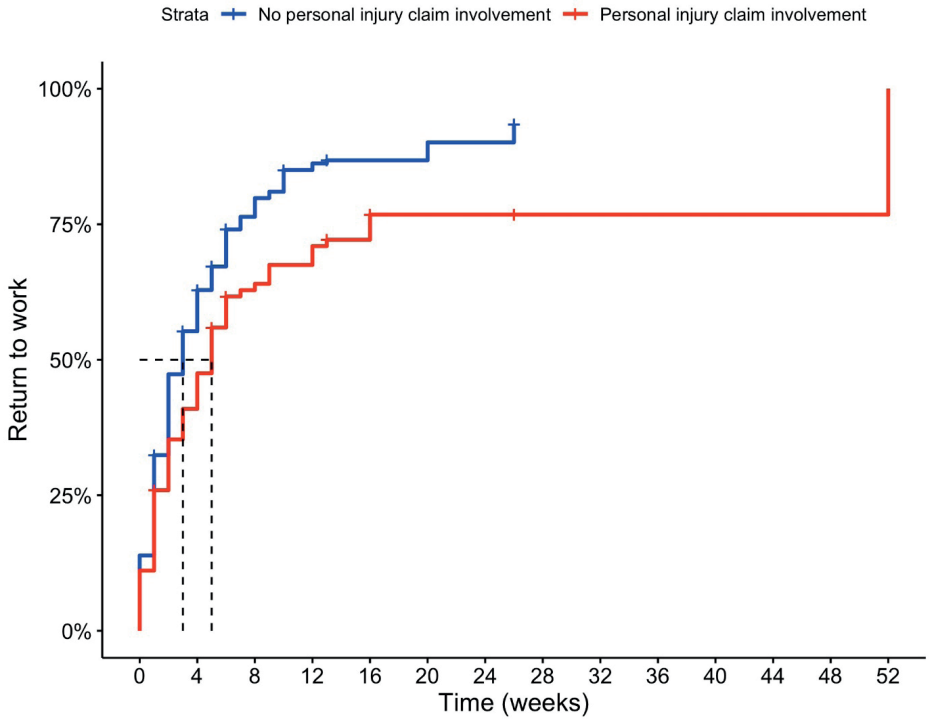


Figure 5. This Kaplan Meier curve represents the time to return to work after minor surgery for patients involved in a personal injury claim and matched controls. The dashed line indicates the median time to return to work.

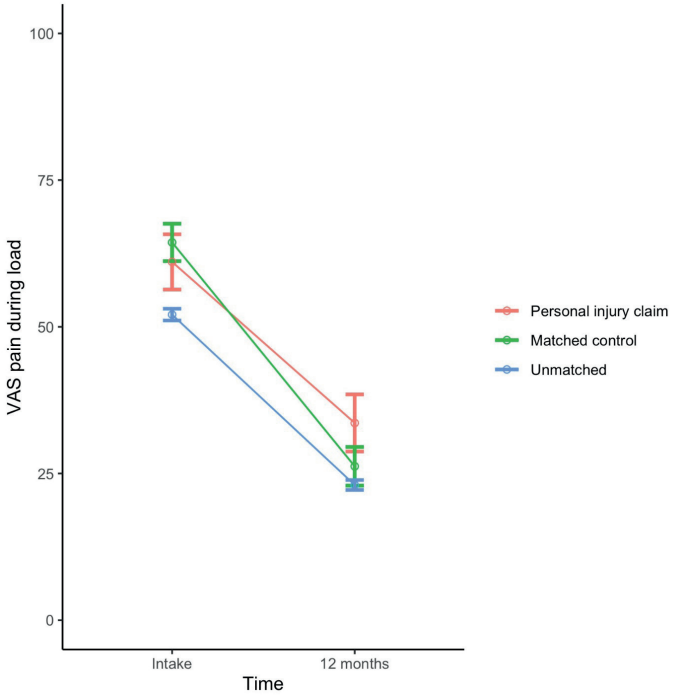

100
75
50
25
0
VAS pain during load
Intake
12 months
Time
Personal injury claim
Matched control
Unmatched

Figure 6A

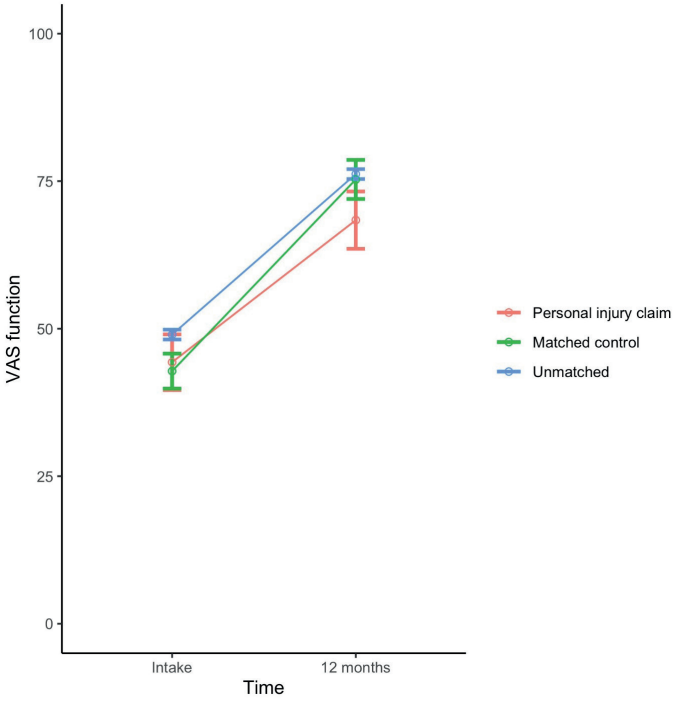


Figure 6B

Figure 6A-B. These graphs show the mean VAS scores for (A) pain and (B) hand function during load over time after major surgery. The error bars represent the 95% CI. Values are shown for patients involved in a personal injury claim, matched control patients, and unmatched patients not involved in a personal injury claim.

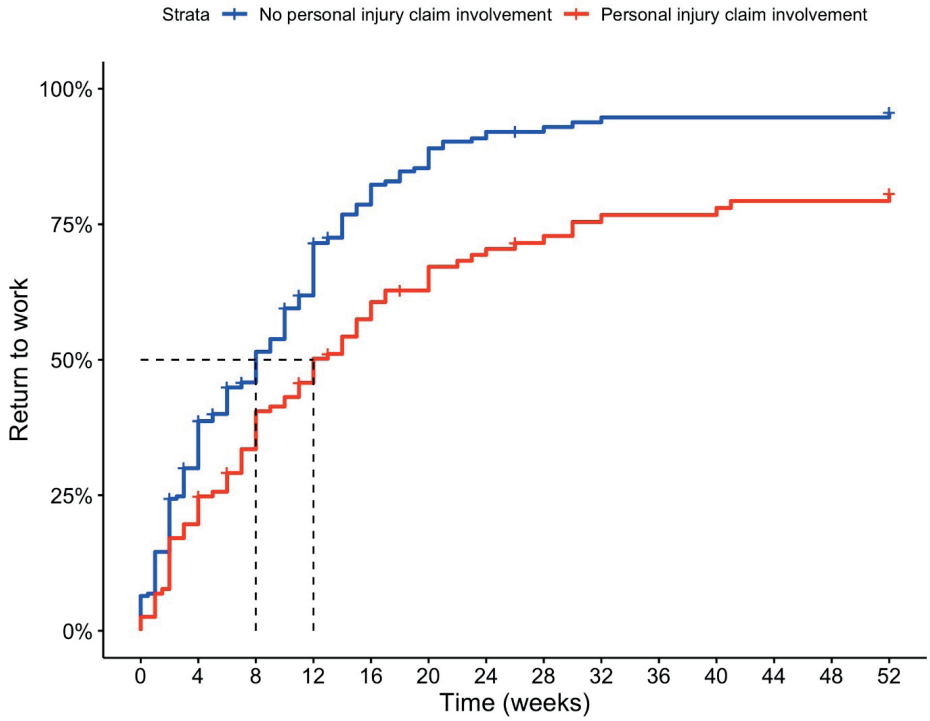


Figure 7. This Kaplan Meier curve represent the time to return to work after major surgery for patients involved in a personal injury claim and matched controls. The dashed line indicates the median time to return to work.

Table 1. Patient characteristics and VAS scores before treatment for patients with and without involvement in a personal injury claim

Parameter	**Patients with a personal injury claim (n = 651)**	**Patients without a personal injury claim (n = 28,450)**	**P-value**
Age in years, mean ± SD	46 ± 15	53 ± 15	< 0.001
Women, % (n)	62 (406)	64 (18,289)	0.57
Duration of symptoms in months, mean ± SD	17 ± 31	21 ± 46	0.02
Type of work, % (n)			0.003
Unemployed	28 (185)	34 (9782)	
Light physical labor	28 (183)	28 (8021)	
Moderate physical labor	28 (185)	26 (7286)	
Heavy physical labor	15 (98)	12 (3361)	
Dominant side, % (n)			0.44
Left	10 (64)	9 (2477)	
Right	86 (563)	88 (25,070)	
Ambidextrous	4 (24)	3 (903)	
Type of treatment, % (n)			< 0.001
Nonsurgical treatment	29 (189)	32 (9235)	
Minor surgery (e.g., release of the first extensor compartment)[a]	40 (259)	45 (12,698)	
Major surgery (e.g., TFCC reinsertion)[a]	31 (203)	23 (6517)	
VAS pain during load	66 ± 24	56 ± 28	< 0.001
VAS hand function	43 ± 24	51 ± 26	< 0.001

[a]For an overview of all treatments categorized as minor or major surgery, see Supplementary Table 1.

Table 2. Improvement over time (between baseline and 3 months) and the proportion of patients reaching the MIC for VAS pain and function [19] after nonsurgical treatment

Nonsurgical treatment	**Patients with a personal injury claim (n = 115)**	**Matched controls (n = 230)**	**P-value**
Change in VAS pain	13 ± 34	22 ± 30	0.01
Change in VAS function	12 ± 31	19 ± 31	0.048
Reaching the MIC of VAS pain	45 (52)	55 (127)	0.10
Reaching the MIC of VAS function	50 (57)	58 (134)	0.16

Data presented as % (n).
MIC = Minimally Important Change; VAS = Visual Analog Scale

Table 3. Improvement over time (between baseline and 3 months) and the proportion of patients reaching the MIC for VAS pain and function [19] after minor surgery

Minor surgery	**Patients with personal injury claim (n = 172)**	**Matched controls (n = 344)**	**P-value**
Change in VAS pain	25 ± 30	34 ± 30	0.001
Change in VAS function	20 ± 35	29 ± 32	0.005
Reaching the MIC of VAS pain	42 (72)	61 (210)	< 0.001
Reaching the MIC of VAS function	53 (91)	64 (220)	0.02

Data presented as % (n).
MIC = Minimally Important Change; VAS = Visual Analog Scale

Table 4. Improvement over time (between baseline and 12 months) and the proportion of patients reaching the MIC for VAS pain and function [19] after minor surgery

Major surgery	**Patients with personal injury claim (n = 129)**	**Matched controls (n = 258)**	**P-value**
Change in VAS pain	29 ± 34	39 ± 33	0.004
Change in VAS function	23 ± 33	34 ±33	0.003
Reaching the MIC of VAS pain, % (n)	55 (71)	67 (174)	0.02
Reaching the MIC of VAS function, % (n)	59 (76)	70 (180)	0.04

Data presented as % (n).
MIC = Minimally Important Change; VAS = Visual Analog Scale

Part IV

DEVELOPMENT, VALIDATION, AND ADDED VALUE OF PREDICTION MODELS IN HAND SURGERY

15

PREDICTING CLINICALLY RELEVANT PATIENT-REPORTED SYMPTOM IMPROVEMENT AFTER CARPAL TUNNEL RELEASE: A MACHINE LEARNING APPROACH

Hoogendam, L., Bakx, J. A. C., Souer, J. S., Slijper, H. P., Andrinopoulou, E. R., Selles, R. W., The Hand-Wrist Study Group (2022). Predicting Clinically Relevant Patient-Reported Symptom Improvement After Carpal Tunnel Release: A Machine Learning Approach. *Neurosurgery*, *90*(1), 106–113.

ABSTRACT

Background

Symptom improvement is an important goal when considering surgery for carpal tunnel syndrome. There is currently no prediction model available to predict symptom improvement for patients considering a carpal tunnel release (CTR).

Objective

Therefore, we developed a model to predict the probability of clinically-relevant symptom improvement at six months after CTR.

Methods

We split a cohort of 2119 patients who underwent a mini open CTR and completed the Boston Carpal Tunnel Questionnaire (BCTQ) preoperatively and six months postoperatively, into a training (75%) and validation dataset (25%). Patients who improved more than the minimal clinically important difference of 0.8 at the BCTQ-symptom severity scale were classified as "improved".

Logistic regression, Random Forests, and Gradient Boosting Machines were considered to train prediction models. The best model was selected based on discriminative ability (AUC) and calibration in the validation dataset. This model was further assessed in a hold-out dataset (N=397).

Results

A Gradient Boosting Machine with five predictors was chosen as optimal trade-off between discriminative ability and number of predictors. In the hold-out dataset, this model had an AUC of 0.723, good calibration, sensitivity of 0.77 and specificity of 0.55. Positive predictive value was 0.50 and negative predictive value was 0.81.

Conclusion

We developed a prediction model for clinically-relevant symptom improvement six months after a CTR, which required five patient-reported predictors (18 questions), has reasonable discriminative ability and good calibration. The model is available online and might help shared decision-making when patients are considering a CTR.

INTRODUCTION

Carpal tunnel syndrome (CTS) is a common peripheral nerve entrapment, with an incidence of 3.5 per 1000 person-years.[1] Symptoms include pain, tingling, and numbness in the thumb, index, and middle finger.[2] Conservative treatment options, such as a night splint or a corticosteroid injection, can be effective.[3] However, in severe cases or when conservative treatment fails, carpal tunnel release surgery (CTR) is indicated.[4]

While CTR is often considered a minor surgery, patients do not always experience it as such, and it can take months for patients to resume work.[5,6] Moreover, previous studies have shown that not all patients experience improved CTS symptoms after CTR.[7-9] Therefore, it would be very relevant to preoperatively identify which patients benefit from CTR.

Previous studies have investigated prognostic factors for outcomes of CTR and identified pretreatment symptom severity, other hand conditions and psychosocial factors as associated with outcomes of CTR.[10-13] While this knowledge is useful, it might be difficult to translate the effect of prognostic factors to decision-making for individual patients. Additionally, it has not been studied how well these factors predict outcomes of CTR. Therefore, models that predict outcomes for individual patients are needed to help patients and clinicians in their decision-making.

Besides classical statistical techniques, machine learning algorithms are available and are increasingly used to build prediction models.[14,15] Machine learning algorithms differ from classical techniques because of their ability to better identify non-linear effects, interactions, or other patterns, particularly in large datasets.[16] These properties can lead to better predictive performance of prediction models trained using machine learning algorithms.[17,18]

Therefore, the aim of this study was to develop and validate a prediction model using one statistical algorithm and two machine learning algorithms to predict the probability of clinically-relevant patient-reported symptom improvement for individual patients six months after CTR.

METHODS

Source of data

Data from the Hand and Wrist cohort was used. This cohort and data collection have previously been described in more detail.[19] Reporting was done following the TRIPOD statement.[20] We used data collected from November 2011 to November 2019 as development dataset. Data collected between November 29th, 2019 and May 6th, 2021 were used as a hold-out test dataset.

Participants

All patients were invited to complete patient-reported outcome measures (PROMs) preoperatively and postoperatively as part of routine outcome measurements. We included all patients who were surgically treated with a mini-open CTR[21] at Xpert Clinics between 2011 and 2019 and who completed at least the Boston Carpal Tunnel Questionnaire (BCTQ) before and six months after surgery. Xpert Clinics comprises 25 locations and 23 European Board certified hand surgeons. All patients provided written informed consent. Ethics approval from the local ethics board was obtained.

Outcome

The BCTQ is a validated PROM specifically developed for CTS.[22] The BCTQ contains the subscales "Symptom severity scale (SSS)" and "Functional status scale (FSS)", consisting of eleven and eight questions, respectively. Scores for each subscale range from 1-5 (1=no functional impairment, 5=extreme functional impairment). We calculated a change score for each subscale and dichotomized these change scores as "reached minimal clinically important difference (MCID)" or "did not reach MCID". A threshold of 0.8 was used as the MCID for the BCTQ-SSS, as previously suggested.[23]

Predictors

Patient characteristics, medical history, and baseline PROM scores were used as predictors (Supplemental Table 1). Besides the BCTQ, patients treated in our clinic are invited to complete questionnaires on medical history, pain and hand function, and psychosocial factors. Pain and hand function are measured using Visual Analog Scales (0-100). From September 2017 onwards, we invited patients to complete the Euro-QoL-5D-5L (EQ5D), the Credibility and Expectancy questionnaire (CEQ), the Patient Health Questionnaire-4 (PHQ), the Pain Catastrophizing Scale (PCS) and the Brief Illness Perception Questionnaire (B-IPQ) to measure mindset.[24-28] The subscales of these questionnaires were used in the analysis.

Sample size

The number of patients treated until December 2019 determined the sample size of the development dataset. For the test dataset, patients were included until this dataset consisted of at least 10% of the patients in the development dataset.

Missing data

We expected missingness in the mindset questionnaires, since these data are only collected since September 2017. Additionally, missing data were expected in medical history, since some patients complete this questionnaire on paper in the waiting room. Because these data are most likely missing at random, and data on mindset and medical history might improve predictions, we imputed these. To help distinguish between missing data of patients who were not invited to complete the mindset questionnaires and patients who were invited, but did not participate, we created an indicator variable, which was used during imputation. Previously, it has been reported that datasets with up to 90% missingness can be safely imputed with multiple imputation.[29] Variables with more than 90% missing were excluded from analysis rather than imputed (Supplemental Table 1).

We imputed missing data using "K-nearest neighbors (KNN)"-imputation from the Caret package,[30] with K=5. We tested whether imputation would negatively affect predictive performance of the models, which was not the case (Supplemental material A). We therefore used the above-mentioned imputation in all models.

Statistical analysis

The development dataset was randomly split in a training (75%) and validation dataset (25%). After splitting the data, standardization and KNN-imputation was performed. Predictors with near-zero variance were removed (Supplemental Table 1). Three algorithms (one statistical algorithm and two machine learning algorithms) were considered: logistic regression, Random Forest, and Gradient Boosting Machines, respectively. We selected Random Forest and Gradient Boosting Machines because they are frequently used, reasonably intuitive because of similarities with decision trees, and good performance has been reported.[31-33]

The two outcome classes (reached MCID or not) did not occur equally often, therefore downsampling and upsampling were incorporated into the variable selection to allow the algorithms to also learn from the minority class.[34,35] Variable selection was performed using recursive feature elimination with 5-repeats of 10-fold cross-validation. This is a type of backwards selection, where variables are ranked on their importance.[36] For each algorithm, from the three recursive feature elimination models trained with either upsampling, downsampling, or without sampling, the most promising (lowest number of variables and

best discriminative ability (AUC)) was selected for further performance evaluation (Figure 1). DeLong tests were used to detect significant differences in AUC.

For each algorithm, one model with the variables selected in recursive feature elimination was fitted on the original training dataset. The models were optimized by selecting the best values for the hyperparameters (e.g., number of trees in a random forest), which were obtained from 5-repeats of 10-fold cross-validation. Afterwards, the performance of the models was assessed in the validation dataset in terms of AUC and calibration. Calibration was visually assessed using calibration belts[37], which is a method to assess reliability of the predictions.

We then compared performance of the three optimized models (i.e., one for each algorithm) to establish which model had the best AUC while also having a limited number of predictors and good calibration. This model was further evaluated on a test dataset that was held out from the start of model building. In this test dataset, we evaluated model performance in terms of AUC value, calibration, sensitivity, and specificity.

We present the final model as online app, which was created in R Shiny (R Studio). We followed the recommendations of the Dutch National Healthcare Institute on presenting quality information, such as PROMs, to consumers.[38] This resulted in presenting the predicted probability of improvement using population icons. We designed the app for joint usage by clinicians and patients during the consultation, as part of the shared decision-making process, to allow the clinician to answer any questions the patient has. The online app currently does not have CE-marking.

Analyses were performed in R statistical programming (The R project for statistical computing), version 3.6.3. Prediction models were trained in the Caret package version 6.0-86. A p-value < 0.05 was considered statistically significant.

RESULTS

Between November 2011 and November 28th, 2019, 5095 patients were treated with CTR in our clinic. After applying the eligibility criteria, 2119 patients were included in the training dataset (N=1590) and validation dataset (N=529). Patient characteristics in the three groups are shown in Table 1. Patient-reported outcomes at baseline and six months postoperatively are reported in Table 2 and 3, respectively. 1152 patients (72.4%) in the training dataset and 384 patients (72.6%) in the validation dataset reached the MCID threshold of 0.8 on the BCTQ-SSS. The hold-out test dataset consisted of the 397 most recently treated patients, who may be more representable of new patients considering CTR.

Recursive feature elimination with either upsampling, downsampling or without sampling, resulted in three models per algorithm with 5–30 predictors (Table 4). All models identified

baseline BCTQ-SSS as the most important variable. Based on performance during cross-validation in the training dataset, we chose for each algorithm which variables should be used for model training. After training three models (one for each algorithm) on the original training dataset, AUC in the validation dataset ranged between 0.735–0.782 (Table 4). Sensitivity ranged between 0.66–0.91 and specificity ranged between 0.46–0.77. For all models, visual inspection of calibration belts (Supplementary Figure 1ABC) showed adequate calibration in the validation dataset.

The downsampled Gradient Boosting Machine model required the lowest number of predictors, while having the highest AUC (0.782) in the validation dataset, and was therefore selected for further testing on a held-out test dataset of new patients. In the hold-out test dataset, AUC of this model was 0.723, sensitivity was 0.77, and specificity was 0.55 at a threshold of 0.75 (Table 5). Positive and negative predictive value were 0.50 and 0.81, respectively. This model's calibration belt indicates good calibration (Figure 2).

The final model is presented as Shiny web application, accessible via: https://analyse.equipezorgbedrijven.nl/shiny/cts-prediction-model-Eng/. On the left, all mandatory variables can be filled in. The model requires five PROM scores as mandatory input. In case one or more scores are unknown, they can be calculated by completing the individual questions (maximum of 18 when all scores are unknown). After clicking "Calculate your prediction", a prediction for the probability of at least 0.8 improvement on the BCTQ-SSS will be generated for an individual patient, based on the information supplied.

DISCUSSION

In this study, we developed and validated a model to predict the probability of clinically-relevant patient-reported symptom improvement at six months after CTR. The best model was a Gradient Boosting Machine with five baseline PROM scores as predictors. The model had a reasonable discriminative ability, good calibration, high sensitivity and reasonable specificity in a hold-out test dataset.

Previous studies have investigated outcome prediction after CTR.[11,39-42] However, these studies generally had a small sample size with a maximum of 145 patients.[39-41] None of these studies validated their model on a separate dataset, which is essential before a model can be used in clinical practice.[43] In contrast to these studies, Bowman et al.[44] have developed and validated two prediction models for self-reported improvement one to three years after CTR. These prediction models, however, each require over twenty predictors, including nerve conduction outcomes, and have an AUC of 0.69–0.70. The model we developed has a similar discriminative ability (AUC 0.72) while requiring only five PROM scores as predictors.

Previously, similar prediction models have been developed for clinically-relevant improvement following lumbar spinal stenosis decompression (AUC 0.68–0.79),[45] surgery for degenerative cervical myelopathy (AUC 0.70),[46] and total joint arthroplasty (AUC 0.78-0.97).[47-49] While the performance of our model is comparable to Siccoli et al.,[45] and Bowman et al.,[44] some prediction models currently available in orthopaedics appear to have better discrimination.[47-49] A possible explanation could be that in general, larger datasets are available in the field of orthopaedics. Alternatively, it could be that treatment outcomes of CTR vary more than outcomes of total joint arthroplasty, and are therefore more difficult to accurately predict.

Strengths and limitations

Strengths of this study were that we developed a prediction model for symptom improvement after CTR on a large real-life dataset, which leads to high generalizability. Additionally, we validated this model on a hold-out test dataset of the most recently treated patients, therefore testing how the model would perform when used for new patients considering CTR. Moreover, the developed model is publicly available online.

However, this study also has some limitations. We defined clinically-relevant improvement in symptoms as reaching at least the MCID of 0.8 on the BCTQ-SSS.[23] However, an improvement of 0.8 may not be equally relevant for all individual patients. We therefore believe that is important to communicate what our prediction model exactly predicts. Additionally, because this MCID threshold has been calculated for a large group of CTR patients, we believe that it is still a relevant outcome for most patients.

The probability of reaching the MCID is dependent on the baseline score.[47,50] There are studies that provide threshold values of baseline scores as decision aid in the choice for surgery.[51-53] Indeed, our models identified baseline BCTQ-SSS as most important variable. Still, we argue that it makes sense to use this data to predict the probability of clinically-relevant improvement since the baseline BCTQ-SSS is preoperatively available to clinicians and indicates symptom severity.

Additionally, missing data affected the dataset used to develop this prediction model. All patients treated in our clinic are invited to complete PROMs, which leads to a higher non-response proportion than may be expected in a controlled trial. However, in our opinion, these data reflect daily clinical practice. Moreover, the psychosocial factors had a large proportion of missing data, mainly because they have only been collected since 2017. Because we believed that these factors could be relevant for the prediction model, we applied imputation. Even though the missing data mechanisms might differ between the development dataset and the more recent hold-out test dataset, model performance was still reasonable in the hold-out dataset, which best represents new patients considering CTR. We therefore believe this will not be an issue.

In our dataset, patient demographics, patient-reported outcomes, and certain psychosocial variables were available. Other potentially relevant variables, such as nerve conduction outcomes or neurological examination, were not available. During variable selection, only patient-reported outcomes were selected as best predictors, which may be counterintuitive for clinicians. Further studies comparing accuracy of predictions made by clinicians to predictions made by prediction models may help to assess the ecological validity of our prediction model. Moreover, adding clinical variables might further improve the performance of our model. Possibly, our model can be updated and validated further in independent datasets with other relevant variables available.[54,55]

Implications and future research

Previously, pretreatment expectations have been related to treatment outcomes, such as satisfaction.[56] Kadzielski et al., therefore, conclude that patients should receive realistic information on the expected outcome of treatment. Our prediction model helps to provide patients with realistic, individualized outcome information for patients considering CTR.

Possibly, providing patients with a prediction of their outcome might impact pretreatment expectations, patients' experiences with the treatment, and possibly even treatment outcomes. Whether providing patients with an individualized prediction affects preoperative and postoperative patient-reported outcomes should be evaluated in future studies.[57]

CONCLUSION

In this study, we developed and validated a prediction model for the probability of patient-reported improvement in symptoms six months after CTR. The model had a reasonable discriminative ability, good calibration, high sensitivity and reasonable specificity in a hold-out test dataset. The model is available online. Possibly, using the prediction model in pretreatment counselling will help shared decision making and expectation management.

ACKNOWLEDGEMENTS

The following collaborators are members of the Hand-Wrist Study Group.

R Arjen M Blomme MD, Berbel JR Sluijter MD PhD, Dirk-Jan JC van der Avoort MD, Alexander Kroeze MD, Jeroen Smit MD PhD, Jan Debeij MD PhD, Erik T Walbeehm MD PhD, Gijs M van Couwelaar MD, Guus M Vermeulen MD PhD, Hans de Schipper MD, Hans Temming MD, Jeroen H van Uchelen MD PhD, H Luitzen de Boer MD, Nicoline de Haas MD, Kennard Harmsen MD, Oliver T Zöphel MD PhD, Reinier Feitz MD, Richard Koch MD, Steven ER Hovius MD PhD, Thybout M Moojen MD PhD, Xander Smit MD

PhD, Rob van Huis PT, Pierre Y Pennehouat PT, Karin Schoneveld PT MSc, Yara E van Kooij PT MSc, Robbert M Wouters PT PhD, Paul Zagt PT, Folkert J van Ewijk PT, Joris J Veltkamp PT, Alexandra Fink PT, Willemijn A de Ridder PT MSc, Miguel C Jansen MD PhD, Mark JW van der Oest PhD, Pepijn O Sun MD, Joris S Teunissen BSc, Jak Dekker MSc, Marlies L Jansen-Landheer MD MSc, Marloes HP ter Stege MSc

REFERENCES

1. Nordstrom DL, DeStefano F, Vierkant RA, Layde PM. Incidence of diagnosed carpal tunnel syndrome in a general population. *Epidemiology.* 1998;9(3):342-345.

2. Middleton SD, Anakwe RE. Carpal tunnel syndrome. *BMJ.* 2014;349:g6437.

3. Shi Q, MacDermid JC. Is surgical intervention more effective than non-surgical treatment for carpal tunnel syndrome? A systematic review. *J Orthop Surg Res.* 2011;6:17.

4. Bland JD. Carpal tunnel syndrome. *BMJ.* 2007;335(7615):343-346.

5. Newington L, Brooks C, Warwick D, Adams J, Walker-Bone K. Return to work after carpal tunnel release surgery: a qualitative interview study. *BMC Musculoskelet Disord.* 2019;20(1):242.

6. Newington L, Harris EC, Walker-Bone K. Carpal tunnel syndrome and work. *Best Pract Res Clin Rheumatol.* 2015;29(3):440-453.

7. Katz JN, Keller RB, Simmons BP, et al. Maine Carpal Tunnel Study: outcomes of operative and nonoperative therapy for carpal tunnel syndrome in a community-based cohort. *J Hand Surg Am.* 1998;23(4):697-710.

8. Nancollas MP, Peimer CA, Wheeler DR, Sherwin FS. Long-term results of carpal tunnel release. *J Hand Surg Br.* 1995;20(4):470-474.

9. Tang CQY, Lai SWH, Tay SC. Patient-reported outcomes of carpal tunnel release surgery in patients with bilateral severe carpal tunnel syndrome. *J Hand Surg Eur Vol.* 2017;42(9):932-936.

10. Alimohammadi E, Bagheri SR, Hadidi H, Rizevandi P, Abdi A. Carpal tunnel surgery: predictors of clinical outcomes and patients' satisfaction. *BMC Musculoskelet Disord.* 2020;21(1):51.

11. Jansen MC, Evers S, Slijper HP, et al. Predicting Clinical Outcome After Surgical Treatment in Patients With Carpal Tunnel Syndrome. *J Hand Surg Am.* 2018;43(12):1098-1106 e1091.

12. Turner A, Kimble F, Gulyas K, Ball J. Can the outcome of open carpal tunnel release be predicted?: a review of the literature. *Anz J Surg.* 2010;80(1-2):50-54.

13. Fernandez-de-Las-Penas C, de-la-Llave-Rincon AI, Cescon C, Barbero M, Arias-Buria JL, Falla D. Influence of Clinical, Psychological, and Psychophysical Variables on Long-term Treatment Outcomes in Carpal Tunnel Syndrome: Evidence From a Randomized Clinical Trial. *Pain Pract.* 2019;19(6):644-655.

14. Cabitza F, Locoro A, Banfi G. Machine Learning in Orthopedics: A Literature Review. *Front Bioeng Biotechnol.* 2018;6:75.

15. Myers TG, Ramkumar PN, Ricciardi BF, Urish KL, Kipper J, Ketonis C. Artificial Intelligence and Orthopaedics: An Introduction for Clinicians. *J Bone Joint Surg Am.* 2020;102(9):830-840.

16. Bayliss L, Jones LD. The role of artificial intelligence and machine learning in predicting orthopaedic outcomes. *Bone Joint J.* 2019;101-B(12):1476-1478.

17. Fernandez-Delgado M, Cernadas E, Barro S, Amorim D. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *J Mach Learn Res.* 2014;15:3133-3181.

18. Liu NT, Salinas J. Machine Learning for Predicting Outcomes in Trauma. *Shock.* 2017;48(5):504-510.

19. Selles RW, Wouters RM, Poelstra R, et al. Routine health outcome measurement: Development, design, and implementation of the Hand and Wrist Cohort. *Plast Reconstr Surg.* 2020.

20. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ.* 2015;350:g7594.

21. van den Broeke LR, Theuvenet WJ, van Wingerden JJ. Effectiveness of mini-open carpal tunnel release: An outcome study. *Arch Plast Surg.* 2019;46(4):350-358.

22. Levine DW, Simmons BP, Koris MJ, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Joint Surg Am.* 1993;75(11):1585-1592.

23. Leite JC, Jerosch-Herold C, Song F. A systematic review of the psychometric properties of the Boston Carpal Tunnel Questionnaire. *BMC Musculoskelet Disord.* 2006;7:78.

24. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. *J Psychosom Res.* 2006;60(6):631-637.

25. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry.* 2000;31(2):73-86.

26. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.* 2011;20(10):1727-1736.

27. Kroenke K, Spitzer RL, Williams JB, Lowe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics.* 2009;50(6):613-621.

28. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. *Psychol Assessment.* 1995;7(4):524-532.

29. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol.* 2019;110:63-73.

30. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw.* 2008;28(5):1-26.

31. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5-32.

32. Friedman JH. Stochastic gradient boosting. *Comput Stat Data An.* 2002;38(4):367-378.

33. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak.* 2019;19(1):281.

34. Japkowicz N. Learning from imbalanced data sets: A comparison of various strategies. *AAAI Technical Report.* 2000.

35. Provost F. Machine learning from imbalanced datasets 101. *AAAI Technical Report.* 2000.

36. Kuhn M, Johnson K. *Feature engineering and selection : a practical approach for predictive models.* 2021.

37. Finazzi S, Poole D, Luciani D, Cogo PE, Bertolini G. Calibration belt for quality-of-care assessment based on dichotomous outcomes. *PLoS One.* 2011;6(2):e16110.

38. Damman OC, Malanda UL. Handreiking presentatie van kwaliteitsinformatie aan consumenten. 2015; https://www.zorginzicht.nl/binaries/content/assets/zorginzicht/ontwikkeltools-ontwikkelen/Handreiking_presentatie_van_kwaliteitsinformatie_aan_consumenten_V1+0.pdf. Accessed 23-03-2021, 2021.

39. Beck JD, Brothers JG, Maloney PJ, Deegan JH, Tang XQ, Klena JC. Predicting the Outcome of Revision Carpal Tunnel Release. *J Hand Surg-Am.* 2012;37a(2):282-287.

40. Sun PO, Selles RW, Jansen MC, Slijper HP, Ulrich DJO, Walbeehm ET. Recurrent and persistent carpal tunnel syndrome: predicting clinical outcome of revision surgery. *J Neurosurg.* 2019:1-9.

41. Bland JDP, Rudolfer SM. Ultrasound Imaging of the Median Nerve as a Prognostic Factor for Carpal Tunnel Decompression. *Muscle & Nerve*. 2014;49(5):741-744.

42. Watchmaker JD, Watchmaker GP. Independent Variables Affecting Outcome of Carpal Tunnel Release Surgery. *Hand (N Y)*. 2017:1558944717703739.

43. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14:40.

44. Bowman A, Rudolfer S, Weller P, Bland JDP. A prognostic model for the patient-reported outcome of surgical treatment of carpal tunnel syndrome. *Muscle Nerve*. 2018;58(6):784-789.

45. Siccoli A, de Wispelaere MP, Schroder ML, Staartjes VE. Machine learning-based preoperative predictive analytics for lumbar spinal stenosis. *Neurosurg Focus*. 2019;46(5):E5.

46. Merali ZG, Witiw CD, Badhiwala JH, Wilson JR, Fehlings MG. Using a machine learning approach to predict outcome after surgery for degenerative cervical myelopathy. *PLoS One*. 2019;14(4):e0215133.

47. Fontana MA, Lyman S, Sarker GK, Padgett DE, MacLean CH. Can Machine Learning Algorithms Predict Which Patients Will Achieve Minimally Clinically Important Differences From Total Joint Arthroplasty? *Clin Orthop Relat R*. 2019;477(6):1267-1279.

48. Kumar V, Roche C, Overman S, et al. What Is the Accuracy of Three Different Machine Learning Techniques to Predict Clinical Outcomes After Shoulder Arthroplasty? *Clin Orthop Relat R*. 2020;478(10):2351-2363.

49. Kunze KN, Karhade AV, Sadauskas AJ, Schwab JH, Levine BR. Development of Machine Learning Algorithms to Predict Clinically Meaningful Improvement for the Patient-Reported Health State After Total Hip Arthroplasty. *J Arthroplasty*. 2020;35(8):2119-2123.

50. Copay AG, Subach BR, Glassman SD, Polly DW, Jr., Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J*. 2007;7(5):541-546.

51. Berliner JL, Brodke DJ, Chan V, SooHoo NF, Bozic KJ. Can Preoperative Patient-reported Outcome Measures Be Used to Predict Meaningful Improvement in Function After TKA? *Clin Orthop Relat Res*. 2017;475(1):149-157.

52. Conti MS, Caolo KC, Nguyen JT, Deland JT, Ellis SJ. Preoperative Patient-Reported Outcome Measures Relationship With Postoperative Outcomes in Flexible Adult-Acquired Flatfoot Deformity. *Foot Ankle Int*. 2021;42(3):268-277.

53. Nwachukwu BU, Beck EC, Kunze KN, Chahla J, Rasio J, Nho SJ. Defining the Clinically Meaningful Outcomes for Arthroscopic Treatment of Femoroacetabular Impingement Syndrome at Minimum 5-Year Follow-up. *Am J Sports Med*. 2020;48(4):901-907.

54. Davis SE, Greevy RA, Lasko TA, Walsh CG, Matheny ME. Comparison of Prediction Model Performance Updating Protocols: Using a Data-Driven Testing Procedure to Guide Updating. *AMIA Annu Symp Proc*. 2019;2019:1002-1010.

55. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98(9):683-690.

56. Kadzielski J, Malhotra LR, Zurakowski D, Lee SG, Jupiter JB, Ring D. Evaluation of preoperative expectations and patient satisfaction after carpal tunnel release. *J Hand Surg Am*. 2008;33(10):1783-1788.

57. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagn Progn Res*. 2018;2:11.

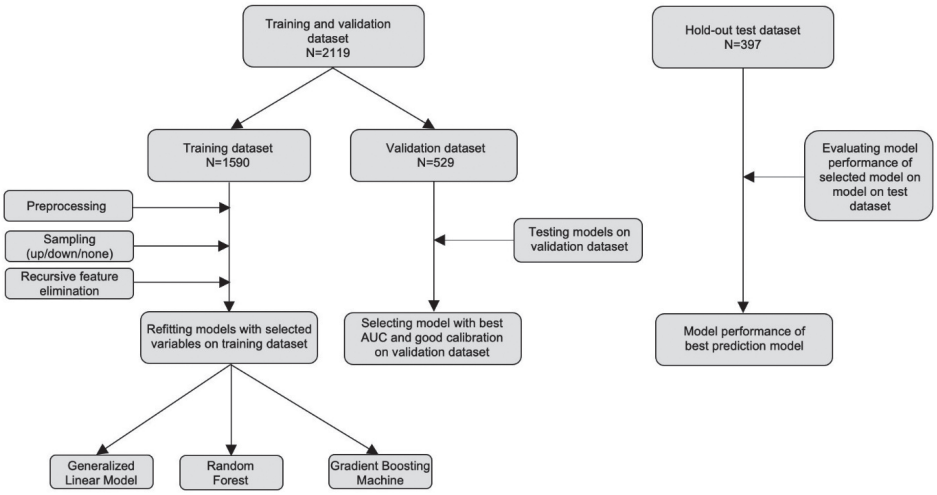


Figure 1. Flowchart of training, selecting, and validating prediction models

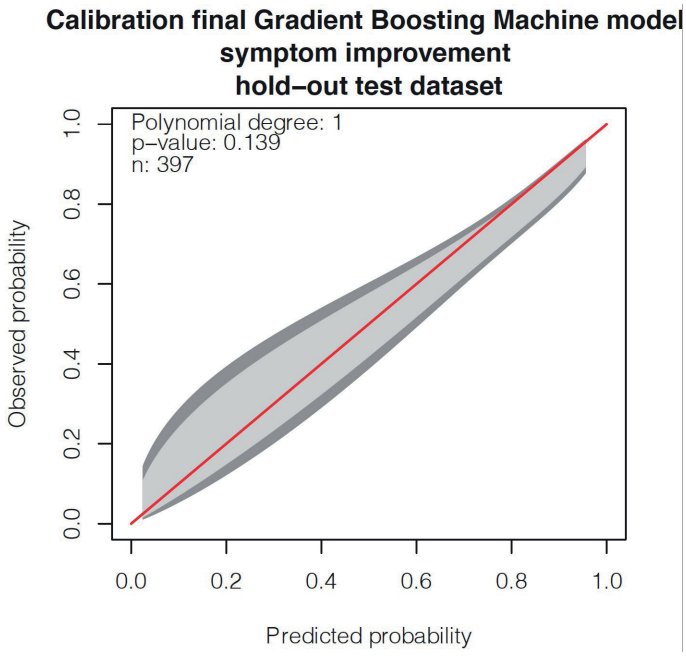


Figure 2. Calibration belt of the selected prediction model in the hold-out test dataset. The predicted probability of symptom improvement compared to the observed probability of symptom improvement is shown. Perfect calibration is indicated as red line and the confidence interval of the calibration of the prediction model is shown in grey. The entire red line falls within the confidence interval, indicating no significant deviation in calibration between the predicted and observed probability of improvement.

Table 1. Patient characteristics in train, validation, and test data.

	Training dataset (N=1590)	% missing	Validation dataset (N=529)	% missing	Test dataset (N=397)	% missing
Age (in years)	55.6 (12.9)	0	55.6 (12.7)	0	55.8 (12.9)	0
Sex (%)		0		0		0
Female	1135 (71.4)		365 (69.0)		278 (70.0)	
Duration of symptoms (in months)	24.9 (37.4)	0	29.5 (55.0)	0	21.1 (29.4)	0
Primary CTS (%)		0		0		0.3
Yes	1442 (90.7)		486 (91.9)		359 (90.7)	
Hand dominance (%)		0		0		0.3
Right	1409 (88.6)		464 (87.7)		349 (88.1)	
Left	137 (8.6)		44 (8.3)		40 (10.1)	
Both	44 (2.8)		21 (4.0)		7 (1.8)	
Affected side (%)		0		0		0
Right	930 (58.5)		311 (58.8)		216 (54.4)	
Occupational intensity (%)		0		0		0
Not employed	602 (37.9)		206 (38.9)		154 (38.8)	
Light	333 (20.9)		125 (23.6)		88 (22.2)	
Moderate	430 (27.0)		133 (25.1)		108 (27.2)	
Severe	225 (14.2)		65 (12.3)		47 (11.8)	
Body Mass Index	27.5 (5.2)	13.6	27.5 (4.7)	11.9	28.0 (4.8)	11.6
Diabetes mellitus (%)		13.6		11.9		11.6
Yes	85 (6.2)		23 (5.2)		29 (8.3)	
Smoking (%)		30.0		26.5		16.1
Never	580 (52.1)		193 (50.4)		150 (45.0)	

* Values reported as mean (SD) unless stated otherwise
CTS = Carpal tunnel syndrome; SD = Standard deviation

Table 2. Patient-reported outcome scores at baseline

	Range	Training dataset (N=1590)	% missing	Validation dataset (N=529)	% missing	Test dataset (N=397)	% missing
BCTQ SSS	1-5	2.85 (0.65)	0	2.88 (0.65)	0	2.89 (0.69)	0
BCTQ FSS	1-5	2.46 (0.78)	0	2.44 (0.78)	0	2.48 (0.81)	0
VAS average pain	0-100	53 (25)	2.5	53 (26)	2.1	51 (26)	2.0
VAS pain at rest	0-100	49 (26)	2.5	48 (26)	2.1	46 (26)	2.0
VAS pain during load	0-100	57 (27)	2.5	56 (26)	2.1	57 (28)	2.0
VAS hand function	0-100	47 (24)	2.5	46 (24)	2.1	50 (24)	2.0
VAS satisfaction	0-100	30 (23)	2.5	30 (22)	2.1	35 (25)	2.0
EQ5D mobility	1-5	1.27 (0.65)	66.2	1.37 (0.78)	60.9	1.33 (0.72)	3.0
EQ5D selfcare	1-5	1.33 (0.67)	66.2	1.37 (0.67)	60.9	1.39 (0.69)	3.0
EQ5D activities of daily living	1-5	2.28 (1.00)	66.2	2.32 (1.07)	60.9	2.36 (1.03)	3.0
EQ5D pain and discomfort	1-5	2.80 (0.85)	66.2	2.79 (0.89)	60.9	2.78 (0.87)	3.0
EQ5D anxiety and depression	1-5	1.34 (0.72)	66.2	1.41 (0.72)	60.9	1.32 (0.68)	3.0
PCS magnification	0-24	2.0 (2.2)	70.3	1.9 (2.2)	65.2	1.9 (2.2)	4.0
PCS helplessness	0-12	6.1 (5.1)	70.3	5.7 (4.9)	65.2	5.9 (5.2)	4.0
PCS rumination	0-16	5.2 (3.9)	70.3	4.9 (3.7)	65.2	5.1 (3.9)	4.0
PHQ-4 depression	0-6	0.77 (1.33)	70.4	0.64 (1.15)	64.7	0.70 (1.32)	2.3
PHQ-4 anxiety	0-6	0.92 (1.48)	70.4	0.74 (1.20)	64.7	0.83 (1.43)	2.3
CEQ expectancy	3-27	23.4 (3.2)	70.3	23.3 (3.2)	65.0	23.3 (3.3)	3.0
CEQ credibility	3-27	23.9 (3.1)	70.3	23.9 (2.9)	65.0	24.2 (2.7)	3.0
B-IPQ consequences	1-10	6.9 (2.3)	70.1	6.8 (2.2)	64.7	7.0 (2.1)	4.0
B-IPQ timeline	1-10	4.9 (2.5)	70.1	4.6 (2.4)	64.7	5.1 (2.4)	4.0
B-IPQ personal control	1-10	4.1 (2.5)	70.1	4.1 (2.5)	64.7	4.2 (2.3)	4.0
B-IPQ treatment control	1-10	8.5 (1.3)	70.1	8.5 (1.4)	64.7	8.5 (1.3)	4.0
B-IPQ identity	1-10	6.6 (2.3)	70.1	6.7 (2.1)	64.7	6.6 (2.4)	4.0
B-IPQ concern	1-10	5.6 (2.7)	70.1	5.8 (2.5)	64.7	5.5 (2.8)	4.0
B-IPQ understanding	1-10	8.0 (2.2)	70.1	8.1 (2.0)	64.7	8.1 (2.0)	4.0
B-IPQ emotional response	1-10	4.5 (3.1)	70.1	4.6 (3.0)	64.7	4.3 (3.0)	4.0

* Values reported as mean (SD) unless stated otherwise

BCTQ = Boston carpal tunnel syndrome; B-IPQ = Brief illness perception questionnaire; CEQ = Credibility/Expectancy Questionnaire; EQ5D = Euro-QoL-5D-5L; FSS = Functional status scale; PCS = Pain catastrophizing scale; PHQ = Patient health questionnaire-4; SD = Standard deviation; SSS = Symptom severity scale; VAS = Visual analog scale

Table 3. Patient-reported outcome scores six months postoperatively.

	Range	Training dataset (N=1590)	% missing	Validation dataset (N=529)	% missing	Test dataset (N=397)	% missing
BCTQ SSS	1-5	1.60 (0.62)	0	1.60 (0.63)	0	1.67 (0.73)	0
BCTQ FSS	1-5	1.61 (0.68)	0	1.58 (0.67)	0	1.66 (0.72)	0
VAS average pain	0-100	17 (23)	0.9	17 (23)	1.3	20 (26)	1.3
VAS pain at rest	0-100	14 (22)	0.9	14 (21)	1.3	17 (26)	1.3
VAS pain during load	0-100	21 (26)	0.9	22 (26)	1.3	24 (29)	1.3
VAS hand function	0-100	75 (29)	0.9	76 (28)	1.3	77 (27)	1.3
VAS satisfaction	0-100	74 (30)	0.9	75 (28)	1.3	75 (30)	1.3

* Values reported as mean (SD) unless stated otherwise
BCTQ = Boston carpal tunnel syndrome; FSS = Functional status scale; SD = Standard deviation; SSS = Symptom severity scale; VAS = Visual analog scale

Table 4. Model properties in training and validation dataset. For each algorithm, three recursive feature elimination models are trained using different sampling strategies. From these three models per algorithm, one model is selected for further evaluation (AUC value, sensitivity, and specificity in bold) based on number of predictors and AUC. P-values are obtained from comparisons in AUC between the model with the highest AUC (downsampled Gradient Boosting Machine, reference) and the two models that were not selected for further evaluation.

Training method		Results					
Algorithm	**Sampling technique**	**Number of variables**	**AUC (SD) in cross-validation in training dataset**	**AUC in validation dataset**	**P-value**	**Sensitivity in validation dataset**	**Specificity in validation dataset**
Generalized Linear Model	Upsampling	6	**0.7452 (0.036)**	0.7809	0.9473	0.66	0.77
	Downsampling	9	0.7462 (0.038)				
	No sampling	10	0.7465 (0.038)				
Random Forest	Upsampling	30	**0.7217 (0.042)**	0.7350	0.01763	0.91	0.46
	Downsampling	71	0.7157 (0.047)				
	No sampling	71	0.7099 (0.042)				
Gradient Boosting Machine	Upsampling	71	0.7337 (0.043)				
	Downsampling	5	**0.7385 (0.040)**	**0.7820**	**Reference**	**0.84**	**0.55**
	No sampling	15	0.7417 (0.041)				

AUC = Area under the curve; SD = Standard deviation

Table 5. Model properties in test dataset of the selected prediction model

Training method		Results						
Algorithm	**Sampling technique**	**Number of variables**	**AUC in test dataset**	**Sensitivity in test dataset**	**Specificity in test dataset**	**Threshold for improvement/no improvement**	**Positive predictive value**	**Negative predictive value**
Gradient Boosting Machine	Downsampling	5	0.7229	0.77	0.55	0. 75	0.50	0.81

AUC = Area under the curve

16

MACHINE LEARNING CAN BE USED TO PREDICT FUNCTION BUT NOT PAIN AFTER SURGERY FOR THUMB CARPOMETACARPAL OSTEOARTHRITIS

Loos, N. L., Hoogendam, L., Souer, J. S., Slijper, H. P., Andrinopoulou, E. R., Coppieters, M. W., Selles, R. W., The Hand-Wrist Study Group (2022). *Clinical orthopaedics and related research*, *480*(7), 1271–1284.

ABSTRACT

Background

Surgery for thumb carpometacarpal osteoarthritis is offered to patients who do not benefit from nonoperative treatment. Although surgery is generally successful in reducing symptoms, not all patients benefit. Predicting clinical improvement after surgery could provide decision-support and enhance preoperative patient selection.

Questions/purposes

This study aimed to develop and validate prediction models for clinically important improvement in (1) pain and (2) hand function 12 months after surgery for thumb carpometacarpal osteoarthritis.

Methods

Between November 2011 and June 2020, 2653 patients were surgically treated for thumb carpometacarpal osteoarthritis. Patient-reported outcome measures were used to preoperatively assess pain, hand function, and satisfaction with hand function, as well as the general mental health of patients and mindset toward their condition. Patient characteristics, medical history, patient-reported symptom severity, and patient-reported mindset were considered as possible predictors.

Patients who had incomplete Michigan Hand outcomes Questionnaires at baseline or 12 months post-surgery were excluded, as these scores were used to determine clinical improvement. The Michigan Hand outcome Questionnaires provides subscores for pain and hand function. Scores range from 0 to 100, with higher scores indicating less pain and better hand function. An improvement of at least the minimum clinically important difference (MCID) of 14.4 for the pain score and 11.7 for the function score were considered "clinically relevant". These values were derived from previous reports that provided triangulated estimates of two anchor-based and one distribution-based MCID. Data collection resulted in a dataset of 1489 patients for the pain model and 1469 patients for the hand function model. The data were split into training (60%), validation (20%), and test (20%) datasets. The training dataset was used to select the predictive variables and to train our models. The performance of all models was evaluated in the validation dataset, after which one model was selected for further evaluation. Performance of this final model was evaluated on the test data set. We trained the models using logistic regression, random forest, and gradient boosting machines and compared their performance. We chose these algorithms because of their relative simplicity, which makes them easier to implement and interpret. Model performance was assessed using discriminative ability and qualitative visual inspection of calibration curves. Discrimination was measured using area under the curve (AUC) and is a measure of how well the model can differentiate between the outcomes (improvement or

no improvement), with an AUC of 0.5 being equal to chance. Calibration is a measure of the agreement between the predicted probabilities and the observed frequencies and was assessed by visual inspection of calibration curves. We selected the model with the most promising performance for clinical implementation (that is, good model performance and a low number of predictors) for further evaluation in the test dataset.

Results

For pain, the random forest model showed the most promising results based on discrimination, calibration, and number of predictors in the validation dataset. In the test dataset, this pain model had a poor AUC (0.59) and poor calibration. For function, the gradient boosting machine showed the most promising results in the validation dataset. This model had a good area under the curve (0.74) and good calibration in the test dataset. The baseline Michigan Hand outcomes Questionnaire hand function score was the only predictor in the model. For the hand function model, we made a web application that can be accessed via: https://analyse.equipezorgbedrijven.nl/shiny/cmc1-prediction-model-Eng/.

Conclusion

We developed a promising model that may allow clinicians to predict the chance of functional improvement in an individual patient undergoing surgery for thumb carpometacarpal osteoarthritis, which would thereby help in the decision-making process. However, caution is warranted because our model has not been externally validated. Unfortunately, the performance of the prediction model for pain is insufficient for application in clinical practice.

INTRODUCTION

Thumb carpometacarpal (CMC1) osteoarthritis (OA) is common and increases in frequency with age. The symptomatic prevalence is 2% in men and 7% in women older than 50 years [20, 35, 46]. The disorder can lead to impaired hand function because of pain, weakness, loss of motion, and progressive deformity [3, 4]. Initial treatment options are nonsurgical, but surgical treatment might be indicated for a subset of patients who have persistent pain and disability. Although surgical treatment is generally successful in reducing symptoms, not all patients experience benefits from surgery, and some may not be satisfied with their treatment [3, 34, 48].

Pain reduction is generally the most important reason for patients seeking surgical treatment for CMC1 OA, followed by improving hand function [17]. Therefore, it would be useful to be able to accurately predict whether a patient will experience a clinically meaningful reduction in pain and improvement in hand function after surgery. This would help in the decision-making process, help manage patients' expectations, and assist clinicians in determining which patients will benefit from surgery; this may lead to better preoperative patient selection and may improve the likelihood that patients will be pleased with their surgical results [34]. However, determining which patients will benefit from surgical treatment is challenging because many factors may influence outcomes, such as demographics, clinical characteristics, and psychosocial profiles [12, 26, 36, 47]. Thus, developing and implementing tools that accurately predict clinical improvement would be valuable. At present, there are no prediction models available to predict clinically meaningful improvement after the surgical treatment of CMC1 OA.

Machine learning is a type of artificial intelligence that is seeing wider use in healthcare and is increasingly being used to develop prediction models [9]. In a recent editorial, *Clinical Orthopaedics and Related Research*® highlighted the potential value of machine learning in clinical research [29]. Machine learning is based on algorithms that can build models which learn from sample data to make predictions without being explicitly programmed to do so [53]. The models are trained on a training dataset and then evaluated on one or two other datasets (validation and test datasets) [53]. Machine-learning methods can develop models based on large quantities of possible predictive variables and process large amounts of data [6]. Therefore, these algorithms may be better able to identify patterns in large datasets than traditional statistical methods, which may lead to better predictive performance [31].

We aimed to develop and validate two prediction models using machine-learning methods to forecast the probability of clinically meaningful improvement in (1) pain and (2) hand function of patients 12 months after surgery for CMC1 OA. More specifically, in separate models, we focused on predicting pain reduction and improvement in hand function.

We trained and validated our models using different algorithms: one traditional statistical method and two commonly used machine-learning algorithms.

PATIENTS AND METHODS

Study Population

We conducted a retrospective study using longitudinally maintained data from the Hand-Wrist Study Group, which is a collaboration between the Xpert Clinics Hand and Wrist Care and the Departments of Rehabilitation Medicine and Plastic and Reconstructive Surgery of Erasmus Medical Centre in Rotterdam, the Netherlands. The Xpert Clinics comprise 25 locations and 23 European board-certified (Federation of European Societies for Surgery of the Hand) hand surgeons. The cohort and data collection methods have been described in more detail elsewhere [41]. We used data collected between November 2011 and June 2020. All patients were asked to complete patient-reported outcome measures before surgery and 12 months after surgery as part of routine outcome measurements.

We included all patients who received surgery for CMC1 OA (trapeziectomy with ligament reconstruction and tendon interposition (LRTI) [8, 16, 42, 51]), and completed the Michigan Hand outcomes Questionnaire (MHQ) at baseline and 12 months after surgery. Patients with an incomplete MHQ were included if they had a complete MHQ pain score for the development of our pain model and a complete MHQ function score for the development of our hand function model. We excluded patients who underwent revision surgery and patients treated with a different surgical technique than trapeziectomy with ligament reconstruction and tendon interposition because these procedures are not performed routinely in our clinics.

Patients

Diagnosis and Treatment

Diagnoses were made by European board-certified hand surgeons based on clinical symptoms and, when required, additional radiographs of the CMC1 joint. In general, surgery was recommended to patients who did not improve after at least 3 months of nonoperative treatment consisting of hand therapy and braces. In our clinics, this is about 15% of patients [45], and surgery generally consists of trapeziectomy with ligament reconstruction and tendon interposition. The choice of tendonplasty after trapeziectomy depended on the surgeon's preference, which is most likely influenced by the location of residency. The most performed procedure in our clinics is the Weilby sling [51]. Given that the type of tendonplasty was not structurally recorded with sufficient detail in the dataset, we did not include this as possible predictor for our models. We did not expect this to influence our results because there is no evidence that one tendonplasty is superior

regarding improvement in pain, hand function, and patient-reported outcome measures [48, 49].

During the data collection period, 2653 patients were surgically treated for CMC1 OA, excluding revision surgeries. Of those patients, 68% (1794) were treated with the Weilby sling procedure. After excluding patients who did not have complete MHQ score data for pain or function, 1489 patients were included in the prediction model development for pain and 1469 patients were included in the model development for hand function (Fig. 1). We performed two nonresponder analyses. One nonresponder analysis was conducted between all patients who were surgically treated for CMC1 OA and those who were included in our datasets for pain (Supplementary Table 1; supplemental materials are available with the online version of *CORR*®) and for function (Supplementary Table 2; supplemental materials are available with the online version of *CORR*®); the other was for patients with missing MHQ scores at 12-months for pain (Supplementary Table 3; supplemental materials are available with the online version of *CORR*®) and for function (Supplementary Table 4; supplemental materials are available with the online version of *CORR*®). Although we found differences in both analyses in symptom duration, second opinion (yes/no), and smoking, these were small effects with a maximum effect size of 0.12.

In the dataset for pain, mean age was 61 ± 8 years and 79% (1178 of 1489) were women. Of these patients, 47% (704 of 1489) were unemployed. Average preoperative MHQ pain score was 34.3 and average preoperative MHQ hand function score was 48.9. The most common comorbidities were other disorders of the locomotor system (23% [345 of 1489] of patients) and rheumatic disorders (17% [251 of 1489] of patients) (Table 1). Patient characteristics for the hand function model development were similar to those for the pain model (Table 2).

Primary Outcome

To assess symptom relief after surgery, we used the difference in the MHQ scores between baseline and 12 months after surgery [11]. The MHQ is a self-reported questionnaire developed for all conditions of the hand. It provides a summary score and subscores for pain, hand function, ability to complete daily activities, work performance, aesthetics, and satisfaction separately. Scores range from 0 to 100, with higher scores indicating better health [11]. In this study, we focused on the pain and function scores of the MHQ. The Dutch-language version of the MHQ was used [11, 25].

We defined the threshold for a clinically meaningful improvement as having an increase of at least 14.4 and 11.7 points on the MHQ pain score and function score, respectively. These thresholds are based on the minimum clinically important differences (MCID). The MCIDs we used are triangulated estimates of three calculation methods: two anchor-based question methods and one statistical distribution method. They were calculated

for patients with atraumatic hand and wrist conditions [32]. These MCIDs were chosen because determination of MCIDs based on triangulation of multiple calculation methods is recommended [40]. Furthermore, to our knowledge, there are currently no MCIDs available for the MHQ that are specifically determined for CMC1 OA [32]. We dichotomized each patient's change in score between baseline and 12 months as threshold reached or threshold not reached. The prediction models were trained to predict whether a patient would reach the threshold and thus benefit from surgery. The outcome of each model represents the predicted probability of reaching the threshold for the individual patient.

We also included patients with an MHQ score at intake higher than 85.6 for pain and greater than 88.3 for hand function. These patients could not experience an improvement of 14.4 or 11.7, respectively, because of a ceiling effect. Therefore, the chance of these patients reaching the MCID after 12 months was, per definition, zero. However, to provide our model with the opportunity to also learn from these patients, we decided not to exclude them.

Measurements

We considered several baseline measurements as possible predictors for our models (Supplementary Table 5; supplemental materials are available with the online version of *CORR*®). Sociodemographic characteristics such as age, gender, BMI, and occupation as well as medical history, including comorbidities, were collected at intake.

Strength was measured at intake using a Biometrics E-link handgrip dynamometer (Biometrics Ltd). Strength measurements included grip strength, key pinch strength, and tip pinch strength. All measurements were performed according to the guidelines of the American Society of Hand Therapists [14].

Patient-reported outcome measure questionnaires were sent by email after consultation with the hand surgeon. The VAS was used to measure pain at rest and during loading, hand function, and satisfaction with hand function. Each subscale ranged from 0 to 100, with a higher score representing more pain but better hand function and greater satisfaction. The patient's mindset toward their condition and treatment as well as their general mental health and quality of life were measured using several patient-reported outcome measures: the Brief Illness Perception Questionnaire, Credibility and Expectancy Questionnaire, the Patient Health Questionnaire-4, the Pain Catastrophizing Scale, and Euro-QoL-5D-5L. The Dutch-language versions of all questionnaires were used [7, 13, 21, 27, 44]. We continue to improve our data collection and add new variables to the routine measurements. For example, the psychological questionnaires were added in September 2017. Therefore, only patients treated after September 2017 were invited to complete the psychological questionnaires. For the pain model, 248 patients were enrolled after September 2017 and for function 234 patients were enrolled.

Missing Data

There was a substantial proportion of missing data on mindset because only patients who were included between September 2017 and June 2020 were asked to complete these questionnaires. Furthermore, there was also a substantial number of nonresponders to the other measurements because these measures were collected as part of daily clinical practice. Therefore, we performed a nonresponder analysis by comparing baseline characteristics. We imputed data using the k-nearest-neighbor imputation implementation in the Caret package [28], as missingness is most likely missing completely at random or missing at random. Madley-Dowd et al. [33] reported that datasets with up to 90% of missing data can be reasonably imputed using multiple imputation.

Data Splitting and Measurements of Performance

Patient-reported outcome measures, sociodemographic characteristics, and strength of the affected hand before surgery were considered as possible predictors in our models. To avoid overfitting and to base decisions for the most promising model(s) on, we split the resulting data into training (60%), validation (20%), and test datasets (20%) (Fig. 2). Both the validation dataset and test dataset refer to a sample of the dataset that is separate from the training dataset [53]. To select the algorithm for our final model, we applied all algorithms to the validation dataset and selected the one with the most promising performance in terms of (1) discrimination, (2) calibration, and (3) the number of predictors as our final model. We also took the number of predictors into account because we believe a low number of predictors will make the model easier to implement and use in daily practice. The test dataset was then used to evaluate the performance of this final model based on discrimination and calibration.

The random split of the dataset for the pain model (1489) resulted in a training dataset of 894 patients, a validation dataset of 298 patients, and a test dataset of 297 patients. In the training dataset, 73% (653 of 894) of patients reached the MCID threshold of 14.4 points on the MHQ pain scale. In both the validation (218 of 298) and test (217 of 297) datasets, 73% of patients reached the MCID threshold.

The dataset for the hand function model (1469) was randomly split into a training dataset of 883 patients, a validation dataset of 293, and a test dataset of 293. In the training dataset, 56% (494 of 883) of patients reached the MCID threshold of 11.7 points on the MHQ function scale. In both the validation dataset and test dataset, 56% (164 of 293) of patients reached the MCID threshold.

For the dataset for the pain model development (Supplementary Table 6; supplemental materials are available with the online version of *CORR*®) and the dataset for the hand function development (Supplementary Table 7; supplemental materials are available with

the online version of *CORR*®), there were no important differences in patient characteristics and preoperative PROMs values between the train, validation, and test datasets.

Ethical Approval

The medical ethics review board at Erasmus Medical Centre approved the study. This study was reported according to the guidelines of the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement [37]. All patients provided written informed consent for their data to be used for research purposes.

Statistical Analysis and Machine Learning

After splitting the data, we standardized the data and imputed missing data. Standardization consisted of centering and scaling the data [53]. When standardization and imputation are performed before splitting the data, the validation and test dataset are not completely independent, which can result in a model performance that is too optimistic. Therefore, we performed standardization and imputation after splitting.

We compared three algorithms: logistic regression (generalized linear models), random forest, and gradient boosting machines. Logistic regression is a traditional regression-based statistical model. Random forest and gradient boosting machines are decision tree-based machine-learning models. We decided to use gradient boosting machine and random forest algorithms as our machine learning algorithms for several reasons. First, they are relatively simple to implement, and because of their similarities with decision trees, they are easier to interpret than other, more complex algorithms [53]. Second, they are computationally less expensive and require less extensive datasets [6, 31, 53]. We selected variables for our models using recursive feature elimination with five repeats of 10-fold cross-validation in each training dataset (Fig. 2). Recursive feature elimination can be considered as backward selection of predictive variables. It starts by building a model that includes all variables as predictors. For each predictor, an importance score is computed, and predictors with the lowest score are removed. Then the model is rebuilt, and the process is repeated until model performance decreases by removing another variable [19].

Because the surgical treatment of CMC1 OA is generally successful [48], we expected more patients in the threshold-reached group than in the threshold-not-reached group. This was confirmed by a preliminary analysis for the MCID for pain, with 73% in the threshold-reached group, and 27% of patients in the threshold-not-reached group. To account for this imbalance, we incorporated resampling in the feature elimination process. Thus, for our pain model, we performed recursive feature elimination three times for each machine-learning algorithm: without sampling, with up-sampling, and with down-sampling. With up-sampling, randomly selected patients are duplicated in the minority group, and with down-sampling, randomly selected patients are removed from the majority group. Because

the resampling methods have disadvantages [10], we tested both. A preliminary analysis of the MCID for hand function showed the data were sufficiently balanced, with 56% in the threshold-reached group and 44% of patients in the threshold-not-reached group. Therefore, resampling was not needed in the function dataset. For each machine-learning algorithm for pain, the best-performing resampling method was selected based on area under the curve (AUC) values and the number of predictors. The AUC is a measure of the discriminative ability of a model; that is, the ability of the model to classify the two different groups correctly [23]. The models with the most promising performance were then used for further analysis.

The selected models, one from each machine-learning algorithm and with the predictive variables selected using recursive feature elimination, were trained in the original training set. After training the models, we analyzed performance in the validation set using AUC values and calibration. Calibration is a measure of the model's fit and refers to the agreement between predicted probabilities and the observed frequency of the outcome [15]. In other words, this indicates whether, for example, out of 10 patients with a predicted probability to improve of 0.6, we observe that six patients actually improved. Calibration was visually assessed using calibration curves [15, 24, 38]. The model performs well on calibration when the calibration curve is close to the bisector. If the calibration curve lies above the bisector, it means the model underestimates the probability of the patient reaching the MCID; if the calibration curve lies under the bisector, the model overestimates the probability. The confidence belts represent the estimated degree of uncertainty of the calibration curve [15]. We then selected the algorithm with the best AUC and calibration. Additionally, we considered the number of predictors. This model was further evaluated in our test dataset using discriminative ability (AUC) and calibration (visual inspection of calibration curves). An AUC between 0.7 and 0.8 was considered acceptable discrimination, an AUC between 0.8 and 0.9 excellent, and an AUC above 0.9 outstanding [23]. Furthermore, we determined the sensitivity and specificity, positive predictive value, and negative predictive value.

The analysis was performed using R statistical programming, version 1.3.1073 (R Foundation). Prediction models were trained using the Caret package, version 6.0-86 [28]. A p value < 0.05 was considered statistically significant.

RESULTS

Pain Model

In the validation dataset, the random forest model with down-sampling showed the most promising performance in terms of discrimination, visual inspection of calibration curves (Supplementary Fig. 1A-C; supplemental materials are available with the online version of *CORR*®), and number of predictors (Supplementary Table 8; supplemental materials

are available with the online version of *CORR®*). This model was evaluated further (Supplementary File 1; supplemental materials are available with the online version of *CORR®*). Unfortunately, in the test dataset, it performed poorly with an AUC of 0.59 (95% confidence interval 0.52 to 0.66), which is hardly better than chance. Sensitivity was 0.67 and specificity was 0.49 at a threshold of 0.72 (Table 3). In addition, a visual inspection of the calibration curve also indicated poor calibration (Fig. 3). We therefore believe this model should not be used in clinical practice.

Function Model

In the validation dataset, the gradient boosting machines model was selected for further evaluation because it showed the most promising performance in terms of discrimination, calibration (Supplementary Fig. 2A-C), and the fact that it required only a single predictor variable (Supplementary File 1); the MHQ function score at baseline (Supplementary Table 8). In the test dataset, it had a good discriminative ability, with an AUC of 0.74 (95% CI 0.69 to 0.80) (Table 3). Sensitivity was 0.62 and specificity was 0.72 at a threshold of 0.62 (Table 3). A visual inspection of the calibration curve showed good calibration (Fig. 4). We have made this model available as web application. The model predicts the change of reaching the MCID for hand function for an individual patient 12 months after surgery, given the patient pre-operative MHQ hand function score (Supplementary Fig. 3; supplemental materials are available with the online version of *CORR®*).

The final hand function model is presented as a Shiny internet application, accessible at: https://analyse.equipezorgbedrijven.nl/shiny/cmc1-prediction-model-Eng/. The app currently does not have the Conformité Européenne (CE) mark and has not yet been externally validated; therefore, caution is warranted when using the application.

The R code of the models is available via GitHub [39]. Because of the poor predictive ability, we did not make an internet application for the pain model.

DISCUSSION

Assessing the likelihood of success is an important part of the decision to undergo a certain treatment, especially when the treatment is invasive and elective in nature such as the surgical treatment of CMC1 OA. Thus, communicating the chance of a successful outcome can help the decision-making process. However, predicting which patients will improve in symptoms is difficult. Therefore, a model that predicts the probability of improvement for each patient might contribute to the shared decision-making process. It could also help manage patients' expectations of the treatment outcome. This study aimed to develop prediction models for the probability of clinically meaningful improvement in pain and hand function 12 months after trapeziectomy with ligament reconstruction

and tenson interposition for CMC1 OA. Unfortunately, despite the relatively large dataset with many variables, we considered the performance of the pain model as insufficient for clinical practice. However, the hand function model had a good discriminative ability and good calibration in our test dataset. This model was a gradient boosting machines model with the baseline MHQ hand function score as the only predictor. We have made an internet application of our hand function model, which can be accessed via: https://analyse.equipezorgbedrijven.nl/shiny/cmc1-prediction-model-Eng/. To calculate the prediction of an individual patient, the preoperative MHQ function score of the patient is submitted. The model then calculates the probability of improvement 12 1months after surgery. If the MHQ function score of the patient is unknown, it can be calculated in the application by answering five questions.

Limitations

This study has some limitations. Although the models were internally validated in a separate test dataset, no external validation was performed. Evaluating predictive performance of these models in a prospective setting with new patients could be a valuable addition. The current models are generalizable to settings where patients with thumb base osteoarthritis are first treated nonsurgically, and trapeziectomy with LRTI is considered when symptoms are not sufficiently relieved. No distinction was made between different tendonplasties that can be considered as LRTI because they are very similar and there is little evidence for differences in patient-reported outcomes between these techniques [41, 49]. Before generalizing predictions from our model to other surgical treatment options than trapeziectomy with LRTI, such as arthrodesis and prosthetics, additional validation is needed.

Another important limitation is the relatively high proportion of missing data, which is inherent to the to the nature of the dataset, where all patients are invited to complete multiple patient-reported outcome measures as part as routine outcome measures. Specifically, there was a high proportion of missing data for the psychological variables, which have only been collected since September 2017. We have chosen to still include these because previous studies have shown that these are associated with treatment outcomes of thumb base osteoarthritis [18, 52]. Additionally, we compared patient characteristics and preoperative symptom severity to assess whether nonresponders differed from responders and found only small differences with negligible effect sizes. Missing data was imputed using k-nearest neighbors, but multiple imputation would be preferable. However, to our knowledge, this is currently not implemented in R. When this implementation is available, this may possibly result in better prediction models in the future.

Our models predict the probability that a patient will reach a clinically meaningful improvement, defined as reaching the MCID for the MHQ pain score and the MHQ function

score [32]. These MCID scores were calculated for patients with atraumatic hand and wrist conditions, not specifically for patients with CMC1 OA. Additionally, the MCID threshold is determined for patient populations and may be less relevant for individual patients. Therefore, we believe it is important to clearly communicate our definition of improvement to clinicians and patients when using this model, emphasizing that this improvement is considered relevant for most patients but not all. Furthermore, we used random forest and gradient boosting as machine learning algorithms in our study. It is, however, possible that more complex algorithms, such as artificial neural networks, have a better predictive performance. In our case, the relatively small dataset compared with other studies on machine learning limited our choice of algorithms. Further, the use of additional variables such as preoperative goniometric measurements, genetics, or comorbidities might have improved the performance of our prediction models [12, 22, 50]. However, we did not have sufficient data to evaluate these variables. In our opinion, the inclusion of variables such as genetics would make it harder to implement our model in daily clinical practice.

Finally, we judged that the AUC of the pain model (0.59) was insufficient since it was only slightly better than chance, and the AUC of the hand function model (0.74) was sufficient for application in clinical practice, given that it met the threshold for acceptable discrimination [23]. However, what is considered "sufficient" might be debatable and dependent on the action that will follow from the prediction on the model. We therefore strongly recommend that the model is only used as a decision aid that provides additional insight into the expected outcome of surgery.

Pain Model

The performance of our best performing pain model was insufficient. The model performed poorly on both discrimination and calibration and should therefore, not be used in clinical practice. This is in line with reports on surgery for OA of other joints [30, 43], and the finding that pain after total joint arthroplasty cannot accurately be predicted using clinical variables [5]. The nature of postoperative pain might be different from preoperative OA-related pain, and this is therefore more difficult to predict. In clinical practice, we have noticed that many patients indicate they still experience pain but not the familiar OA-related pain that was the surgical indication. Because there are indications that genetic factors play a role in chronic and neuropathic pain [22, 50], this may be a direction for further research into predicting postoperative pain.

Function Model

Our hand function model showed reasonable performance in terms of discrimination and calibration and required only one predictor. This model was used for the development of a web application that is publicly available online and can be used to help guide the decision-

making process. However, since our model has not been externally validated, caution is warranted. The model was a gradient boosting machines algorithm.

In our study, machine-learning algorithms had a better predictive performance in both our datasets than the traditional statistical logistic regression model. Although the discriminative ability of the logistic regression model for hand function was only marginally worse than that of the gradient boosting machine model in the validation dataset, it required almost 80 variables as input, whereas the gradient boosting machines model only required one (Supplementary Table 8). Machine-learning algorithms might be better equipped to deal with the nonlinearities in datasets [1, 31] that are often present in real-world data. It might, however, be possible to fit these nonlinear effects using statistical methods such as generalized additive models or nonlinear effects in logistic regression.

Although some studies have reported prognostic factors influencing the outcome of surgical treatment of CMC1 OA [2, 12, 26, 36], the development of a prediction model is new. One study examined the prognostic value of preoperative patient-reported disability and psychological characteristics for early postoperative outcomes with a mean follow-up of 14 weeks [26]. The authors found that patients with greater preoperative disability experienced more improvement after surgery but did not find an association between psychological factors and outcomes. This is in line with the results of our hand function model, which only requires baseline function as a predictor. Another study evaluated the relationship between the duration of symptoms and surgical outcomes [2]. The authors found that patients with an increased duration of symptoms had a poorer postoperative outcome. Although the duration of symptoms was one of the variables in our dataset, this was not one of the predicting variables in our models. This indicates that in our dataset, the duration of symptoms did not have sufficient predictive power.

Conclusion

In conclusion, we developed a model to predict the probability of improvement in hand function 12 months after trapeziectomy and ligament reconstruction with tendon interposition for CMC1 OA. The model had good discriminative ability and good calibration in our test dataset. Unfortunately, the performance of our pain model was insufficient for use in practice. The final model for hand function was used to develop an online application that can be used to estimate the chance of survival for an individual patient. However, our model does not have CE marking and has not been externally validated. By making our model available online, we encourage others to validate the model in their patient populations.

ACKNOWLEDGMENTS

We would like to thank all patients who have filled out questionnaires as part of their clinical care and who agreed that their data could be anonymously used for the present study. In addition, we would like to acknowledge the caregivers and personnel of Xpert Clinic, Handtherapie Nederland, and Equipe Zorgbedrijven for assisting in the routine outcome measurements that are the basis for this manuscript.

GROUP AUTHORSHIP

[a]Members of the Hand-Wrist Study include:

R Arjen M Blomme MD, Berbel JR Sluijter MD PhD, Dirk-Jan JC van der Avoort MD, Alexander Kroeze MD, Jeroen Smit MD PhD, Jan Debeij MD PhD, Erik T Walbeehm MD PhD, Gijs M van Couwelaar MD, Guus M Vermeulen MD PhD, Hans de Schipper MD, Hans Temming MD, Jeroen H van Uchelen MD PhD, H Luitzen de Boer MD, Nicoline de Haas MD, Kennard Harmsen MD, Oliver T Zöphel MD PhD, Reinier Feitz MD, Richard Koch MD, Steven ER Hovius MD PhD, Thybout M Moojen MD PhD, Xander Smit MD PhD, Rob van Huis PT, Pierre Y Pennehouat PT, Karin Schoneveld PT MSc, Yara E van Kooij PT MSc, Robbert M Wouters PT PhD, Paul Zagt PT, Folkert J van Ewijk PT, Joris J Veltkamp PT, Alexandra Fink PT, Willemijn A de Ridder PT MSc, Miguel C Jansen MD PhD, Mark JW van der Oest PhD, Pepijn O Sun MD, Joris S Teunissen BSc, Jak Dekker MSc, Marlies L Jansen-Landheer MD MSc, Marloes HP ter Stege MSc

REFERENCES

1. Auret L, Aldrich C. Interpretation of nonlinear relationships between process variables by use of random forests. *Minerals Engineering.* 2012;35:27-42.
2. Baca ME, Rozental TD, McFarlane K, Hall MJ, Ostergaard PJ, Harper CM. Trapeziometacarpal joint arthritis: Is duration of symptoms a predictor of surgical outcomes? *J Hand Surg Am.* 2020;45:1184.e1181-1184.e1187.
3. Baker RH, Al-Shukri J, Davis TR. Evidence-based medicine: thumb basal joint arthritis. *Plast Reconstr Surg.* 2017;139:256e-266e.
4. Bakri K, Moran SL. Thumb carpometacarpal arthritis. *Plast Reconstr Surg.* 2015;135:508-520.
5. Barroso J, Wakaizumi K, Reckziegel D, et al. Prognostics for pain in osteoarthritis: Do clinical measures predict pain after total joint replacement? *PLoS One.* 2020;15:e0222370.
6. Bayliss L, Jones LD. The role of artificial intelligence and machine learning in predicting orthopaedic outcomes. *Bone Joint J.* 2019;101:1476-1478.
7. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. *J Psychosom Res.* 2006;60:631-637.
8. Burton RI, Pellegrini VD, Jr. Surgical management of basal joint arthritis of the thumb. Part II. Ligament reconstruction with tendon interposition arthroplasty. *J Hand Surg Am.* 1986;11:324-332.
9. Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. *Front Bioeng Biotechnol.* 2018;6:75.
10. Chawla NV. Data Mining for Imbalanced Datasets: An Overview. In: Maimon O, Rokach L, ed. *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer US; 2010:875-886.
11. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am.* 1998;23:575-587.
12. Degreef I, De Smet L. Predictors of outcome in surgical treatment for basal joint osteoarthritis of the thumb. *Clin Rheumatol.* 2006;25:140-142.
13. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry.* 2000;31:73-86.
14. Fess E, Moran C. *American Society of Hand Therapists clinical assessment recommendations*: American Society of Hand Therapists; 1981.
15. Finazzi S, Poole D, Luciani D, Cogo PE, Bertolini G. Calibration belt for quality-of-care assessment based on dichotomous outcomes. *PLoS One.* 2011;6:e16110.
16. Froimson AI. Tendon arthroplasty of the trapeziometacarpal joint. *Clin Orthop Relat Res.* 1970;70:191-199.
17. Frouzakis R, Herren DB, Marks M. Evaluation of expectations and expectation fulfillment in patients treated for trapeziometacarpal osteoarthritis. *J Hand Surg Am.* 2015;40:483-490.
18. Giesinger JM, Kuster MS, Behrend H, Giesinger K. Association of psychological status and patient-reported physical outcome measures in joint arthroplasty: a lack of divergent validity. *Health Qual Life Outcomes.* 2013;11:64.
19. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Machine Learning.* 2002;46:389-422.

20. Haugen IK, Englund M, Aliabadi P, et al. Prevalence, incidence and progression of hand osteoarthritis in the general population: the Framingham Osteoarthritis Study. *Ann Rheum Dis.* 2011;70:1581-1586.

21. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.* 2011;20:1727-1736.

22. Hoofwijk DM, van Reij RR, Rutten BP, Kenis G, Buhre WF, Joosten EA. Genetic polymorphisms and their association with the prevalence and severity of chronic postsurgical pain: a systematic review. *Br J Anaesth.* 2016;117:708-719.

23. Hosmer DW LS, Sturdivant RX. Assessing the fit of the model. *Applied Logistic Regression*. New York: NY John Wiley and Sons; 2000:143-202.

24. Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. *J Am Med Inform Assoc.* 2020;27:621-633.

25. Huijsmans RS, H.; Aufdemkampe, G. Michigan Hand Outcomes Questionnaire-Dutch Language Version; een vragenlijst voor patienten met handfunctieproblemen. *Fysiopraxis.* 2001;9: 38-41.

26. Kazmers NH, Grasu B, Presson AP, Ou Z, Henrie NB, Tyser AR. The prognostic value of preoperative patient-reported function and psychological characteristics on early outcomes following trapeziectomy with ligament reconstruction tendon interposition for treatment of thumb carpometacarpal osteoarthritis. *J Hand Surg Am*. 2020;45:469-478.

27. Kroenke K, Spitzer RL, Williams JB, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics.* 2009;50:613-621.

28. Kuhn M. Building Predictive Models in R Using the caret Package. *2008.* 2008;28:26.

29. Leopold SS, Porcher R, Gebhardt MC, et al. Editorial: Opposites attract at CORR®-machine learning and qualitative research. *Clin Orthop Relat Res.* 2020;478:2193-2196.

30. Lewis GN, Rice DA, McNair PJ, Kluger M. Predictors of persistent pain after total knee arthroplasty: a systematic review and meta-analysis. *Br J Anaesth.* 2015;114:551-561.

31. Liu NT, Salinas J. Machine learning for predicting outcomes in trauma. *Shock*. 2017;48:504-510.

32. London DA, Stepan JG, Calfee RP. Determining the Michigan Hand Outcomes Questionnaire minimal clinically important difference by means of three methods. *Plast Reconstr Surg.* 2014;133:616-625.

33. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol.* 2019;110:63-73.

34. Marks M, Audigé L, Reissner L, Herren DB, Schindele S, Vliet Vlieland TP. Determinants of patient satisfaction after surgery or corticosteroid injection for trapeziometacarpal osteoarthritis: results of a prospective cohort study. *Arch Orthop Trauma Surg.* 2015;135:141-147.

35. Marshall M, van der Windt D, Nicholls E, Myers H, Dziedzic K. Radiographic thumb osteoarthritis: frequency, patterns and associations with pain and clinical assessment findings in a community-dwelling population. *Rheumatology (Oxford).* 2011;50:735-739.

36. Moineau G, Richou J, Liot M, Le Nen D. Prognostic factors for the recovery of hand function following trapeziectomy with ligamentoplasty stabilisation. *Orthop Traumatol Surg Res.* 2009;95:352-358.

37. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162:W1-73.

38. Nattino G, Finazzi S, Bertolini G. A new calibration test and a reappraisal of the calibration belt for the assessment of prediction models based on dichotomous outcomes. *Stat Med.* 2014;33:2390-2407.

39. Ninalouisa. Ninalouisa/CMC1-Prediction: Release CMC1 Prediction (Version V2.0). Zenodo. Available at: https://doi.org/10.5281/zenodo.5032347 (Accessed: June 25 2021)

40. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61:102-109.

41. Selles RW, Wouters RM, Poelstra R, et al. Routine health outcome measurement: development, design, and implementation of the Hand and Wrist Cohort. *Plast Reconstr Surg.* 2020;146:343-354.

42. Sigfusson R, Lundborg G. Abductor pollicis longus tendon arthroplasty for treatment of arthrosis in the first carpometacarpal joint. *Scand J Plast Reconstr Surg Hand Surg.* 1991;25:73-77.

43. Spekreijse K, Steyerberg E, Tsehaie J, et al. Predicting outcome after surgery for carpometacarpal osteoarthritis: a prospective study. *HAND.* 2016;11:1S-2S.

44. Sullivan M, Bishop S, Pivik J. The Pain Catastrophizing Scale: Development and validation. *Psychological Assessment.* 1995;7:524-532.

45. Tsehaie J, Spekreijse KR, Wouters RM, et al. Outcome of a hand orthosis and hand therapy for carpometacarpal osteoarthritis in daily practice: a prospective cohort study. *J Hand Surg Am.* 2018;43:1000-1009.e1001.

46. van der Oest MJW, Duraku LS, Andrinopoulou ER, et al. The prevalence of radiographic thumb base osteoarthritis: a meta-analysis. *Osteoarthritis Cartilage.* 2021.

47. van der Oest MJW, Teunissen JS, Poelstra R, Feitz R, Burdorf A, Selles RW. Factors affecting return to work after surgical treatment of trapeziometacarpal joint osteoarthritis. *J Hand Surg Eur Vol.* 202146:979-984.

48. Vermeulen GM, Slijper H, Feitz R, Hovius SE, Moojen TM, Selles RW. Surgical management of primary thumb carpometacarpal osteoarthritis: a systematic review. *J Hand Surg Am.* 2011;36:157-169.

49. Wajon A, Vinycomb T, Carr E, Edmunds I, Ada L. Surgery for thumb (trapeziometacarpal joint) osteoarthritis. *Cochrane Database Syst Rev.* 2015;2015:Cd004631.

50. Warner SC, van Meurs JBJ, Schiphof D, et al. Genome-wide association scan of neuropathic pain symptoms post total joint replacement highlights a variant in the protein-kinase C gene. *Eur Journal Hum Gene.* 2017;25:446-451.

51. Weilby A. Tendon interposition arthroplasty of the first carpo-metacarpal joint. *J Hand Surg Br.* 1988;13:421-425.

52. Wouters RM, Porsius JT, van der Oest MJW, et al. Psychological characteristics, female sex, and opioid use predict acute postoperative pain in patients surgically treated for thumb base osteoarthritis: a cohort study. *Plast Reconstr Surg.* 2020;146:1307-1316.

53. Zhang X-D. Machine Learning. In: Zhang X-D, ed. *A Matrix Algebra Approach to Artificial Intelligence*. Singapore: Springer Singapore; 2020:223-440.

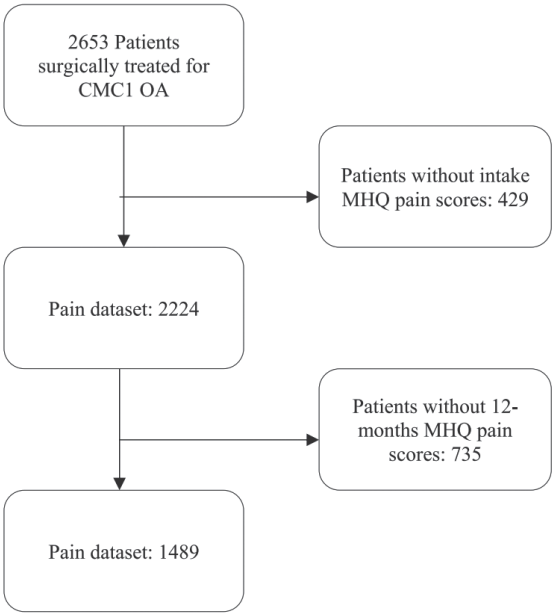


Figure 1A

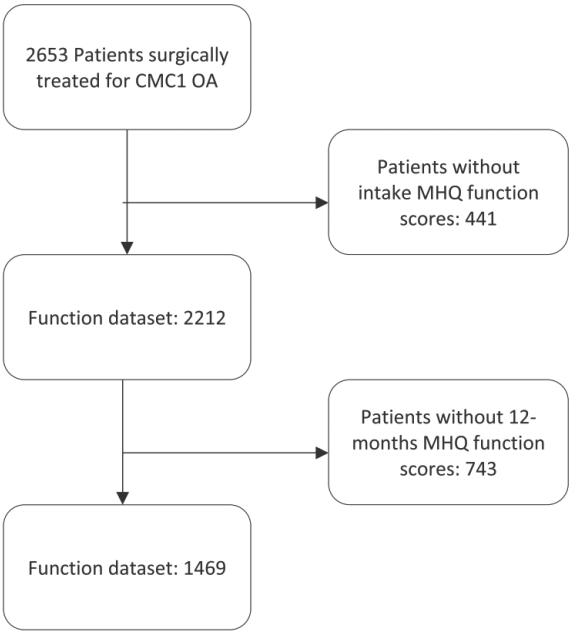


Figure 1B

Figure 1A-B. Flow diagram of patient selection for the pain dataset (A) and function dataset (B). During the inclusion period, 2653 patient were surgically treated with primary trapeziectomy with LRTI. Of these patients, 429 and 441 patients were excluded because they did not have baseline scores for MHQ pain and MHQ function, respectively; and 735 and 743 patients were excluded because of missing MHQ scores at 12 months.

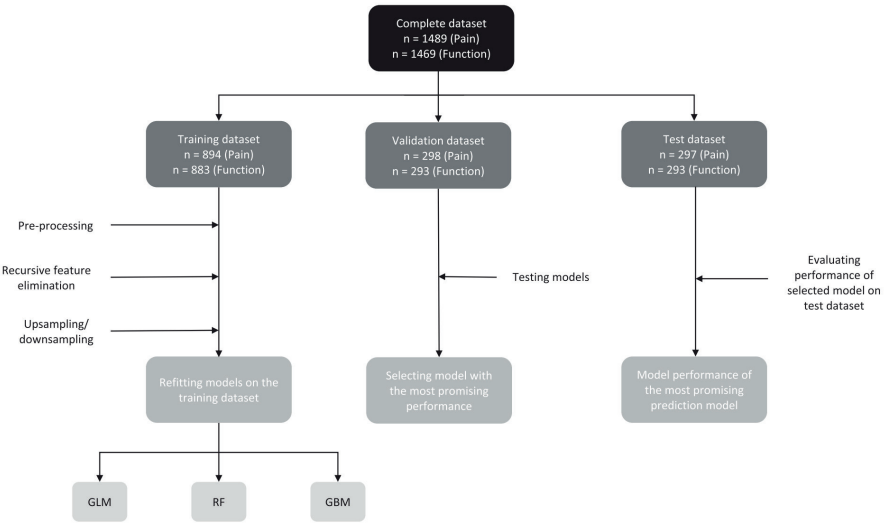


Figure 2. This flow diagram shows the selection of prediction models. The complete dataset was split into training (60%), validation (20%), and test (20%) datasets. The training set was used for feature elimination, resampling, and training of the prediction models. The best performing models of each algorithm were evaluated in the validation dataset. The performance of the model with the best AUC and calibration in the validation dataset was further evaluated in the test dataset; GLM = generalized linear model, RF = random forest, GBM = gradient boosting machine; AUC = area under the curve.

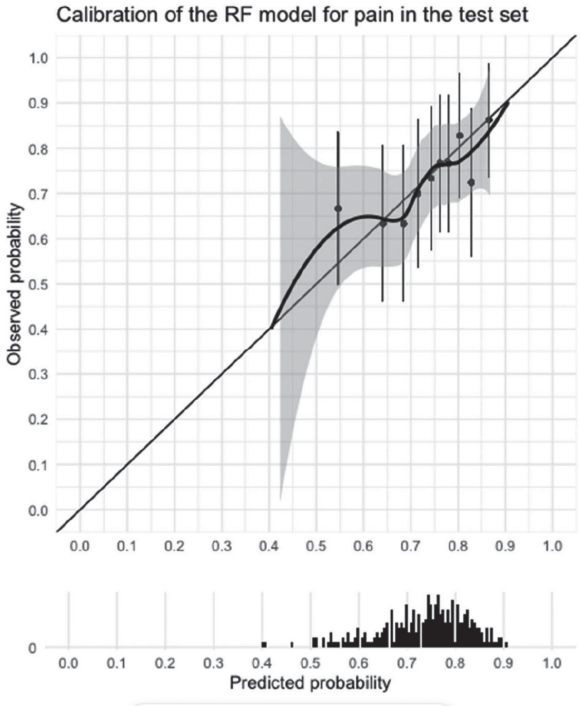


Figure 3. This graph shows the calibration curve of the selected prediction model (random forest) for pain in the test dataset and a histogram of the distribution of the predicted probabilities of improvement. Calibration refers to agreement between the predicted probabilities and observed probabilities. In other words, if 10 people had a probability of improvement of 0.6, did six people actually improve? The model performs well on calibration when the calibration curve lies close to the bisector. Calibration for our pain model was insufficient because of the wide CI and because the curve does not cover the lower probability range.

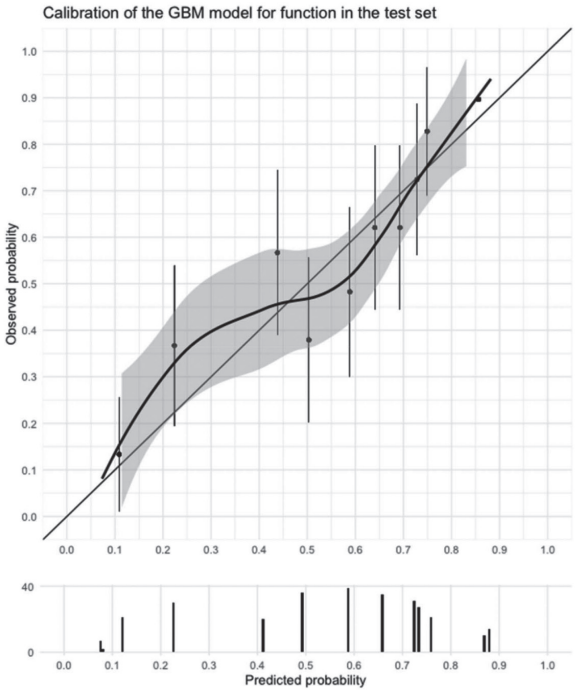


Figure 4. This graph shows the calibration curve of the selected prediction model (gradient boosting machines) for function in the test dataset and a histogram of the distribution of the predicted probabilities of improvement. Calibration refers to agreement between the predicted probabilities and observed probabilities. In other words, if 10 people had a probability of improvement of 0.6, did six people actually improve? The model performs well on calibration when the calibration curve lies close to the bisector. Our model for function shows good calibration.

Table 1. Characteristics of the patients in the training, validation, and test dataset for pain

Parameter	Complete dataset (n = 1489)	% missing	Training (n = 894)	Validation (n = 298)	Test (n = 297)
Age in years	61 ± 8	0	60.5 ± 8.2	60.9 ± 7.5	60.3 ± 7.9
Gender, women	79 (1178)		80 (718)	78 (231)	77 (229)
Duration of symptoms in months	37.2 ± 35.5	2	37.5 ± 35.0	35.9 ± 36.5	37.6 ± 36.0
Second opinion	89 (1325)	0	90 (805)	87 (259)	88 (261)
Hand dominance		0			
Right	85 (1269)		85 (762)	85 (252)	86 (255)
Left	10 (143)		9 (83)	11 (33)	9 (27)
Both	5 (77)		5 (49)	4 (13)	5 (15)
Dominant hand treated	47 (695)	0	45 (405)	50 (149)	48 (141)
Occupational intensity		0			
Not employed	47 (704)		46 (414)	52 (155)	45 (135)
Light	19 (278)		21 (184)	17 (52)	14 (42)
Moderate	22 (333)		21 (190)	20 (59)	28 (84)
Heavy	12 (174)		12 (106)	11 (32)	12 (36)
BMI in kg/m^2	26.6 ± 3.9	35	26.7 ± 3.9	26.4 ± 3	26.7 ± 4.2
Smoking		45			
Never	24 (358)		24 (214)	23 (69)	25 (75)
Disease severity					
Preoperative MHQ Pain score	34.3 ± 14.1	0	34.12 ± 13.98	34.95 ± 14.75	34.04 ± 13.61
Preoperative MHQ Function Score	48.9 ± 17.0	0.6	48.41 ± 16.25	49.95 ± 18.28	49.51 ± 17.94
Medical history		35			
Diabetes	4 (53)		3 (30)	2 (7)	5 (16)
Cardiovascular system	7 (104)		7 (60)	8 (24)	7 (20)

Parameter	Complete dataset (n = 1489)	% missing	Training (n = 894)	Validation (n = 298)	Test (n = 297)
Thrombosis/vasculitis	1 (13)		1 57 (7)	1 (3)	1 (3)
Respiratory system	8 (119)		9 (82)	5 (15)	7 (22)
Liver/kidneys	1 (12)		1 (5)	2 (5)	1 (2)
Cranial nerves	2 (24)		2 (18)	1 (2)	1 (4)
Locomotor system	23 (345)		24 (215)	21 (63)	23 (67)
Rheumatic disorders	17 (251)		18 (157)	14 (41)	18 (53)
Hemorrhoids/varicosities	11 (166)		10 (87)	11 (33)	16 (46)
Allergies	17 (252)		17 (156)	14 (42)	18 (54)
Hematomas	3 (49)		3 (29)	3 (9)	4 (11)

Data presented as mean ± SD or % (n); the training dataset was used to select the predictive variables and to train our models; the performance of all models was evaluated in the validation dataset, after which one model was selected for further evaluation; performance of this final model was evaluated on the test data set.

Table 2. Characteristics of the patients in the training, validation and test dataset for function

Parameter	Complete dataset (n = 1469)	% missing	Training dataset (n = 883)	Validation dataset (n = 293)	Test dataset (n = 293)
Age in years	61 ± 8	0	61 ± 7	60 ± 8	61 ± 8
Gender, women	79 (1167)	0	78 (689)	81 (236)	83 (242)
Duration of symptoms in months	37 ± 46	2	38 ± 36	38 ± 36	34 ± 32
Second opinion	89 (1305)	0	89 (789)	88 (258)	88 (258)
Hand dominance		0			
Right	85 (1251)		85 (748)	89 (260)	83 (243)
Left	10 (141)		10 (90)	7 (21)	10 (30)
Both	5 (77)		5 (45)	4 (12)	7 (20)
Dominant hand treated	47 (686)	0	49 (428)	46 (136)	42 (122)
Occupational intensity		0			
Not employed	47 (691)		49 (428)	44 (130)	45 (133)
Light	19 (277)		18 (156)	20 (58)	22 (63)
Moderate	22 (329)		23 (199)	20 (58)	25 (72)
Heavy	12 (172)		11 (100)	16 (47)	9 (25)
BMI in kg/m^2	26.6 ± 3.9	36	26.6 ± 4.0	26.4 ± 3.6	26.9 ± 3.9
Smoking		45			
Never	24 (351)		25 (218)	20 (58)	26 (75)
Disease severity					
Preoperative MHQ Pain score	34.2 ± 14.0	0.3	33.8 ± 14.1	34.5 ± 14.0	35.3 ± 13.8
Preoperative MHQ Function Score	48.9 ± 17.0	0	49.1 ± 17.2	47.9 ± 16.1	49.2 ± 17.1
Medical history		36			
Diabetes	4 (51)		3 (30)	3 (9)	4 (12)
Cardiovascular system	7 (104)		7 (62)	7 (20)	8 (22)

Parameter	Complete dataset (n = 1469)	% missing	Training dataset (n = 883)	Validation dataset (n = 293)	Test dataset (n = 293)
Thrombosis/vasculitis	1 (13)		1 (10)	0.3 (1)	1 (2)
Respiratory system	8 (118)		8 (70)	7 (21)	9 (27)
Liver/kidneys	1 (12)		1 (8)	1 (2)	1 (2)
Cranial nerves	2 (22)		2 (15)	2 (5)	1 (2)
Locomotor system	23 (337)		24 (208)	19 (55)	25 (74)
Rheumatic disorders	17 (243)		18 (158)	13 (37)	16 (48)
Hemorrhoids/varicosities	11 (162)		11 (101)	11 (33)	10 (28)
Allergies	17 (249)		17 (149)	16 (47)	18 (53)
Hematomas	3 (49)		3 (30)	4 (13)	2 (6)

Data presented as mean ± SD or % (n); the training dataset was used to select the predictive variables and to train our models, the performance of all models was evaluated in the validation dataset, after which one model was selected for further evaluation, performance of this final model was evaluated on the test data set.

Table 3. Model properties in the test dataset of the selected prediction models for pain and function

Outcome	Training method		Results						
	Algorithm	Sampling method	Number of variables	AUC in the test dataset	Sensitivity in the test dataset	Specificity in the test dataset	Positive predictive value	Negative predictive value	Threshold for improvement or no improvement
Pain	Random forest	Downsampling	27	0.59	0.67	0.49	0.78	0.35	0.72
Function	Gradient boosting machine	No sampling	1	0.74	0.62	0.72	0.73	0.60	0.62

17

ALGORITHM VERSUS EXPERT: MACHINE LEARNING VERSUS SURGEON-PREDICTED SYMPTOM IMPROVEMENT AFTER CARPAL TUNNEL RELEASE

Loos, N. L., Hoogendam, L., Souer, J. S., van Uchelen, J. H., Slijper, H. P., Wouters, R. M., Selles, R. W., The Hand-Wrist Study Group (2024). *Neurosurgery*, *95*(1), 110–117.

ABSTRACT

Background and Objectives

Surgeons rely on clinical experience when making predictions about treatment effects. Incorporating algorithm-based predictions of symptom improvement after carpal tunnel release (CTR) could support medical decision-making. However, these algorithm-based predictions need to outperform predictions made by surgeons to add value. We compared predictions of a validated prediction model for symptom improvement after CTR with predictions made by surgeons.

Methods

This cohort study included 97 patients scheduled for CTR. Preoperatively, surgeons estimated each patient's probability of improvement six months post-surgery, defined as reaching the Minimally Clinically Important Difference on the Boston Carpal Tunnel Syndrome Symptom Severity Score. We assessed model and surgeon performance using calibration (calibration belts), discrimination (area under the curve (AUC)), sensitivity, and specificity. Additionally, we assessed the net benefit of decision-making based on the prediction model's estimates versus the surgeon's judgement.

Results

The surgeon predictions had poor calibration and suboptimal discrimination (AUC 0.62, 95%-CI 0.49-0.74), while the prediction model showed good calibration and appropriate discrimination (AUC 0.77, 95%-CI 0.66-0.89, p=0.05). The accuracy of surgeon predictions was 0.65 (95%-CI 0.37-0.78) versus 0.78 (95%-CI 0.67-0.89) for the prediction model (p=0.03). The sensitivity of surgeon predictions and the prediction model were 0.72 (95%-CI 0.15-0.96) and 0.85 (95%-CI 0.62-0.97), respectively (p=0.04). The specificity of the surgeon predictions was similar to the model's specificity (p=0.25). The net benefit analysis showed better decision-making based on the prediction model compared to the surgeons' decision making (i.e., more correctly predicted improvements and/or fewer incorrectly predicted improvements).

Conclusion

The prediction model outperformed surgeon predictions of improvement following CTR in terms of calibration, accuracy, and sensitivity. Furthermore, the net benefit analysis indicated that using the prediction model instead of relying solely on surgeon decision-making increases the number of patients who will improve after CTR, without increasing the number of unnecessary surgeries.

INTRODUCTION

Surgical treatment of carpal tunnel syndrome (CTS) generally has superior outcomes than nonsurgical treatment[1]. Although surgical treatment is low-risk and generally successful[2,3], not all patients improve, and for some returning to work takes months[2,4]. Therefore, predicting the individual probability of a successful outcome would be useful. This will help inform patients about treatment options, supporting decision-making of both patients and surgeons. This may help increase patient satisfaction and treatment success and avoid unnecessary healthcare consumption[5,6].

We previously developed a prediction model to predict the probability of patient-reported symptom improvement following carpal tunnel release (CTR)[7]. The model predicts the patients' probability of achieving a clinically relevant improvement six months post-surgery, using five patient-reported outcome measure (PROM) scores as predictors. In temporal validation, our model showed good calibration and appropriate discriminative ability with an area under the curve (AUC) of 0.71. The model is available online[9].

Besides sufficient performance in internal and external validation, determining the clinical usefulness of prediction models is important, but often overlooked[10]. Several factors contribute to the success of prediction models in clinical practice[11-14]: 1) whether the model's predictions are more accurate than predictions made by clinicians, 2) how prediction results are integrated into clinical practice, and 3) use and acceptance of prediction models by clinicians. Furthermore, for prediction models to be of added value, they should result in better medical decision-making. This study aims to assess our prediction model's clinical value by comparing its performance to the performance of surgeons predicting improvement following CTR and evaluate the effect on decision-making.

PATIENTS AND METHODS

We conducted a prospective observational study at Xpert Clinics between January 1st and October 13th, 2021. Xpert Clinics comprises 25 locations for hand and wrist care in The Netherlands, employs 23 European Board (FESSH) certified hand surgeons and offers a hand surgery fellowship program. Patients treated at Xpert Clinics are invited to routinely complete PROMs. This cohort and data collection have been described elsewhere[15]. The study was approved by the local institutional review board and reported according to The Strengthening the Reporting of Observational Studies in Epidemiology Statement[16]. All patients provided informed consent.

We considered patients scheduled for CTR eligible for inclusion. Patients who did not complete the Boston Carpal Tunnel Questionnaire (BCTQ) before treatment and six months postoperatively, were excluded, since these scores were used to assess clinical improvement.

Patients were diagnosed and treated by hand surgeons. Diagnoses were based on clinical symptoms, findings on physical examination, and nerve conduction studies when required. All patients underwent a mini-open CTR and received standard postoperative care[17].

Our primary outcome was the Symptom Severity Scale (SSS) of the BCTQ, and additionally, we used visual analogue scales (VAS) to assess hand function and pain. The BCTQ is specifically developed for CTS. It consists of the SSS and the functional status scale (FSS). Both scales range from 1-5, with higher scores representing more severe CTS[18]. Since relieving symptoms (tingling or pain) is generally most important, we focused on the SSS[19]. We used the MCID of 0.8 to define a clinically relevant improvement[8].

Additionally, patients were invited to complete PROMs on their mindset towards their condition and treatment, and on their mental health. These questionnaires included the Brief Illness Perception Questionnaire (B-IPQ), the Credibility and Expectancy questionnaire (CEQ), the Pain Catastrophizing Scale (PCS), and the Patient Health Questionnaire-4 (PHQ-4)[20-23]. All PROMs were sent via email after the first consultation.

The prediction model for clinical improvement after CTR requires five subscales of previously mentioned PROMs at baseline as predictors: the BCTQ-SSS, VAS hand function score, CEQ Expectancy Score, PHQ-4 depression score, and the B-IPQ illness comprehension score. These predictors were selected through recursive feature elimination from multiple potential predictors, including demographic factors and PROM scores[7]. Missing data on these items were imputed through K-nearest neighbors by the prediction model. The predictions were made retrospectively, but were based on PROMs completed by patients prior to surgery (i.e., the model only had access to preoperative data).

Surgeons received information on the study protocol, the prediction model, the BCTQ-SSS and the MCID. After the consultation but before the point where patients completed the baseline PROMs, surgeons answered three questions on their expectations about the improvement of each new patient scheduled for CTR. Surgeons completed these questions on the day of the first consultation.

1. What do you think the patient scores on the SSS at this moment (from 1 to 5)?
2. What do you think the patient will score on the SSS 6 months post-surgery (from 1 to 5)?
3. What is, according to you, the probability that the patient will improve with at least the MCID (0.8 points) on the SSS, 6 months post-surgery?

The first two questions assessed the surgeons' understanding of the SSS, because insufficient understanding might affect their predictions. The third question was used for comparison with model predictions.

Six months post-surgery, patients were invited to complete the BCTQ and VAS pain and function. Satisfaction with treatment result was assessed as part of routine outcome

measurements[24]. Patients failing to complete the BCTQ after two e-mail reminders were contacted by phone once. The difference between the SSS at intake and six months post-surgery was calculated to determine whether the patient reached a clinically relevant symptom improvement.

We aimed to detect a difference in accuracy of 80% correct model predictions and 70% correct surgeon predictions with a power of 0.80 and a two-sided alpha of 0.05. This resulted in a sample size of 146 patients. To account for potential drop-out, we aimed for a sample size of approximately 200 patients. However, our inclusion stopped prematurely due to an administrative change in the workflow, preventing us to collect the surgeon predictions in the same way. Therefore, the study inclusion was stopped before reaching the required sample size.

We performed a non-responder analysis comparing patients with complete BCTQ-scores and patients with missing BCTQ-scores at baseline and six months postoperatively. T-tests were used for normally distributed variables and Wilcoxon tests were used for nonnormally distributed variables.

We used the Wilcoxon signed rank test to compare surgeons' estimates of the SSS to observed scores at baseline and six months post-surgery. Paired t-tests were used to compare estimated SSS change scores to observed scores.

We evaluated the prediction model's and surgeons' performance using calibration, discrimination, and accuracy. Calibration measures the accordance between predicted probabilities and observed frequencies of events. This indicates whether, for example, of 10 patients with a predicted probability of 0.6, we observe that six patients actually improve.We considered calibration the most important outcome, since using estimations from models with poor calibration can lead to misinforming patients on the likely success of a treatment[25]. Calibration was visually assessed using calibration belts[26]. Discrimination refers to the ability to distinguish between patients who will improve and patients who will not[27]. Discrimination was assessed with the area under the curve (AUC). Similar to previous literature, we considered an AUC below 0.70 as suboptimal, 0.70-0.79 as good, and equal or above 0.80 as excellent[28]. We used the DeLong's test to compare the AUCs of the model and of surgeons[29]. Accuracy refers to the percentage of correctly predicted outcomes (i.e., improved or not improved). Accuracy, sensitivity, and specificity were compared using McNemar's test.

Finally, the net benefit of using the prediction model was assessed using decision curve analysis (DCA)[30]. With DCA, we evaluated the effect of different decision-making strategies: 1) "treat none", 2) "current decision-making", and 3) "deciding based on the prediction model". Current practice indicates that surgeons identify patients eligible for CTR based on their expertise. DCA compares the net benefit of each strategy, combining the benefits

and harms associated with each strategy, calculated as a weighted difference between the true positives and false positives. True positives represent cases where the decision-maker correctly identified patients who benefit from CTR, while false positives indicate cases where the decision-maker suggested CTR, but patients did not benefit. A higher net benefit indicates that a decision strategy results in more true positives or fewer false positives than other strategies. This implies that more patients who truly benefit from surgery are correctly scheduled for CTR, while fewer patients who will not benefit from CTR are incorrectly scheduled for CTR[30]. The net benefit is calculated across a range of threshold probabilities, representing the predicted probability at which the surgeon is willing to choose one option (e.g., CTR) over the other (e.g., no CTR). So, in other words, it reflects the point at which the benefits of a decision outweigh the potential risks. A threshold of 10% indicates that the surgeon feels the benefits of the CTR outweigh the risks if the patient has more than 10% chance of improvement following CTR. The threshold probability varies depending on individual perspectives and the specific context of the decision. For example, a higher threshold probability indicates that surgeons are more cautious and require a higher level of confidence before choosing CTR.

Analyses were performed using R statistical programming (version 4.2.2). Statistical significance was determined at $p<0.05$.

RESULTS

During the study period, 205 patients were eligible for inclusion. Ninety-seven patients completed the BCTQ before surgery and six months postoperatively and were included for analysis (Supplementary Figure 1). No significant differences were found in the non-responder analysis (Supplementary Table 1).

The mean age of included patients was 57 years (SD 12) and 66% were female (Table 1). The mean preoperative BCTQ-SSS was 2.9 (SD 0.7) (Supplementary Table 2). Patients who reached the MCID were more satisfied with their treatment result and had better VAS pain and function scores (Supplementary Table 3).

Twelve hand surgeons and one fellow participated in the study. Ninety-two percent were male and 85% were trained as a plastic surgeon (Table 2). Surgeons were FESSH-certified, or fellows trained in hand surgery.

Surgeons overestimated the preoperative symptom severity of patients (Wilcoxon effect size 0.61, $p<0.001$) (Supplemental Figure 2). There was no difference between the surgeon-predicted and observed SSS six months post-surgery (Wilcoxon effect size 0.10, $p=0.90$). The surgeon-predicted improvement from baseline to six months post-surgery was larger than the observed improvement (Cohen's d 0.94, $p<0.001$). This indicates that surgeons

believed patients had more severe symptoms before surgery and that surgeons predicted a greater improvement post-surgery than was observed.

The model's calibration curve did not show any deviations, indicating good calibration, while we found that the surgeon predictions had poor calibration (Figure 1). Specifically, the surgeons' calibration curve deviated at predicted probabilities below 0.3 and above 0.8. This means that surgeon predictions with probabilities of symptom improvement below 0.3 or above 0.8 are unreliable.

Accuracy (0.78 (95%-CI 0.67-0.89) versus 0.65 (95%-CI 0.37-0.78), p=0.03) and sensitivity (0.85 (95%-CI 0.62-0.97) versus 0.72 (95%-CI 0.15-0.96), p=0.04) were higher for the prediction model (Figure 2). However, DeLong's test showed no difference in discriminative ability in surgeon (AUC 0.62, 95%-CI 0.49-0.74) and model predictions (AUC 0.77, 95%-CI 0.66-0.89, p=0.05). Similarly, we found no difference in specificity of the model (0.62, 95%-CI 0.42-0.88) compared to specificity of surgeons (0.46, 95%-CI 0.19-1.00, p=0.25).

The DCA showed a higher net benefit of "deciding based on the prediction model" compared to current decision-making ("Treat all"; Figure 3).

DISCUSSION

We compared surgeon predictions to a previously developed prediction model for patient-reported symptom improvement after CTR. Our model outperformed surgeons in terms of calibration, sensitivity and accuracy. Surgeons overestimated the baseline symptom severity of patients and the treatment effect after surgery. Furthermore, the DCA showed better decision-making when using the model instead of relying solely on surgeon expertise. This suggests that our model could be of added value in clinical practice by providing individual information on expected outcomes and supporting decision-making. Model performance was similar to the performance we found in previous temporal validation[7].

Evaluating patients and informing them of the expected outcomes of CTR is routine for surgeons, so it is likely they perform well on this task. Consequently, for a prediction model to be valuable, it should at least perform equally well as, and preferably better than, surgeons.

The model had access to several PROM scores while the surgeons do not systematically collect this information, which could be considered an unequal comparison. However, surgeons usually have years of experience in assessing the prognosis and symptom severity of their patients, know the actual patient behind the data, and generally obtain more detailed anamnestic and diagnostic information from their consultations and physical examination, such as whether thenar atrophy is present. Therefore, we believe that both the model and

surgeons have access to sufficient relevant data to be able to predict the probability of improvement.

Although there is ample research on the development of prediction models for outcomes after treatment, we were unable to find previous studies assessing the value of prognostic prediction models by comparing the model-based predictions with surgeon estimates. There is, however, some research on this subject for other applications, such as for medical imaging[31,32], outcomes, such as predicting patient survival in cancer patients[33,34], disease progression[35], and complications after cardio-thoracic surgery[36]. In a systematic review on prediction models in medical imaging for diagnostics, Nagendran et al. found that most studies reported prediction models to have similar or superior performance compared to surgeons, with only two studies reporting prediction models performing worse[31]. Similarly, the results of Kuo et al. suggest that prediction models are not only noninferior to surgeons in diagnostic performance in fracture detection on medical imaging, but also that surgeon performance improved with the assistance of an artificial intelligence-based prediction model[32].

The results of this study were presented and discussed with the participating surgeons. Surgeons indicated that, for standard procedures, they are confident to rely on their own knowledge and experience to educate patients about the expected improvement. Therefore, they only find prediction models useful when they are uncertain about the treatment outcome. The prediction model may also have more difficulty predicting the probability of improvement for these relatively difficult cases. Future research should evaluate whether the prediction model has added benefit for these cases. However, our results indicate that, even for a standard procedure like CTR, using the prediction model resulted in improved decision-making compared to relying only on the surgeons' expertise for all patients. Additionally, surgeons indicated that they sometimes schedule surgery to prevent (further) nerve damage, instead of symptom improvement. Finally, surgeons expressed skepticism about prediction models' ability to capture the entire context of a patient's situation. However, we do not intend for prediction models to replace the shared-decision making, but rather to complement this process by providing individualized information. We believe prediction models can be valuable in patient education, enabling patients to become more involved.

We were unable to reach our intended sample size of 146 because of an administrative change in workflow, resulting in our study being underpowered. Therefore, we consider our findings "preliminary conclusions" and recommend repeating this study in a larger sample. Our model predicts the probability of reaching an improvement of 0.8 on the BCTQ-SSS, which may not be the best indicator for a successful outcome. However, patients who improved beyond 0.8 points were more likely to be satisfied with their treatment result and

had less pain six months post-surgery, suggesting it can be considered as an appropriate indicator of successful outcome.

We only included patients who were scheduled for CTR. In our clinics, there are differences in PROMs depending on whether patients are treated surgically or nonsurgically. Consequently, the BCTQ at the relevant time points for comparison with predictions was only available for patients scheduled for CTR. Limiting the study to patients scheduled for surgery could affect the surgeons' responses because of cognitive dissonance avoidance: "I choose to perform CTR, therefore it must help". If surgeons were less sure, they could induce nocebo in patients. Future studies should evaluate the clinical benefit and external validity of the prediction model in a broader population of symptomatic patients with CTS. This will provide valuable information for the assessment of clinical usefulness and implementation.

CONCLUSION

Predicting clinical improvement after CTR could help manage patient expectations and improve pre-operative patient selection. This preliminary study comparing surgeon predictions of symptom improvement after CTR to model predictions, indicated that surgeons showed a worse performance than our model, based on calibration, accuracy, and sensitivity. The decision curve analysis indicated that if surgeons would use our prediction model during clinical practice, they would more frequently select patients who will benefit from surgery without increasing the number of unnecessary procedures. While future studies with larger sample sizes are needed to validate our findings, we believe our results show that prediction models are promising tools with added value for decision-support.

REFERENCES

1. Jarvik JG, Comstock BA, Kliot M, et al. Surgery versus non-surgical therapy for carpal tunnel syndrome: a randomised parallel-group trial. *Lancet*. Sep 26 2009;374(9695):1074-81. doi:10.1016/s0140-6736(09)61517-8
2. Katz JN, Keller RB, Simmons BP, et al. Maine Carpal Tunnel Study: outcomes of operative and nonoperative therapy for carpal tunnel syndrome in a community-based cohort. *J Hand Surg Am*. Jul 1998;23(4):697-710. doi:10.1016/s0363-5023(98)80058-0
3. Louie DL, Earp BE, Collins JE, et al. Outcomes of open carpal tunnel release at a minimum of ten years. *J Bone Joint Surg Am*. Jun 19 2013;95(12):1067-73. doi:10.2106/jbjs.L.00903
4. Newington L, Stevens M, Warwick D, Adams J, Walker-Bone K. Sickness absence after carpal tunnel release: a systematic review of the literature. *Scand J Work Environ Health*. Nov 1 2018;44(6):557-567. doi:10.5271/sjweh.3762
5. Marks M, Audigé L, Reissner L, Herren DB, Schindele S, Vliet Vlieland TP. Determinants of patient satisfaction after surgery or corticosteroid injection for trapeziometacarpal osteoarthritis: results of a prospective cohort study. *Arch Orthop Trauma Surg*. Jan 2015;135(1):141-7. doi:10.1007/s00402-014-2119-0
6. Swarup I, Henn CM, Gulotta LV, Henn RF, 3rd. Patient expectations and satisfaction in orthopaedic surgery: A review of the literature. *J Clin Orthop Trauma*. Jul-Aug 2019;10(4):755-760. doi:10.1016/j.jcot.2018.08.008
7. Hoogendam L, Bakx JAC, Souer JS, Slijper HP, Andrinopoulou ER, Selles RW. Predicting Clinically Relevant Patient-Reported Symptom Improvement After Carpal Tunnel Release: A Machine Learning Approach. *Neurosurgery*. Jan 1 2022;90(1):106-113. doi:10.1227/neu.0000000000001749
8. Leite JC, Jerosch-Herold C, Song F. A systematic review of the psychometric properties of the Boston Carpal Tunnel Questionnaire. *BMC Musculoskelet Disord*. Oct 20 2006;7:78. doi:10.1186/1471-2474-7-78
9. Hoogendam L. Prediction of symptom improvement after carpal tunnel surgery. https://analyse.equipezorgbedrijven.nl/shiny/cts-prediction-model-Eng/
10. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. Jan 2010;21(1):128-38. doi:10.1097/EDE.0b013e3181c30fb2
11. Parikh RB, Manz CR, Nelson MN, et al. Clinician perspectives on machine learning prognostic algorithms in the routine care of patients with cancer: a qualitative study. *Supportive Care in Cancer*. 2022/05/01 2022;30(5):4363-4372. doi:10.1007/s00520-021-06774-w
12. O'Neill C, Hutchings HA, Abbott Z, Battle C. Prognostic prediction tools and clinician communication: a qualitative study of the effect of the STUMBL tool on clinical practice. *BMC Emergency Medicine*. 2020/05/11 2020;20(1):36. doi:10.1186/s12873-020-00331-0
13. Sharma V, Ali I, van der Veer S, Martin G, Ainsworth J, Augustine T. Adoption of clinical risk prediction tools is limited by a lack of integration with electronic health records. *BMJ Health & Care Informatics*. 2021;28(1):e100253. doi:10.1136/bmjhci-2020-100253
14. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *Bmj*. Jun 4 2009;338:b606. doi:10.1136/bmj.b606

15. Selles RW, Wouters RM, Poelstra R, et al. Routine Health Outcome Measurement: Development, Design, and Implementation of the Hand and Wrist Cohort. *Plast Reconstr Surg*. Aug 2020;146(2):343-354. doi:10.1097/prs.0000000000007008

16. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. Oct 20 2007;370(9596):1453-7. doi:10.1016/s0140-6736(07)61602-x

17. van den Broeke LR, Theuvenet WJ, van Wingerden JJ. Effectiveness of mini-open carpal tunnel release: An outcome study. *Arch Plast Surg*. Jul 2019;46(4):350-358. doi:10.5999/aps.2018.00535

18. Levine DW, Simmons BP, Koris MJ, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Joint Surg Am*. Nov 1993;75(11):1585-92. doi:10.2106/00004623-199311000-00002

19. Gong HS, Baek GH, Oh JH, Lee YH, Jeon SH, Chung MS. Factors affecting willingness to undergo carpal tunnel release. *J Bone Joint Surg Am*. Sep 2009;91(9):2130-6. doi:10.2106/jbjs.H.01221

20. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry*. Jun 2000;31(2):73-86. doi:10.1016/s0005-7916(00)00012-4

21. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. *Psychological Assessment*. 1995;7:524-532. doi:10.1037/1040-3590.7.4.524

22. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. *J Psychosom Res*. Jun 2006;60(6):631-7. doi:10.1016/j.jpsychores.2005.10.020

23. Kroenke K, Spitzer RL, Williams JB, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics*. Nov-Dec 2009;50(6):613-21. doi:10.1176/appi.psy.50.6.613

24. De Ridder WA, Wouters RM, Hoogendam L, Vermeulen GM, Slijper HP, Selles RW. Which Factors Are Associated With Satisfaction With Treatment Results in Patients With Hand and Wrist Conditions? A Large Cohort Analysis. *Clin Orthop Relat Res*. Jul 1 2022;480(7):1287-1301. doi:10.1097/corr.0000000000002107

25. Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. *BMC Medicine*. 2019/12/16 2019;17(1):230. doi:10.1186/s12916-019-1466-7

26. Nattino G, Finazzi S, Bertolini G. A new calibration test and a reappraisal of the calibration belt for the assessment of prediction models based on dichotomous outcomes. *Stat Med*. Jun 30 2014;33(14):2390-407. doi:10.1002/sim.6100

27. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. Apr 1982;143(1):29-36. doi:10.1148/radiology.143.1.7063747

28. D'Agostino RB, Sr., Pencina MJ, Massaro JM, Coady S. Cardiovascular Disease Risk Assessment: Insights from Framingham. *Glob Heart*. Mar 2013;8(1):11-23. doi:10.1016/j.gheart.2013.01.001

29. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. Sep 1988;44(3):837-45.

30. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res*. 2019;3:18. doi:10.1186/s41512-019-0064-7

31. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *Bmj*. Mar 25 2020;368:m689. doi:10.1136/bmj.m689

32. Kuo RYL, Harrison C, Curran TA, et al. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. *Radiology*. Jul 2022;304(1):50-62. doi:10.1148/radiol.211785

33. Hui D, Park M, Liu D, et al. Clinician prediction of survival versus the Palliative Prognostic Score: Which approach is more accurate? *Eur J Cancer*. Sep 2016;64:89-95. doi:10.1016/j.ejca.2016.05.009

34. Farinholt P, Park M, Guo Y, Bruera E, Hui D. A Comparison of the Accuracy of Clinician Prediction of Survival Versus the Palliative Prognostic Index. *J Pain Symptom Manage*. Mar 2018;55(3):792-797. doi:10.1016/j.jpainsymman.2017.11.028

35. Chuah A, Walters G, Christiadi D, et al. Machine Learning Improves Upon Clinicians' Prediction of End Stage Kidney Disease. *Front Med (Lausanne)*. 2022;9:837232. doi:10.3389/fmed.2022.837232

36. Rank N, Pfahringer B, Kempfert J, et al. Deep-learning-based real-time prediction of acute kidney injury outperforms human predictive performance. *NPJ Digit Med*. 2020;3:139. doi:10.1038/s41746-020-00346-8

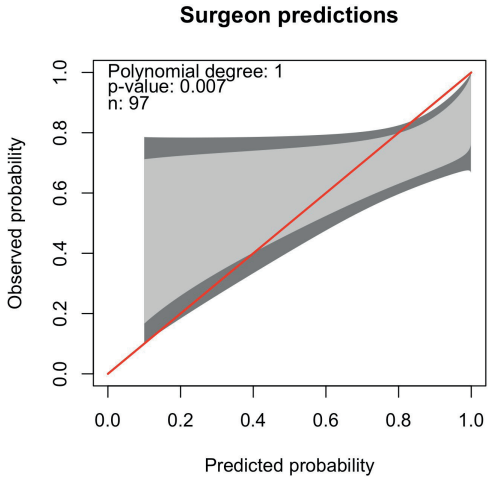


Figure 1A

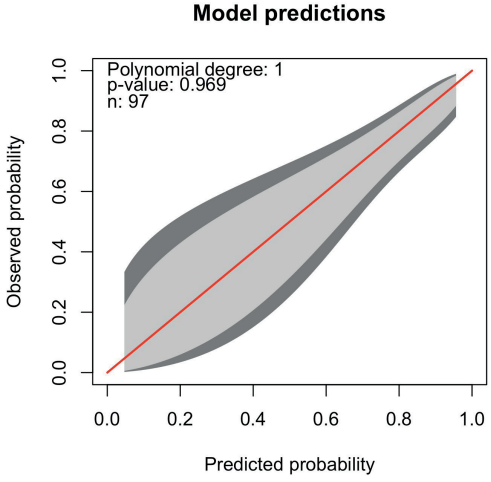


Figure 1B

Figure 1. Calibration belts of the surgeon predictions (A) and the prediction model (B). The predicted probability of symptom improvement compared to the observed probability of symptom improvement is shown. The red line indicates perfect calibration and the 80% CI (light grey) and 95% CI (dark grey) of the calibration is shown in grey. The model or surgeons perform well on calibration if the belt is close to the bisector. If the belt falls above the bisector, the model or surgeons underestimate the probability of improvement. If the belt falls under the bisector, the model or surgeons overestimate the probability. **A**. Surgeon predictions: the red line does not fall within the 80% CI (light grey) and 95% CI (dark grey) over the whole range of predicted probability, indicating significant deviation in calibration between the predicted and observed probability of improvement. This is also confirmed by the p-value < 0.05. Patients for whom

surgeons predict the probability of symptom improvement to be higher than 80%, the observed probability of symptom improvement is lower, indicating an overly optimistic surgeon prediction. **B**. Prediction model: the red line falls within the 80% CI (light grey) and 95% CI (dark grey), indicating no significant deviation in calibration between the predicted and observed probability of improvement. This is also confirmed by the p-value > 0.05.

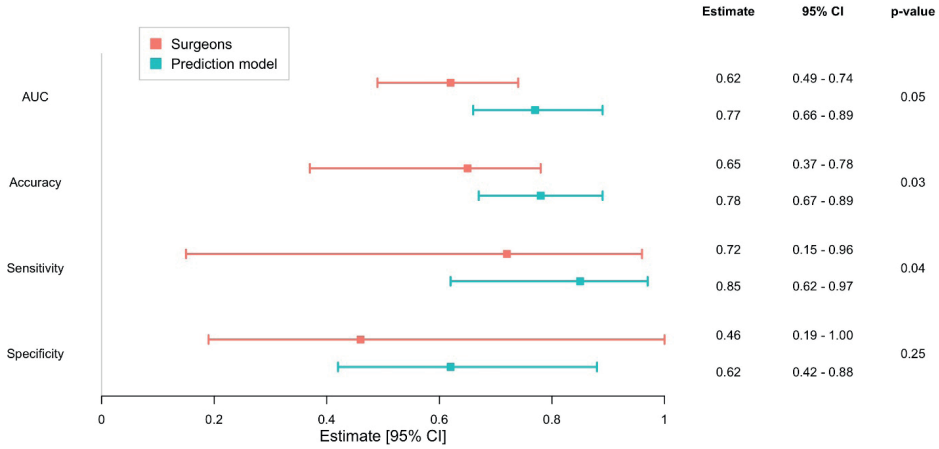


Figure 2. Comparison of performance measures (AUC, accuracy, sensitivity and specificity) of surgeon predictions and model predictions; Higher scores indicate better performance. For all measures, estimates are displayed with 95% confidence intervals. Additionally, the p-value for the comparisons between surgeon and model predictions is shown. A significantly higher sensitivity for the prediction model compared to surgeon predictions is seen.

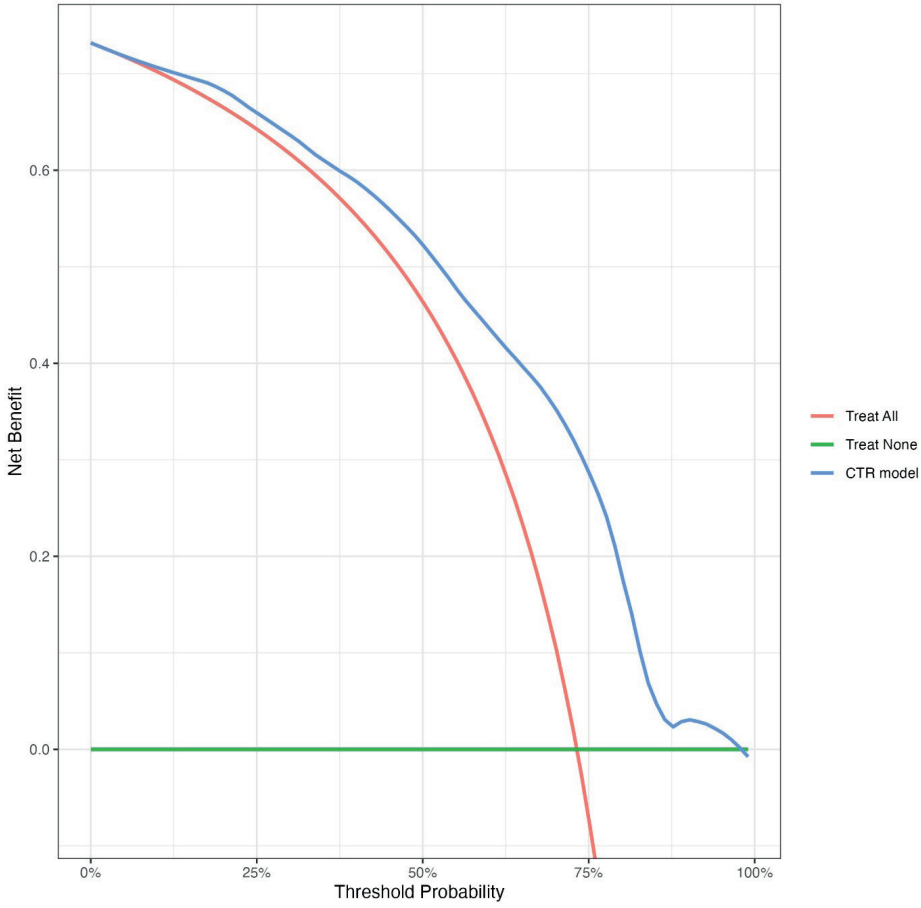


Figure 3. Net benefit curve for three decision-making strategies: "Treat all", "Treat none", and "Deciding based on the CTR prediction model". "Treat all" is, in our sample equal, to the current decision-making of surgeons. The net benefit weighs the benefits (i.e., true positives) and harms (i.e., false positives) of a decision strategy over a range of threshold probabilities. The threshold probability (on the x-axis) reflects the point at which the benefits of a particular decision or strategy outweigh the potential harm. We observe a benefit of "Deciding based on the CTR prediction model" compared to "Treat all" from a threshold probability of 10% onwards. A threshold of 10% would indicate that the surgeon feels the benefits of the CTR outweigh the risks if the patient has more than 10% chance of improvement following CTR. Given the elective nature of CTR, it is likely that surgeons would only schedule their patients for this procedure when the patient has a high probability of improving. Therefore, it is likely that the threshold probability for choosing CTR lies above 10%. The higher net benefit of "Deciding based on the CTR prediction model" indicates that this strategy results in more true positives and/or less false positives compared to current decision-making ("Treat all").

Table 1. Characteristics of the included patients

	Included patients (n=97)
Age (in years)	57.0 (12.4)
Gender, n (%)	
Female	64 (66)
Duration of symptoms (in months)	9.0 [6.0, 24.0]
Type of work, n (%)	
Unemployed	34 (35)
Light physical labor	29 (30)
Moderate physical labor	21 (22)
Heavy physical labor	13 (13)
Affected side, n (%)	
Left	41 (42)
Right	55 (57)
Both	1 (1)
Recurrent CTS, n (%)	
Yes*	3 (3)
Second opinion, n (%)	
Yes	1 (1)

* Patients with recurrent CTS were previously treated elsewhere

Table 2. Characteristics of the participating hand surgeons

	Participating surgeons (N = 13)
Gender, n (%)	
Female	1 (8)
Years of experience as specialist	9.69 (5.25)
Type of employment, n (%)	
Fellow hand surgery	1 (8)
Trained hand surgeon	12 (92)
Type of training, n (%)	
Plastic surgery	11(85)
Orthopaedic surgery	2 (15)
Solely focused on hand surgery	
Yes	9 (70)
No	4 (30)
European board (FESSH) certified	
Yes	12 (96)
No	1 (8)

FESSH = Federation of European Societies for Surgery of the Hand

20

GENERAL DISCUSSION

GENERAL DISCUSSION

To optimize the shared decision-making process, patients considering elective treatment for hand or wrist conditions must have relevant and unbiased information about the treatment options, risks, and benefits. Therefore, this thesis aimed to provide patients and their clinicians with clinically relevant, personalized, and data-driven outcome information, helping them to make well-informed treatment decisions for hand and wrist conditions.

This general discussion is divided into four parts, following the structure of this thesis: I) Determining clinically important outcome values, II) Clinically important outcomes of treatment (strategies) for thumb base osteoarthritis, III) Prognostic factors for clinically important treatment outcomes (for thumb base osteoarthritis), and IV) Development, validation, and added value of prediction models in hand surgery. The main findings, limitations, implications, and future perspectives are discussed below for each part.

PART I - DETERMINING CLINICALLY IMPORTANT OUTCOME VALUES

In the first part of this thesis, we determined clinically important outcome values for Patient Reported Outcome Measures (PROMs) that are frequently used to measure outcomes of hand and wrist treatments. In **Chapter 2**, we provided estimates of the minimal improvement patients require to consider it a meaningful change (Minimally Important Change (MIC)) based on two frequently used methods. Notably, we found that the MIC was significantly lower for nonsurgically treated patients compared to surgically treated patients, indicating that surgically treated patients want to improve more following treatment. In **Chapter 3**, we presented a new anchor-based method to calculate Patient Acceptable Symptom State (PASS) estimates using item response theory. We demonstrated that this method outperformed previous, frequently used methods in terms of reliability, precision, measuring error, non-normally distributed outcomes, and class imbalance of the anchor. In **Chapter 4**, we used this new method to provide estimates of the PASS for three multi-item PROMs for 35 diagnosis-treatment combinations for hand and wrist disorders.

Implications and future perspectives of Part I

Our results from Chapters 2, 3, and 4 demonstrate that the choice of calculation method for a clinically important outcome value has a notable effect on the obtained estimates. Calculation methods based on item response theory, such as those presented for the PASS in Chapter 3 (1) and the MIC elsewhere (2, 3), are considered the most reliable methods available at the moment and are therefore recommended to be used when the assumptions of these methods are met. However, further method development is needed to provide reliable estimates for single-item scores, such as Visual Analog Scales.

The PASS values we provided in Chapter 4 for PROMs used in hand-wrist care can be considered the currently preferred when interpreting treatment results on a group level, based on this simulation study we performed in Chapter 3. The MIC values we provided in Chapter 2 can also be considered the best available estimates but should be interpreted with caution in the cases where the improved and not-improved groups differed remarkably in size, as these estimates may be biased (4). In that case, we recommend re-estimating the MIC with the newly developed methods based on Item Response Theory (5). There is also a need for Substantial Clinical Benefit estimates, referring to the change a patient needs to experience to consider themselves substantially improved, for hand and wrist conditions using the abovementioned calculation methods (6, 7), which may be addressed in future studies.

The MIC and PASS values described in Chapters 2, 3, and 4 are relevant on a group level but cannot be applied directly to individual patients. As each patient has a personal internal threshold where they will consider themself to be "relevantly improved" (in case of the Minimally Important Change) or consider the treatment effect to be "satisfactory" (in case of the Patient Acceptable Symptom State), the clinically important outcome values we estimated can be viewed as the mean threshold that is relevant for the population of interest. There may be differences within each group. For example, an elderly patient might be satisfied with a slight improvement in functioning, while a professional athlete will likely demand a more considerable functional improvement to be satisfied. This illustrates that each patient has an internal threshold to determine whether a treatment effect is relevant. To overcome this, other personalized clinically important outcome values need to be introduced. An example of such a personalized value is the Personal Meaningful Gain (8), where patients can choose what outcome domain they wish to improve, how they score now on that outcome, and what score they need to be satisfied with the treatment. Using such a personalized measure as an outcome for a prediction model could help align predictions of treatment success with the individual's goals.

PART II - CLINICALLY IMPORTANT OUTCOMES OF TREATMENT (STRATEGIES) FOR THUMB BASE OSTEOARTHRITIS

In the second part of this thesis, we determined the outcomes of several nonsurgical and surgical treatment options for thumb base osteoarthritis. In **Chapter 5**, we demonstrated that such patients who underwent hand therapy, possibly combined with an orthosis, report a sustained positive effect on pain at a median follow-up of 7 years compared to 1 year after treatment, and only 22% underwent surgery. In **Chapter 6**, we found a complication rate of 35% (according to the International Consortium for Health Outcomes Measurement Complications in Hand and Wrist conditions (ICHAW) classification) after trapeziectomy

with Weilby plasty, with 16% being Grade 1 and 19% being Grade 2/3 events. As Grade 1 events were not associated with clinically relevant worse PROM scores at twelve months and their more subjective nature, we proposed reclassifying Grade 1 events as "adverse protocol deviations" and considering Grade 2 and 3 as complications in future studies. In **Chapter 7**, we compared treatment outcomes of trapeziectomy with Anchovy plasty, Weilby plasty, Burton-Pellegrini plasty, and Zancolli plasty. We found no differences in pain at twelve months, although secondary outcomes showed a slight preference for trapeziectomy with Anchovy plasty. In **Chapter 8**, we evaluated the cost-effectiveness of four potential treatment strategies for thumb base osteoarthritis over a 10-year period, including treatment-related and societal costs. We found that starting with nonsurgical treatment, considering surgery if needed, was the most cost-effective treatment strategy.

Implications and future perspectives of Part II

The findings in Chapter 5 and Chapter 8 emphasize the importance of starting with nonsurgical treatment (i.e., hand therapy, possibly combined with an orthosis) for patients with primary thumb base osteoarthritis. We have shown that nonsurgical treatment has a long-term positive effect on pain, that few patients elect surgery afterward, and that this is the most cost-effective treatment strategy (of those we compared), making it a promising first-line treatment. Still, in clinical practice, we see variation in the exact treatment content of nonsurgical treatment for primary thumb base osteoarthritis, for example, in the number of hand therapy sessions patients receive or whether an orthosis is prescribed. This allows for future studies regarding the most optimal nonsurgical treatment for primary thumb base osteoarthritis (9), which could improve outcomes of nonsurgical treatment even further.

Despite many experimental and observational studies, no clear superior surgical technique currently exists to treat patients with thumb base osteoarthritis (10). In line with that, we did not find a clear difference in surgical treatment outcomes between trapeziectomy combined with four different tendon plasties in Chapter 7. Considering the comparable surgical treatment outcomes, it becomes more important to take treatment costs and societal costs (e.g., due to work absenteeism or presenteeism) into account. While this receives little attention in the current literature (11), comprehensive studies on (comparing) cost-effectiveness could further inform policymakers, clinicians, and patients and help to improve treatment guidelines. Such studies would provide relevant information, especially for more costly treatment options with hypothesized faster return to work, such as prostheses for the first carpometacarpal joint and lipofilling of the first carpometacarpal joint.

To systematically compare treatment outcomes and perform benchmarking, it is essential that standardized definitions are used in all settings. Particularly for complications, many different definitions and classifications are used, which made it difficult to compare our findings (from Chapter 7 for example) to previously published complication rates.

To overcome these difficulties, the International Consortium for Health Outcomes Measurement (ICHOM) Hand and Wrist working group introduced the ICHAW classification (12), primarily based on the Clavien-Dindo complication classification system that is mainly used in general surgery (13). As the ICHAW classification is based on the action taken in response to the complication, we believe that the ICHAW classification is an important step towards standardizing complication reporting in hand surgery. However, there is still subjectivity in the actions taken by clinicians (i.e., when a treatment will be prescribed and of what invasiveness). Notably, discretionary treatments for minor discomfort (e.g., prescription of silicone gel sheets for scar tenderness) may lead to a high number of ICHAW grade 1 events (14), which we have therefore proposed to reclassify as "adverse protocol deviations" instead of complications (15). We expect that the rating of ICHAW grade 2 and 3 events will be more objective and that this reclassification of the ICHAW events will contribute to a more standardized definition of complications in hand surgery.

PART III - PROGNOSTIC FACTORS FOR CLINICALLY IMPORTANT TREATMENT OUTCOMES (FOR THUMB BASE OSTEOARTHRITIS)

As observed in the first two parts of this thesis, there is considerable variation in patient-reported treatment outcomes. In part III, we investigated which factors explain this variation in treatment outcomes. In **Chapters 9 and 10,** we showed how psychological factors are associated with pain and hand function before and three months after the start of nonsurgical treatment for thumb base osteoarthritis. In **Chapter 11**, we found that higher pre-treatment expectations are associated with increased satisfaction with treatment results following nonsurgical thumb base osteoarthritis treatment. Moreover, in **Chapter 12,** we found that this positive association between pre-treatment expectations and satisfaction with treatment results was also present in a broader population of patients receiving one of six different (non)surgical treatments for varying hand and wrist disorders. Because we found pre-treatment expectations to be associated with satisfaction in Chapters 11 and 12, we studied in **Chapter 13** how pre-treatment expectations are related to patient, surgeon, and treatment characteristics. In **Chapter 14**, we confirmed the hypothesis that patients involved in a personal injury claim have worse treatment outcomes than patients not involved in a personal injury claim following surgical or nonsurgical treatment.

Implications and future perspectives of Part III

The findings from Part III repeatedly indicate the association of psychological factors and expectations with treatment outcomes of thumb base osteoarthritis. These factors have been linked consistently to treatment outcomes across various musculoskeletal conditions.

Therefore, clinicians should be attentive to their patients' mental health status. A first indication of mental health status can easily be obtained by applying the ultrashort mental health screening tool that my colleagues recently introduced (16).

In our observational studies, we could not assess causality between psychological factors and expectations with treatment outcomes of thumb base osteoarthritis. However, some randomized controlled trials found promising effects of optimizing expectations on treatment outcomes, including postoperative satisfaction (17, 18). The optimal strategy for interventions on mental health or expectations in musculoskeletal conditions still needs to be determined. For clinicians, we currently recommend assessing their patients' mental health status', using this as a conversation starter, and considering this in their shared decision-making. Additionally, further studies are needed to determine the potential benefit and implementation of integrating mental health interventions into the treatment plan.

Our findings and the literature (19) show a consistent link between positive pre-treatment outcome expectations and treatment outcomes, such as pain, function, and satisfaction. Trials on optimizing expectations have also shown improved treatment outcomes (17, 18). This consistent association between pre-treatment expectations and treatment outcomes emphasizes the necessity of expectation management. Clinicians would benefit from evidence-based guidelines to aid them in setting positive, realistic expectations for their patients. While future studies are needed for guideline development, implementation, and impact assessment, optimizing expectations seems a promising strategy to improve treatment outcomes.

At the same time, clinicians must recognize the often-undeserved negative expectations surrounding nonsurgical treatments, as we found in Chapter 13. Given that patients might harbor unjustified pessimism towards nonsurgical options, clinicians should educate patients about the potential benefits and provide them with positive, realistic outcome information for these treatments. This might motivate patients for their nonsurgical treatment, improve compliance, and result in better nonsurgical treatment outcomes.

PART IV - DEVELOPMENT, VALIDATION, AND ADDED VALUE OF PREDICTION MODELS IN HAND SURGERY

In the final part of this thesis, we studied the potential of prediction modeling to inform patients of their probability of improvement after treatment. In **Chapters 15 and 16,** we developed and internally validated a prediction model for the probability of a clinically relevant improvement in symptoms following carpal tunnel release surgery and surgical treatment for thumb base osteoarthritis. In **Chapter 17**, we temporally validated the Carpal Tunnel Release model presented in Chapter 15 and showed its added value for decision-making by outperforming surgeon decision-making. In **Chapter 18,** we provided an

overview of prediction models for outcomes of upper extremity surgery and the availability of these models for clinicians and critically assessed their risk of bias. Finally, **Chapter 19** compared condition-treatment-specific prediction models to three generic prediction models for many hand and wrist conditions, showing that the generic prediction models performed similarly to the many condition-treatment-specific models.

Implications and future perspectives of Part IV

Prediction models are typically developed for a single diagnosis-treatment combination (20). However, models for single diagnosis-treatment combinations can be time-consuming to develop, validate, and implement, particularly in medical fields with many diagnoses and treatment options. In Chapter 19, we demonstrated a more generic approach, using diagnosis and treatment as model input, is feasible. This approach also has advantages regarding implementation. As the number of resulting models from this approach is manageable, technical implementation aspects and user training can be handled faster. Additionally, the limited number of prediction models makes temporal validation and routine updating of the models more feasible, thereby ensuring continuous good performance and compliance with (inter)national legislation, such as the Medical Device Regulation (21). Also, there may be situations where a generic model provides better predictions than a diagnosis-treatment-specific model, because the generic model was able to learn from many similar diagnosis-treatment combinations. While there will be limits regarding the applicability of this generic approach for developing prediction models, and while more extensive validation is recommended, we believe this approach seems highly efficient and is promising to explore further.

The findings from Chapter 17 show that prediction models can add value to hand surgery and help clinicians make decisions. This is in line with other prediction models in orthopedics that have shown promising results regarding their effect on clinical decision-making (22, 23). However, many prediction models do not find their way into clinical practice and, therefore, cannot positively impact decision-making.

There are several reasons why prediction models rarely reach clinical practice.

First, temporal and external validation of prediction models is needed but rarely performed (20, 24, 25), also for upper extremity conditions (26). A lack of data interoperability may partially cause this lack of external validation research. We need similar data content and similar data coding and storing to achieve data interoperability. There are already several ongoing initiatives to improve and facilitate data interoperability. ICHOM tries to ensure that the same patient characteristics and outcomes at the same time points are measured for each clinical domain by providing standard measurement sets (27). Additionally, several common data models are used in the medical field, allowing data to be coded and stored similarly (28). The Observational Medical Outcomes Partnership (OMOP) Common Data

Model (CDM)(29) is of particular interest. This OMOP-CDM is historically mainly used for pharmacological research but is also gaining traction in the medical field due to its active community, its generalizability across medical specialties, and the large number of software tools available to interact with datasets in the OMOP-CDM (29-31). Increased adoption of the relevant ICHOM measurement sets and data mapping to the OMOP-CDM will greatly improve data interoperability and allow for external validation of existing prediction models. In our setting, we are currently converting our data to the OMOP-CDM as part of the EHDEN project (32) and plan to use this data for further development and validation of prediction models as part of the PREPARE project (33).

Second, after thoroughly validating prediction models, the models need to be implemented into the clinical setting for clinicians to use. Several factors play a role in the successful implementation of prediction models. It is critical that a model predicts an outcome relevant to patients and clinicians and that clinicians feel using the model is of added value (34, 35). Implementation is facilitated by having a learning climate and providing training material to ensure familiarity with the prediction tool (35, 36). Additionally, technical aspects, such as seamless integration with electronic health care records or data collection systems and ease of use, are important facilitators for successfully implementing a prediction tool (37, 38). As successful implementation is crucial for the model to positively impact healthcare decisions and treatment outcomes, I believe there should be more attention to the specific facilitators and barriers for successful implementation in a clinical setting. These need to be considered earlier in the research process, during the design of the prediction model, and relevant stakeholders (e.g., clinicians, patients, IT-specialists) should be included.

For clinicians to gain trust in the predictions made by the model, it is important that the model is transparent (i.e., showing which model features mainly impact the generated prediction for an individual patient (34, 39). Additionally, it may allow the clinician to explain better to the patient why the probability of improvement with a particular treatment is high or low (40). We did not incorporate this into the models presented in this thesis. Still, we recommend incorporating explainable AI methods like SHapley Additive exPlanations (SHAP)(41) or Local Interpretable Model-agnostic Explanations (LIME)(42) to facilitate clinical uptake for future prediction models.

Finally, when a prediction model is successfully implemented in clinical practice, impact assessment studies are needed to study the effect of having the prediction model available and the impact of using the prediction model on treatment selection, patient experiences, and treatment outcomes (43). Ideally, (stepped wedge) cluster randomized controlled trials are used to study all effects of prediction models compared to usual care. However, considering the high costs and long time required to do such a study, prospective studies addressing whether the prediction model positively affects decision-making should be the first step towards impact studies focusing on treatment outcomes. In future studies, we aim

to study the generic models in Chapter 19 in more detail, assessing their potential impact on decision-making and possibly the effect on treatment selection, patient experiences, and patient outcomes.

LIMITATIONS OF THIS THESIS

The studies in this thesis primarily relied on data from routine outcome measurement, which is prone to missingness. Many patients were excluded because they did not complete all questionnaires at all relevant times. Excluding patients due to missing data can lead to selection bias, particularly when patients who did not answer all the questionnaires did so because of their treatment outcomes (i.e., missing not at random)(44). For example, patients who no longer experience symptoms may not see the need to continue answering questionnaires, resulting in biased outcome estimates. In each study, we conducted non-responder analyses to compare the characteristics of those completing all questionnaires to those who did not. While this only provides circumstantial evidence, we found no meaningful indications of selection bias in pre-treatment characteristics for patients included and excluded from the studies. For developing and validating prediction models, missing data was handled using imputation. This allows using data from patients with complete and incomplete data for model development and validation, but also real-time handling of (some) missing data of future patients, making it still possible to obtain predictions for them. Part of our future research is focused on preventing non-response in our routine outcome measurements.

All studies, including the comparative effectiveness study in Chapter 7, used an observational study design. While randomized controlled trials (RCTs) are the gold standard for comparative effectiveness studies, they pose many practical challenges in surgery, including high costs, long study duration, and the difficulty to blind surgeons and often also patients (45). Moreover, while a new medication is typically compared to a placebo, it is often considered unethical and very unappealing to patients to participate in a trial with sham surgery as a control group. Additionally, RCTs' strict in- and exclusion criteria may limit their generalizability. Still, the RCT is the best study method to answer comparative effectiveness research questions. Although infrequently performed, trials with sham surgery as a control group provide valuable information, as they can distinguish specific surgical effects from contextual effects.

The use of observational data for comparative effectiveness studies, such as in this thesis, requires careful design and analysis to ensure comparability between treatment groups. We used techniques like propensity-score matching and regression analysis to account for differences in patient characteristics, allowing for more reliable comparisons of treatment outcomes. Still, the observational nature of the studies makes it difficult to conclude

causality, as confounding by indication may have occurred. As such, the associations we found, for example, between pre-treatment expectations and treatment outcomes (Chapters 10 and 11), should not be interpreted as causal relations. An RCT actively modifying expectations would be the preferred way to assess whether pre-treatment expectations have a causal relation with treatment outcomes.

Chapter 10 evaluated the cost-effectiveness of four potential treatment strategies for thumb base osteoarthritis based on a microsimulation model. The input parameters for the model were derived from the observational Hand-Wrist Study Group cohort, which caused our results to be specific for our clinical setting. While we believe these results are generalizable to thumb base osteoarthritis patients in similar settings, this may not necessarily apply elsewhere.

The prediction models developed and validated in this thesis face the same issue of generalizability. While the models perform well in our setting, this may not necessarily be true for other clinical settings, such as a tertiary hospital (46). We therefore recommend external validation of the models before they are applied elsewhere.

All prediction models presented in this thesis predict the probability of improvement following treatment for a hand or wrist condition, defined as reaching the MIC on a relevant domain for that hand or wrist condition (e.g., pain or hand function). However, as discussed in Part I, these MIC values are meant to interpret PROM scores on a group level. As such, the outcomes of these prediction models may not always reflect what is most relevant for an individual patient. Future studies may focus on predicting other relevant outcomes, such as the probability of complications, time to return to work, or satisfaction with treatment results. Alternatively, predicting personalized outcomes such as reaching a Personal Meaningful Gain may provide patients and clinicians with relevant information to incorporate in the decision-making for patients considering treatment for their hand or wrist condition.

Conclusions and recommendations

To conclude, this thesis provides an overview of clinically important outcome values, outcomes of treatment (strategies), prognostic factors, and prediction models for hand and wrist conditions, with a focus on thumb base osteoarthritis. This thesis aims to improve the shared decision-making process for patients with hand and wrist conditions by providing insight into clinically relevant, personalized, and data-driven outcome information. The findings highlight nonsurgical treatments as first-line options, the association of psychological factors with outcomes, and the potential benefits of prediction models in clinical practice.

Recommendations for researchers based on this thesis

- **Personalized outcome measures**: Explore the use of personalized measures, like the Personal Meaningful Gain, in research to align information about treatment success with individual patient goals.
- **Cost-effectiveness**: Study the cost-effectiveness of treatment strategies, particularly focusing on integrating new treatments and treatments with proven similar effectiveness.
- **Psychological interventions**: Study the feasibility and (cost)effectiveness of interventions targeting psychological factors, such as illness perceptions, depressive symptoms, or anxiety, and expectations on treatment outcomes.
- **Prediction models**: Conduct extensive external validations and impact assessment studies of existing prediction models and further study the potential added value of generic prediction models.

Recommendations for clinicians based on this thesis

- **Nonsurgical treatments**: Prioritize nonsurgical treatment for primary thumb base osteoarthritis, given the long-term positive effect on pain and its cost-effectiveness.
- **Expectation management**: Incorporate expectation management (particularly for nonsurgical treatment) to counsel patients towards positive, realistic expectations to enhance treatment outcomes.
- **Mental health assessment**: Assess patients' mental health status using tools like the Ultrashort Mental Health Screening Tool and consider these factors in the shared decision-making process.
- **Prediction model integration**: Identify validated prediction models and use these after external validation to inform shared decision-making for patients considering treatment.

Recommendations for policymakers based on this thesis

- **Data interoperability**: Support initiatives aimed at improving data interoperability, such as the adoption of ICHOM measurement sets and mapping data to common data models like the OMOP-CDM, to facilitate benchmarking and increased collaborative real-world studies.
- **Standardized reporting**: Promote the adoption of standardized definitions and classifications, such as the ICHAW classification, to ensure consistency in complication reporting and comparative research.
- **Cost-effectiveness**: Encourage cost-effectiveness studies for various treatment options and strategies (including societal costs) to inform guidelines and healthcare policies.
- **Implementation of prediction models**: Facilitate the integration of validated prediction models into clinical practice by supporting the necessary technical infrastructure, training programs for healthcare providers, and support for continuous performance monitoring, ensuring compliance with (inter)national regulations.

REFERENCES

1. Terluin B, Koopman JE, Hoogendam L, Griffiths P, Terwee CB, Bjorner JB. Estimating meaningful thresholds for multi-item questionnaires using item response theory. Qual Life Res. 2023;32(6):1819-30.
2. Bjorner JB, Terluin B, Trigg A, Hu J, Brady KJS, Griffiths P. Establishing thresholds for meaningful within-individual change using longitudinal item response theory. Qual Life Res. 2023;32(5):1267-76.
3. Terluin B, Trigg A, Fromy P, Schuller W, Terwee CB, Bjorner JB. Estimating anchor-based minimal important change using longitudinal confirmatory factor analysis. Qual Life Res. 2024;33(4):963-73.
4. Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. J Clin Epidemiol. 2017;83:90-100.
5. Terluin B, Griffiths P, van der Wouden JC, Ingelsrud LH, Terwee CB. Unlike ROC analysis, a new IRT method identified clinical thresholds unbiased by disease prevalence. J Clin Epidemiol. 2020;124:118-25.
6. Glassman SD, Copay AG, Berven SH, Polly DW, Subach BR, Carreon LY. Defining substantial clinical benefit following lumbar spine arthrodesis. J Bone Joint Surg Am. 2008;90(9):1839-47.
7. Randall DJ, Zhang Y, Li H, Hubbard JC, Kazmers NH. Establishing the Minimal Clinically Important Difference and Substantial Clinical Benefit for the Pain Visual Analog Scale in a Postoperative Hand Surgery Population. J Hand Surg Am. 2022;47(7):645-53.
8. de Ridder WA, van Kooij YE, Slijper HP, Arends GR, de Roode A, MacDermid JC, et al. Tailoring and evaluating treatment with the Patient-Specific Needs Evaluation: A Patient-Centered Approach. Plast Reconstr Surg. 2023.
9. The THumb Osteoarthritis Exercise TriAl (THETA); a Multicenter, Randomized Controlled Trial on Exercise Therapy With an Orthosis Compared to an Orthosis Alone in Patients With Thumb Base Osteoarthritis [Internet]. 2022. Available from: https://clinicaltrials.gov/study/NCT05772715.
10. Challoumas D, Murray E, Ng N, Putti A, Millar N. A Meta-analysis of Surgical Interventions for Base of Thumb Arthritis. J Wrist Surg. 2022;11(6):550-60.
11. Rajan PV, Qudsi RA, Dyer GSM, Losina E. Cost-utility studies in upper limb orthopaedic surgery: a systematic review of published literature. Bone Joint J. 2018;100-B(11):1416-23.
12. Wouters RM, Jobi-Odeneye AO, de la Torre A, Joseph A, Hand I, Wrist Working G, Hovius SER. A Standard Set for Outcome Measurement in Patients With Hand and Wrist Conditions: Consensus by the International Consortium for Health Outcomes Measurement Hand and Wrist Working Group. J Hand Surg Am. 2021;46(10):841-55 e7.
13. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann Surg. 2004;240(2):205-13.
14. Ring D, Wouters RM, Hoogendam L, van der Oest MJW, Vermeulen GM, Feitz R, et al. Rethinking Complications in Hand Surgery. J Hand Surg Am. 2023;48(10):e13-e4.
15. Hoogendam L, van der Oest MJW, Vermeulen GM, Feitz R, Hovius SER, Zuidam JM, et al. Prevalence of Complications and Association With Patient-Reported Outcomes After Trapeziectomy With a Weilby Sling: A Cohort Study. J Hand Surg Am. 2023;48(5):469-78.

16. Wouters RM, de Ridder WA, Slijper HP, Vermeulen GM, Hovius SER, Selles RW, et al. The Ultrashort Mental Health Screening Tool Is a Valid and Reliable Measure With Added Value to Support Decision-making. Clin Orthop Relat Res. 2023;482(1):59-70.

17. Tolk JJ, Janssen RPA, Haanstra TM, van der Steen MC, Bierma-Zeinstra SMA, Reijman M. The influence of expectation modification in knee arthroplasty on satisfaction of patients: a randomized controlled trial. Bone Joint J. 2021;103-B(4):619-26.

18. Akroyd A, Gunn KN, Rankin S, Douglas M, Kleinstauber M, Rief W, Petrie KJ. Optimizing patient expectations to improve therapeutic response to medical treatment: A randomized controlled trial of iron infusion therapy. Br J Health Psychol. 2020;25(3):639-51.

19. Mooiweer Y, Roling L, Vugrin M, Ansmann L, Stevens M, Seeber GH. Influence of patients' preoperative expectations on postoperative outcomes after total knee or hip arthroplasty: a systematic review. EFORT Open Rev. 2024;9(2):107-18.

20. Lucaroni F, Cicciarella Modica D, Macino M, Palombi L, Abbondanzieri A, Agosti G, et al. Can risk be predicted? An umbrella systematic review of current risk prediction models for cardiovascular diseases, diabetes and hypertension. BMJ Open. 2019;9(12):e030234.

21. Bretthauer M, Gerke S, Hassan C, Ahmad OF, Mori Y. The New European Medical Device Regulation: Balancing Innovation and Patient Safety. Ann Intern Med. 2023;176(6):844-8.

22. Jadresic MC, Baker JF. Predicting complications of spine surgery: external validation of three models. Spine J. 2022;22(11):1801-10.

23. Klemt C, Uzosike AC, Harvey MJ, Laurencin S, Habibi Y, Kwon YM. Neural network models accurately predict discharge disposition after revision total knee arthroplasty? Knee Surg Sports Traumatol Arthrosc. 2022;30(8):2591-9.

24. Kabore EG, Macdonald C, Kabore A, Didier R, Arveux P, Meda N, et al. Risk Prediction Models for Cardiotoxicity of Chemotherapy Among Patients With Breast Cancer: A Systematic Review. JAMA Netw Open. 2023;6(2):e230569.

25. Vuong K, McGeechan K, Armstrong BK, Cust AE. Risk prediction models for incident primary cutaneous melanoma: a systematic review. JAMA Dermatol. 2014;150(4):434-44.

26. Gupta P, Haeberle HS, Zimmer ZR, Levine WN, Williams RJ, Ramkumar PN. Artificial intelligence-based applications in shoulder surgery leaves much to be desired: a systematic review. JSES Rev Rep Tech. 2023;3(2):189-200.

27. Ong WL, Schouwenburg MG, van Bommel ACM, Stowell C, Allison KH, Benn KE, et al. A Standard Set of Value-Based Patient-Centered Outcomes for Breast Cancer: The International Consortium for Health Outcomes Measurement (ICHOM) Initiative. JAMA Oncol. 2017;3(5):677-85.

28. Liyanage H, Liaw ST, Jonnagaddala J, Hinton W, de Lusignan S. Common Data Models (CDMs) to Enhance International Big Data Analytics: A Diabetes Use Case to Compare Three CDMs. Stud Health Technol Inform. 2018;255:60-4.

29. Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J Am Med Inform Assoc. 2015;22(3):553-64.

30. Schuemie M, Reps J, Black A, Defalco F, Evans L, Fridgeirsson E, et al. Health-Analytics Data to Evidence Suite (HADES): Open-Source Software for Observational Research. Stud Health Technol Inform. 2024;310:966-70.

31. Reich C, Ostropolets A, Ryan P, Rijnbeek P, Schuemie M, Davydov A, et al. OHDSI Standardized Vocabularies-a large-scale centralized reference ontology for international data harmonization. J Am Med Inform Assoc. 2024;31(3):583-90.

32. Voss EA, Blacketer C, van Sandijk S, Moinat M, Kallfelz M, van Speybroeck M, et al. European Health Data & Evidence Network-learnings from building out a standardized international health data network. J Am Med Inform Assoc. 2023;31(1):209-19.

33. PREPARE-Transforming rehabilitation: Personalized care for a better quality of life [Available from: https://prepare-rehab.eu/.

34. Kappen TH, van Loon K, Kappen MA, van Wolfswinkel L, Vergouwe Y, van Klei WA, et al. Barriers and facilitators perceived by physicians when using prediction models in practice. J Clin Epidemiol. 2016;70:136-45.

35. Ho V, Brown Johnson C, Ghanzouri I, Amal S, Asch S, Ross E. Physician- and Patient-Elicited Barriers and Facilitators to Implementation of a Machine Learning-Based Screening Tool for Peripheral Arterial Disease: Preimplementation Study With Physician and Patient Stakeholders. JMIR Cardio. 2023;7:e44732.

36. Joshi M, Mecklai K, Rozenblum R, Samal L. Implementation approaches and barriers for rule-based and machine learning-based sepsis risk prediction tools: a qualitative study. JAMIA Open. 2022;5(2):ooac022.

37. Ploegmakers KJ, Medlock S, Linn AJ, Lin Y, Seppala LJ, Petrovic M, et al. Barriers and facilitators in using a Clinical Decision Support System for fall risk management for older people: a European survey. Eur Geriatr Med. 2022;13(2):395-405.

38. Fujimori R, Liu K, Soeno S, Naraba H, Ogura K, Hara K, et al. Acceptance, Barriers, and Facilitators to Implementing Artificial Intelligence-Based Decision Support Systems in Emergency Departments: Quantitative and Qualitative Evaluation. JMIR Form Res. 2022;6(6):e36501.

39. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform. 2021;113:103655.

40. Tonekaboni SJ, S.; McCradden, M.D.; Goldenberg, A.;. What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. Proceedings of Machine Learning Research. 2019:1-21.

41. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. Adv Neur In. 2017;30.

42. Ribeiro MTS, Sameer; Guestrin, Carlos;. "Why Should I Trust You?"

43. Explaining the Predictions of Any Classifier. DeNero JF, Mark; Reddy, Sravana;, editor. San Diego, California: Association for Computational Linguistics; 2016.

44. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ. 2009;338:b606.

45. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581-92.

46. Stirrat GM, Farrow SC, Farndon J, Dwyer N. The challenge of evaluating surgical procedures. Ann R Coll Surg Engl. 1992;74(2):80-4.

47. Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, et al. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ. 2024;384:e074819.

21

SUMMARY

SUMMARY

Within a patient-centered care setting, it is currently best practice that patients, together with their treating clinician, decide on the most suitable treatment option while considering their preferences and values. This decision-making process is commonly referred to as shared decision-making. This is particularly preferred when considering elective treatment, as in many hand or wrist conditions, where many conditions are not considered medical emergencies. While shared decision-making is regarded as best practice and is preferred by patients, it is not always implemented in daily clinical practice. Decisions are often based on clinician experience and established guidelines, which can result in a lack of personalization in decisions and (unwanted) variability between clinicians. To improve this, it is essential that patients receive relevant and unbiased information about the treatment options, risks, and benefits from their clinician.

Therefore, the main aim of this thesis is to provide patients with hand and wrist conditions and their clinicians with clinically relevant, personalized, and data-driven outcome information, to facilitate shared decision-making towards the most suitable treatment option. To this end, we used routine outcome information to describe clinically relevant treatment outcomes in general, to explain variation in treatment outcomes, and to predict treatment outcomes for individual patients with hand and wrist conditions. In the following sections, the main findings are summarized, following the structure of this thesis: I) Determining clinically important outcome values, II) Outcomes of treatment (strategies) for thumb base osteoarthritis, III) Prognostic factors for treatment outcomes, and IV) Development, validation, and added value of prediction models in hand surgery and hand therapy.

PART I - DETERMINING CLINICALLY IMPORTANT OUTCOME VALUES

To describe, explain, and predict outcomes, we first need to know what patients consider relevant outcomes after treatment for their hand or wrist disorder. This can be expressed in concepts such as the minimally important change (MIC) and the patient acceptable symptom state (PASS). Because clinically important outcome values for patients with hand and wrist disorders are limited in the literature, we determined these in Part I for several Patient Reported Outcome Measures frequently used to assess treatment outcomes for hand or wrist disorders.

Chapter 2 provides estimates of the minimal improvement patients consider a meaningful change (i.e., MIC) for the Michigan Hand outcomes Questionnaire, the Boston Carpal Tunnel Questionnaire, the Patient Rated Hand/Wrist Evaluation and the Visual Analog Scale. Notably, we found that patients being treated nonsurgically had lower MICs

(indicating being satisfied with a smaller improvement) compared to patients being treated surgically.

Chapter 3 presents a new anchor-based method using Item Response Theory to calculate Patient Acceptable Symptom State estimates (i.e., at what symptom level patients consider themselves as "good" after treatment). In a simulation study, we showed that this method provides more precise and reliable estimates of the Patient Acceptable Symptom State than previous, frequently used methods. Additionally, we applied this method to calculate the Patient Acceptable Symptom State for pain following trigger finger release.

In **Chapter 4**, we used the method from Chapter 3 to provide estimates of the Patient Acceptable Symptom State for three multi-item Patient Reported Outcome Measures for 35 diagnosis-treatment combinations for hand and wrist disorders. In conclusion, in Part I we have determined the Minimally Important Change and Patient Acceptable Symptom State values for many common hand and wrist disorders and how these values should be calculated.

PART II – CLINICALLY IMPORTANT OUTCOMES OF TREATMENT (STRATEGIES) FOR THUMB BASE OSTEOARTHRITIS

While outcomes of several treatment options for thumb base osteoarthritis are widely described, longer follow-up periods and consensus-based definitions of outcomes such as complications, are needed to improve pretreatment information provision for patients with thumb base osteoarthritis. Additionally, cost-effectiveness is often unknown, while such information is useful to guide treatment strategies from a societal perspective. Therefore, in Part II, we determined long-term outcomes, complications, and cost(effectiveness) of several treatment options for thumb base osteoarthritis, both nonsurgical and surgical.

Chapter 5 presents the long-term outcomes of nonsurgical treatment for thumb base osteoarthritis in our cohort (i.e., hand therapy, with or without an orthosis). Patients reported a sustained positive effect on pain from 1 year after treatment onwards, and only 22% underwent surgery at a median follow-up of 7 years.

In **Chapter 6**, we used the "International Consortium for Health Outcomes Measurement Complications in Hand and Wrist conditions" classification system to study complications following trapeziectomy with a Weilby sling for thumb base osteoarthritis. The complication rate was 35%, with 16% being Grade 1 (e.g., requiring additional hand therapy) and 19% being Grade 2/3 events (e.g., receiving a corticosteroid injection for Flexor Carpi Radialis tendinitis, or revision surgery). As only Grade 2/3 events were associated with worse

patient-reported outcomes twelve months postoperatively, we proposed to reclassify Grade 1 events as “adverse protocol deviations” and consider Grade 2 and 3 as “complications”.

Considering that multiple surgical techniques are being used within our cohort to treat thumb base osteoarthritis, we compared the outcomes of four techniques in **Chapter 7**. Specifically, we compared treatment outcomes of a) trapeziectomy with tendon interposition (Anchovy plasty), b) trapeziectomy with Flexor Carpi Radialis sling (Weilby plasty), c) trapeziectomy with Flexor Carpi Radialis sling and bone tunnel (Burton-Pellegrini plasty), and d) trapeziectomy with Abductor Pollicis Longus sling (Zancolli plasty). We found no differences in pain at twelve months, accounting for potential differences in patient characteristics and symptom severity. Regarding secondary outcomes, patients receiving a trapeziectomy with Anchovy plasty were more likely to experience a clinically relevant improvement in pain.

Given the many treatment strategies for thumb base osteoarthritis, with differing costs, we evaluated cost-effectiveness of four potential treatment strategies for thumb base osteoarthritis over a 10-year period in **Chapter 8**. Both treatment-related costs and societal costs were included. We found that starting with nonsurgical treatment, considering surgery if needed, resulted in 7.01 Quality-Adjusted Life-Years at a cost of 23.175 Euro per patient over a 10-year period. This treatment strategy had the highest probability of being cost-effective in our probabilistic sensitivity analysis.

In summary, Part II shows the importance of treating patients with thumb base osteoarthritis first nonsurgically. Surgery can be considered if nonsurgical treatment fails to relieve symptoms sufficiently, but surgical outcomes vary, and there is a considerable number of complications.

PART III - PROGNOSTIC FACTORS FOR CLINICALLY IMPORTANT TREATMENT OUTCOMES (FOR THUMB BASE OSTEOARTHRITIS)

Treatment outcomes for hand and wrist conditions can vary considerably from patient to patient, but it is poorly understood why certain patients have positive or negative treatment outcomes. Therefore, in Part III, we aimed to better understand which factors explain variation in treatment outcomes. A particular focus was on the relation between psychological factors (including psychological distress, pain catastrophizing, and illness perceptions) and expectations and treatment outcomes.

In **Chapter 9**, we studied to what extent psychological factors explain pain levels prior to nonsurgical treatment for thumb base osteoarthritis. We found that 41% of the variance in pre-treatment pain levels could be explained by psychological factors,

such as pain catastrophizing, and illness perceptions. Notably, only 1% could be explained by radiographical severity of the osteoarthritis (i.e., presence or absence of scaphotrapeziotrapezoid osteoarthritis), whereas previously, this is considered an important factor in making treatment decisions by many clinicians.

In **Chapter 10**, we assessed which pre-treatment factors were associated with outcomes of nonsurgical thumb base osteoarthritis treatment after three months. In addition to lower pre-treatment symptom severity and being a non-smoker, we found that positive outcome expectations and better illness understanding were associated with better treatment outcomes at three months following nonsurgical thumb base osteoarthritis treatment.

In **Chapter 11**, we assessed which factors were associated with satisfaction with treatment results three months after nonsurgical treatment for thumb base osteoarthritis. We found that outcome expectations were the only factor related to satisfaction with treatment results. In As we found an association between outcome expectations and pain and hand function three months post-treatment in Chapter 10, we also examined whether these outcomes mediate the relationship between outcome expectations and satisfaction with treatment results.

We found a partial mediation effect, indicating that there is also a direct association between outcome expectations and satisfaction with treatment outcomes.

In **Chapter 12**, we studied which factors are associated with satisfaction with treatment outcomes in a broader population of patients receiving one of six different (non)surgical treatments for varying hand and wrist conditions (e.g., carpal tunnel syndrome, Dupuytren's disease, and midcarpal laxity). Like Chapter 11, we found a positive association between pre-treatment expectations and satisfaction with treatment results.

Because we found pre-treatment expectations to be associated with treatment outcomes in Chapter 10, and with satisfaction with treatment outcomes in Chapters 11 and 12, we studied how variation in pre-treatment expectations is explained by patient characteristics, surgeon characteristics, and treatment characteristics in **Chapter 13**. We found that surgical treatment and more positive illness perceptions were associated with more positive expectations, indicating that expectation management should be tailored to the specific treatment modality and the specific patient (including the way they perceive their illness).

Personal injury claim involvement is generally considered as a risk factor for poor outcomes in (orthopedic) surgery. Therefore, in **Chapter 14,** we examined whether personal injury claim involvement was negatively associated with pain, hand function, and time to return to work in patients treated for hand or wrist conditions. We found that surgically treated patients involved in a personal injury claim experienced more pain, reduced hand function, and return to work time than matched patients without a claim. For nonsurgically treated

personal injury claim patients, similar negative associations were found for pain and return to work time.

In brief, Part III demonstrates the association between psychological factors and treatment outcomes for hand and wrist conditions. More positive pre-treatment expectations are (independently) related to better treatment outcomes, suggesting that improved expectation management by clinicians may positively affect treatment outcomes.

PART IV - DEVELOPMENT, VALIDATION, AND ADDED VALUE OF PREDICTION MODELS IN HAND SURGERY

Prediction models are promising tools to inform individual patients on their likely treatment outcomes, but they are scarcely available for patients with hand and wrist conditions. Therefore, in Part IV of this thesis, we aimed to develop and validate prediction models to inform patients on their individual probability of improvement after treatment for hand or wrist conditions. Additionally, we studied whether these prediction models are of added value for clinical decision-making.

In **Chapter 15**, we developed and internally validated a prediction model for the probability of a clinically relevant improvement in symptoms following carpal tunnel release surgery. The best performing model using 5 predictors, had a discriminative ability of 0.72 and good calibration in the holdout test dataset. We implemented this prediction model (CTR model) into a web application to facilitate its use.

Using the same methodology, we tried to develop and validate prediction models for improvement in pain and hand function after surgical treatment for thumb base osteoarthritis in **Chapter 16**. While the model predicting hand function had a good discriminative ability (AUC 0.74) and sufficient calibration, the model predicting pain was deemed insufficient for use due to a poor discriminative ability (AUC 0.59) and poor calibration. Like the CTR model, we presented the prediction for hand function using a web application.

In **Chapter 17**, we evaluated the performance of the Carpal Tunnel Release (CTR) model (presented in Chapter 15) over time. Additionally, we assessed the potential added value of the CTR model in clinical decision-making. The model performance in temporal validation was still sufficient and quite similar to the performance in the internal validation (AUC 0.78 and good calibration). The net benefit analysis showed better decision-making based on the prediction model compared with the surgeons' decision-making.

Considering that prediction models are increasingly being developed for outcomes of many surgical treatments, in **Chapter 18,** we provided an overview of published prediction models for outcomes of upper extremity surgery, the availability of these models for clinicians, and their risk of bias. In the literature search up to April 21st, 2022, we found

22 relevant publications about 28 prediction models, with most being applicable to shoulder arthroplasty outcomes. There is a need for external validation of existing models and development of models for outcomes of hand, wrist, and elbow surgery.

We concluded this thesis by investigating whether a single, generic prediction model can be used to predict outcomes for multiple diagnosis-treatment combinations in **Chapter 19**. We found that the generic models were equally suitable for clinical application as diagnosis-treatment-specific models. Given the developmental and implementation advantages, generic models seem preferable and effective in facilitating shared decision-making.

In conclusion, in Part IV, we have shown that we can develop well-performing prediction models for treatment outcomes of hand and wrist conditions, that these have potential to improve decision-making, and that a more general approach with one prediction model for multiple diagnosis-treatment combinations results in sufficient model performance and is feasible to implement.

Chapter 20 discusses these findings, their implications, and the limitations of these studies in more detail.

22

NEDERLANDSE SAMENVATTING

SAMENVATTING

Binnen een patiëntgerichte zorgsetting is het momenteel best practice dat patiënten samen met hun behandelend arts beslissen over de meest geschikte behandeloptie, waarbij rekening gehouden wordt met hun wensen en wat zij belangrijk vinden. Dit besluitvormingsproces wordt ook wel shared decision-making genoemd. Dit is vooral van meerwaarde bij besluitvorming rondom electieve (niet-urgente) behandelingen, zoals bij behandelingen voor veel hand- of polsaandoeningen. Hoewel shared decision-making als best practice wordt beschouwd en patiënten dit prettig vinden, wordt het niet altijd toegepast in de dagelijkse klinische praktijk. Beslissingen zijn vaak gebaseerd op de ervaring van zorgverleners en geldende richtlijnen, wat kan leiden tot een weinig persoonlijke benadering van de patiënt, ongewenste variabiliteit in beslissingen tussen zorgverleners en in hoe zorgverleners hun patiënten informeren. Om dit te verbeteren is het essentieel dat patiënten van hun arts relevante en objectieve informatie krijgen over de behandelopties, met de bijbehorende risico's en voordelen.

Daarom is het hoofddoel van dit proefschrift om patiënten met hand- en polsaandoeningen en hun zorgverleners te voorzien van klinisch relevante, gepersonaliseerde en datagedreven uitkomstinformatie om shared decision-making omtrent de meest geschikte behandeloptie te vergemakkelijken. Hiervoor gebruikten we routinematige uitkomstinformatie om klinisch relevante behandeluitkomsten te beschrijven, om variatie in behandeluitkomsten te verklaren en om behandeluitkomsten voor individuele patiënten met hand- en polsaandoeningen te voorspellen. In de volgende secties worden de belangrijkste bevindingen samengevat, volgens de structuur van dit proefschrift: I) Bepalen van klinisch relevante uitkomstwaarden, II) Uitkomsten van behandeling(strategieën) voor duimbasis artrose, III) Prognostische factoren voor behandeluitkomsten, en IV) Ontwikkeling, validatie en meerwaarde van predictiemodellen voor handchirurgie en handtherapie.

DEEL I - KLINISCH RELEVANTE UITKOMSTWAARDEN BEPALEN

Om uitkomsten te kunnen beschrijven, verklaren en voorspellen, moeten we eerst weten wat patiënten als relevante behandeluitkomsten beschouwen . Dit kan worden uitgedrukt in concepten zoals de Minimally Important Change (MIC) en de Patient Acceptable Symptom State (PASS). Omdat klinisch relevante uitkomstwaarden voor patiënten met hand- en polsaandoeningen beperkt gerapporteerd zijn in de literatuur, hebben we deze in deel I bepaald voor verschillende veel voorkomende patient-gerapporteerde uitkomstmaten (PROMs). Daarnaast hebben we gekeken of de klinisch relevante uitkomstwaarden verschillen tussen hand- en polsaandoeningen.

Hoofdstuk 2 geeft schattingen van de minimale verbetering die patiënten beschouwen als een relevante verbetering (d.w.z. de MIC) voor de Michigan Hand outcomes Questionnaire, de Boston Carpal Tunnel Questionnaire, de Patient Rated Hand/Wrist Evaluation en de Visual Analog Scale. We vonden dat patiënten die conservatief werden behandeld, lagere MIC's hadden (wat aangeeft dat ze tevreden zijn met een kleinere verbetering) vergeleken met patiënten die chirurgisch werden behandeld.

Hoofdstuk 3 introduceert een nieuwe anker-gebaseerde methode die gebruik maakt van Item Response Theory om de PASS te berekenen (d.w.z. wanneer patiënten zichzelf als "goed" beschouwen na behandeling). In een simulatiestudie toonden we aan dat deze methode preciezere en betrouwbaardere schattingen van de PASS geeft dan vorige, vaak gebruikte methoden. Ook pasten we deze methode toe om de PASS te berekenen voor pijn na een trigger finger release.

In **Hoofdstuk 4** gebruikten we de methode uit Hoofdstuk 3 om schattingen te geven van de PASS voor drie multi-item PROMs (de Michigan Hand outcomes Questionnaire, de Boston Carpal Tunnel Questionnaire en de Patient Rated Hand/Wrist Evaluation) voor 35 diagnose-behandelcombinaties voor hand- en polsaandoeningen.

Concluderend hebben we in Deel I nieuwe methoden ontwikkeld om de PASS waarden te berekenen en hebben voor de meest voorkomende hand en pols aandoeningen de MIC en PASS waarden bepaald.

DEEL II - KLINISCH RELEVANTE UITKOMSTEN VAN BEHANDELING(STRATEGIEËN) VOOR DUIMBASISARTROSE

De uitkomsten van verschillende behandelopties voor duimbasisartrose zijn in de literatuur vrij uitgebreid beschreven. Er zijn echter langere follow-up periodes nodig en op consensus gebaseerde uitkomstdefinities, bijvoorbeeld voor complicaties, om de informatievoorziening voorafgaand aan de behandeling voor patiënten met duimbasisartrose te verbeteren. Ook is het vaak niet bekend in hoeverre behandelingen en behandelingstrategieën kosteneffectief zijn, terwijl dergelijke informatie erg nuttig is ter bepaling van behandelrichtlijnen. Daarom hebben we in deel II de lange termijn resultaten, complicaties en kosteneffectiviteit van verschillende (niet-) chirurgische behandelingsopties voor duimbasisartrose bepaald.

In **Hoofdstuk 5** staan de lange termijn uitkomsten van niet-chirurgische behandeling (d.w.z. handtherapie, met of zonder orthese) van duimbasisartrose in ons cohort beschreven. Patiënten rapporteerden een blijvend positief effect op pijn vanaf 1 jaar na de behandeling en slechts 22% onderging een operatie bij een mediane follow-up van 7 jaar.

In **Hoofdstuk 6** hebben we het classificatiesysteem "International Consortium for Health Outcomes Measurement Complications in Hand and Wrist conditions" gebruikt om

complicaties na trapeziectomie met een Weilby sling voor duimbasisartrose te bestuderen. Het complicatiepercentage was 35%, waarvan 16% graad 1 was (bijv. extra handtherapie nodig) en 19% graad 2/3 (bijv. een corticosteroïdinjectie voor Flexor Carpi Radialis tendinitis of een revisie operatie). Aangezien alleen graad 2/3 complicaties geassocieerd waren met slechtere patiënt gerapporteerde uitkomsten twaalf maanden postoperatief, stelden we voor om graad 1 te herclassificeren als "adverse protocol deviations" en graad 2 en 3 te beschouwen als "complicaties".

Aangezien er binnen ons cohort meerdere chirurgische technieken worden gebruikt om duimbasisartrose te behandelen, hebben we in **Hoofdstuk 7** de uitkomsten van vier technieken met elkaar vergeleken. Specifiek vergeleken we de behandelresultaten van a) trapeziectomie met peesinterpositie (Anchovy plastiek), b) trapeziectomie met Flexor Carpi Radialis sling (Weilby plastiek), c) trapeziectomie met Flexor Carpi Radialis sling en bottunnel (Burton-Pellegrini plastiek), en d) trapeziectomie met Abductor Pollicis Longus sling (Zancolli plastiek). We vonden geen verschillen in pijn tussen de chirurgische technieken na twaalf maanden, rekening houdend met verschillen in patiëntkenmerken en ernst van de symptomen. Op secundaire uitkomsten hadden patiënten die een trapeziectomie met Anchovy plastiek een iets grotere kans om een klinisch relevante verbetering in pijn te ervaren.

Gezien de vele behandelstrategieën voor duimbasisartrose, met verschillende kosten, hebben we in **Hoofdstuk 8** de kosteneffectiviteit van vier potentiële behandelstrategieën voor duimbasisartrose over een periode van 10 jaar geëvalueerd. Zowel behandeling gerelateerde kosten als maatschappelijke kosten werden meegenomen. We ontdekten dat beginnen met niet-chirurgische behandeling en indien nodig chirurgie overwegen, resulteerde in 7,01 kwaliteit gecorrigeerde levensjaren tegen kosten van 23,175 euro per patiënt over een periode van 10 jaar. Deze behandelingsstrategie had de hoogste waarschijnlijkheid om kosteneffectief te zijn in onze sensitiviteitsanalyse.

Samengevat laat deel II zien dat het belangrijk het is om patiënten met duimbasisartrose eerst niet-chirurgisch te behandelen. Chirurgie kan overwogen worden als niet-chirurgische behandeling de symptomen onvoldoende verlicht, maar chirurgische resultaten variëren en er is een aanzienlijk aantal complicaties.

DEEL III - PROGNOSTISCHE FACTOREN VOOR KLINISCH RELEVANTE BEHANDELUITKOMSTEN (VOOR DUIMBASIS-ARTROSE)

Behandeluitkomsten voor hand- en polsaandoeningen kunnen van patiënt tot patiënt aanzienlijk verschillen, maar het is vaak onduidelijk waarom bepaalde patiënten goede of slechte behandeluitkomsten hebben. Daarom probeerden we in deel III beter te begrijpen

welke factoren variatie in behandeluitkomsten verklaren. We richtten ons met name op de relatie tussen psychologische factoren (waaronder angst en depressie, pijn catastroferen en ziektepercepties), verwachtingen en behandeluitkomsten.

In **Hoofdstuk 9** onderzochten we in hoeverre psychologische factoren het pijnniveau voorafgaand aan een niet-chirurgische behandeling voor duimbasisartrose verklaren. We vonden dat 41% van de variatie in pijnniveaus voorafgaand aan de behandeling verklaard kon worden door psychologische factoren, zoals pijn catastroferen en ziektepercepties. Opvallend was dat slechts 1% verklaard kon worden door de radiografische ernst van de artrose (d.w.z. aan- of afwezigheid van scaphotrapeziotrapezoïde artrose), terwijl dit door veel zorgverleners als een belangrijke factor wordt beschouwd in de besluitvorming rondom behandelingen van duimbasisartrose.

In **Hoofdstuk 10** hebben we onderzocht welke factoren aanwezig vóór de behandeling geassocieerd waren met uitkomsten van niet-chirurgische duimbasis artrose behandeling op drie maanden. Naast een lagere ernst van de symptomen vóór de behandeling en niet roken, vonden we dat positieve verwachtingen en een beter ziekte-inzicht geassocieerd waren met betere behandeluitkomsten op drie maanden na niet-chirurgische duimbasis artrose behandeling.

In **Hoofdstuk 11** onderzochten we welke factoren geassocieerd waren met tevredenheid met de behandeluitkomsten drie maanden na niet-chirurgische behandeling voor duimbasis artrose. We vonden dat verwachtingen de enige factor was die gerelateerd was aan tevredenheid met de behandeluitkomsten. Omdat we in hoofdstuk 10 een associatie vonden tussen verwachtingen en pijn en handfunctie drie maanden na de behandeling, hebben we ook onderzocht of pijn en handfunctie de relatie tussen verwachtingen en tevredenheid met de behandeluitkomsten mediëren. We vonden een partieel mediatie-effect, wat aangeeft dat er ook een directe associatie is tussen verwachtingen en tevredenheid met de behandeluitkomsten.

In **Hoofdstuk 12** hebben we in een bredere populatie onderzocht welke factoren geassocieerd zijn met tevredenheid met de behandeluitkomsten. We hebben patiënten geïncludeerd die zes verschillende (niet-)chirurgische behandelingen ondergingen voor hun hand- of polsaandoening (incl. carpaal tunnel syndroom, de ziekte van Dupuytren en midcarpale laxiteit). Net als in Hoofdstuk 11 vonden we een positieve associatie tussen verwachtingen en tevredenheid met de behandeluitkomsten.

Omdat we in hoofdstuk 10 vonden dat verwachtingen geassocieerd waren met de behandeluitkomsten en in hoofdstuk 11 en 12 met tevredenheid met de behandeluitkomsten, hebben we in **Hoofdstuk 13** onderzocht hoe variatie in verwachtingen vooraf wordt verklaard door patiëntkenmerken, chirurgkenmerken en behandelkenmerken. We vonden dat chirurgische behandeling en positievere ziektepercepties geassocieerd waren met

positievere verwachtingen, wat aangeeft dat verwachtingsmanagement moet worden afgestemd op de specifieke behandelingsmodaliteit en de specifieke patiënt (inclusief de manier waarop zij tegen hun ziekte aankijken).

Betrokkenheid bij letselschadeclaims wordt over het algemeen beschouwd als een risicofactor voor slechte uitkomsten bij (orthopedische) chirurgie. Daarom onderzochten we in **Hoofdstuk 14** of betrokkenheid bij een letselschadeclaim negatief geassocieerd was met pijn, handfunctie en tijd tot terugkeer naar werk bij patiënten die behandeld werden voor hand- of pols-aandoeningen. We vonden dat chirurgisch behandelde patiënten die betrokken waren bij een letselschadeclaim meer pijn, verminderde handfunctie en werkhervatting ervoeren dan vergelijkbare patiënten zonder claim. Voor niet-chirurgisch behandelde letselschadepatiënten werden vergelijkbare negatieve associaties gevonden voor pijn en werkhervatting.

Kort samengevat toont deel III de associatie aan tussen psychologische factoren en behandeluitkomsten voor hand- en polsaandoeningen. Positievere verwachtingen voorafgaand aan de behandeling zijn gerelateerd aan betere behandeluitkomsten, wat suggereert dat verbeteringen in verwachtingsmanagement door zorgverleners de behandeluitkomsten positief zou kunnen beïnvloeden.

DEEL IV - ONTWIKKELING, VALIDATIE EN MEERWAARDE VAN PREDICTIEMODELLEN VOOR HANDCHIRURGIE

Predictiemodellen zijn veelbelovende hulpmiddelen om individuele patiënten te informeren over hun waarschijnlijke behandeluitkomsten, maar ze zijn nauwelijks beschikbaar voor patiënten met hand- en polsaandoeningen. Daarom richtten we ons in deel IV van dit proefschrift op het ontwikkelen en valideren van predictiemodellen om patiënten te informeren over hun individuele kans op verbetering na een behandeling voor hun hand- of pols-aandoening. Daarnaast onderzochten we of deze voorspellingsmodellen van meerwaarde zijn voor klinische besluitvorming.

In **Hoofdstuk 15** ontwikkelden en valideerden we een predictiemodel voor de kans op een klinisch relevante verbetering van symptomen na een carpaal tunnel release operatie. Het best presterende model, gebaseerd op 5 predictoren, had een discriminerend vermogen van 0,72 en een goede kalibratie in de hold-out test dataset. We hebben dit predictiemodel beschikbaar gesteld in een webapplicatie om het gebruik ervan te vergemakkelijken.

Met dezelfde methodologie als in hoofdstuk 15 hebben we in **Hoofdstuk 16** geprobeerd om predictiemodellen te ontwikkelen en te valideren voor verbetering in pijn en handfunctie na chirurgische behandeling van duimbasisartrose. Hoewel het model dat handfunctie voorspelde een goed discriminerend vermogen (AUC 0,74) en voldoende kalibratie had,

werd het model dat pijn voorspelde als onvoldoende voor gebruik beschouwd vanwege een slecht discriminerend vermogen (AUC 0,59) en slechte kalibratie. Net als bij het carpaal tunnel release model hebben we het model voor handfunctie in een webapplicatie beschikbaar gesteld.

In **Hoofdstuk 17** hebben we de prestaties van het carpaal tunnel release model over de tijd geëvalueerd (gepresenteerd in hoofdstuk 15). Daarnaast hebben we de potentiële toegevoegde waarde van het CTR-model bij klinische besluitvorming onderzocht. De prestaties van het model in de validatie in de tijd waren nog steeds voldoende en vergelijkbaar met de prestaties in de interne validatie (AUC 0,78 en goede kalibratie). De net benefit analyse liet een betere besluitvorming zien op basis van het voorspellingsmodel vergeleken met de huidige besluitvorming van handchirurgen.

Aangezien er steeds meer voorspellingsmodellen worden ontwikkeld voor uitkomsten van veel chirurgische behandelingen, hebben we in **Hoofdstuk 18** een overzicht gegeven van gepubliceerde voorspellingsmodellen voor uitkomsten van chirurgie aan de bovenste extremiteit, de beschikbaarheid van deze modellen voor zorgverleners en hun risico op bias. In de literatuursearch tot 21 april 2022 vonden we 22 artikelen, waarin 26 relevante voorspellingsmodellen werden beschreven. De meeste hiervan waren toepassing op uitkomsten van schouderartroplastiek. We adviseren externe validatie van bestaande voorspellingsmodellen en om nieuwe modellen te ontwikkelen voor de uitkomsten van hand, pols en elleboog chirurgie.

Tenslotte hebben we in **Hoofdstuk 19** onderzocht of een enkel, generiek voorspellingsmodel kan worden gebruikt om uitkomsten te voorspellen voor meerdere diagnose-behandelcombinaties. We vonden dat generieke voorspellingsmodellen net zo goed klinisch toepasbaar waren als diagnose-behandeling specifieke modellen. Vanwege hun voordelen qua ontwikkeling en implementatie, zijn generieke modellen veelbelovend en kunnen ze effectief shared decision-making faciliteren.

Concluderend hebben we in deel 4 laten zien dat we goed presterende voorspellingsmodellen kunnen ontwikkelen voor behandeluitkomsten van hand- en polsaandoeningen, dat deze potentieel hebben om de besluitvorming te verbeteren, en dat een meer algemene benadering met één voorspellingsmodel voor meerdere diagnose-behandelcombinaties ook resulteert in goede modelprestaties en haalbaar is om te implementeren.

In **Hoofdstuk 20** worden deze bevindingen, hun implicaties en de beperkingen van deze studies in meer detail besproken.

A

APPENDICES

LIST OF PUBLICATIONS

Publications in this thesis

Hoogendam, L., van der Oest, M., Tsehaie, J., Wouters, R. M., Vermeulen, G. M., Slijper, H. P., Selles, R. W., Porsius, J. T., the Hand-Wrist Study Group (2021). Psychological factors are more strongly associated with pain than radiographic severity in non-invasively treated first carpometacarpal osteoarthritis. Disability and rehabilitation, 43(13), 1897–1902. https://doi.org/10.1080/09638288.2019.1685602

Hoogendam, L., van der Oest, M., Wouters, R. M., Andrinopoulou, E. R., Vermeulen, G. M., Slijper, H. P., Porsius, J. T., Selles, R. W., The Hand-Wrist Study Group (2021). Patients With Higher Treatment Outcome Expectations Are More Satisfied With the Results of Nonoperative Treatment for Thumb Base Osteoarthritis: A Cohort Study. Archives of physical medicine and rehabilitation, 102(8), 1533–1540. https://doi.org/10.1016/j.apmr.2021.01.080

van der Oest, M. J. W., Hoogendam, L., Wouters, R. M., Vermeulen, G. M., Slijper, H. P., Selles, R. W., Vranceanu, A. M., Porsius, J. T., The Hand-Wrist Study Group (2022). Associations between positive treatment outcome expectations, illness understanding, and outcomes: a cohort study on non-operative treatment of first carpometacarpal osteoarthritis. Disability and rehabilitation, 44(19), 5487–5494. https://doi.org/10.1080/09638288.2021.1936661

Hoogendam, L., Koopman, J. E., van Kooij, Y. E., Feitz, R., Hundepool, C. A., Zhou, C., Slijper, H. P., Selles, R. W., Wouters, R. M., The Hand-Wrist Study Group (2022). What Are the Minimally Important Changes of Four Commonly Used Patient-reported Outcome Measures for 36 Hand and Wrist Condition-Treatment Combinations?. *Clinical orthopaedics and related research*, *480*(6), 1152–1166. https://doi.org/10.1097/CORR.0000000000002094

De Ridder, W. A., Wouters, R. M., Hoogendam, L., Vermeulen, G. M., Slijper, H. P., Selles, R. W., The Hand-Wrist Study Group (2022). Which Factors Are Associated With Satisfaction With Treatment Results in Patients With Hand and Wrist Conditions? A Large Cohort Analysis. *Clinical orthopaedics and related research*, *480*(7), 1287–1301. https://doi.org/10.1097/CORR.0000000000002107

Hoogendam, L., Bakx, J. A. C., Souer, J. S., Slijper, H. P., Andrinopoulou, E. R., Selles, R. W., The Hand-Wrist Study Group (2022). Predicting Clinically Relevant Patient-Reported Symptom Improvement After Carpal Tunnel Release: A Machine Learning Approach. *Neurosurgery*, *90*(1), 106–113. https://doi.org/10.1227/NEU.0000000000001749

Loos, N. L., Hoogendam, L., Souer, J. S., Slijper, H. P., Andrinopoulou, E. R., Coppieters, M. W., Selles, R. W., The Hand-Wrist Study Group (2022). Machine Learning Can be Used to Predict Function but Not Pain After Surgery for Thumb Carpometacarpal

Osteoarthritis. *Clinical orthopaedics and related research*, *480*(7), 1271–1284. https://doi.org/10.1097/CORR.0000000000002105

Hoogendam, L., Bink, T., de Lange, J., Selles, R. W., Colaris, J. W., Zuidam, J. M., Hovius, S. E. R., van der Heijden, B., The Hand-Wrist Study Group (2022). Which Tendon Plasty Has the Best Outcome? A Comparison of Four Tendon Plasty Techniques in a Large Cohort of Patients with Symptomatic Trapeziometacarpal Osteoarthritis. *Plastic and reconstructive surgery*, *150*(2), 364e–374e. https://doi.org/10.1097/PRS.0000000000009350

Hoogendam, L., van der Oest, M. J. W., Souer, J. S., Selles, R. W., Hovius, S. E. R., Feitz, R., The Hand-Wrist Study Group (2023). Involvement in a Personal Injury Claim Is Associated With More Pain and Delayed Return to Work After Elective Nonsurgical or Surgical Treatment for Hand or Wrist Disorders: A Propensity Score-matched Comparative Study. *Clinical orthopaedics and related research*, *481*(4), 751–762. https://doi.org/10.1097/CORR.0000000000002410

de Ridder, W. A., Hoogendam, L., Zeroual, F., Slijper, H. P., Wouters, R. M., Vermeulen, G. M., Selles, R. W., van der Oest, M. J. W., The Hand-Wrist Study Group (2023). Treatment Invasiveness and Illness Perceptions Are Strongly Associated With Outcome Expectations in Patients Treated for Hand or Wrist Conditions: A Cross-sectional Study. *Clinical orthopaedics and related research*, *481*(5), 994–1005. https://doi.org/10.1097/CORR.0000000000002540

Terluin, B., Koopman, J. E., Hoogendam, L., Griffiths, P., Terwee, C. B., Bjorner, J. B. (2023). Estimating meaningful thresholds for multi-item questionnaires using item response theory. *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation*, *32*(6), 1819–1830. https://doi.org/10.1007/s11136-023-03355-8

Hoogendam, L., van der Oest, M. J. W., Vermeulen, G. M., Feitz, R., Hovius, S. E. R., Zuidam, J. M., Slijper, H. P., Selles, R. W., The Hand-Wrist Study Group, Wouters, R. M. (2023). Prevalence of Complications and Association With Patient-Reported Outcomes After Trapeziectomy With a Weilby Sling: A Cohort Study. *The Journal of hand surgery*, *48*(5), 469–478. https://doi.org/10.1016/j.jhsa.2023.01.022

Esteban Lopez, L. M. J., Hoogendam, L., Vermeulen, G. M., Tsehaie, J., Slijper, H. P., Selles, R. W., Wouters, R. M., The Hand-Wrist Study Group (2023). Long-Term Outcomes of Nonsurgical Treatment of Thumb Carpometacarpal Osteoarthritis: A Cohort Study. *The Journal of bone and joint surgery. American volume*, *105*(23), 1837–1845. https://doi.org/10.2106/JBJS.22.01116

Loos, N. L., Hoogendam, L., Souer, J. S., van Uchelen, J. H., Slijper, H. P., Wouters, R. M., Selles, R. W., The Hand-Wrist Study Group (2024). Algorithm Versus Expert: Machine Learning Versus Surgeon-Predicted Symptom Improvement After Carpal Tunnel

Release. *Neurosurgery*, *95*(1), 110–117. Advance online publication. https://doi.org/10.1227/neu.0000000000002848

Hoogendam, L., van der Oest, M. J. W., Souer, J. S., Slijper, H. P., Lane, J. C. E., Furniss, D., Selles, R. W., The Hand-Wrist Study Group, Wouters, R. M. Direct and indirect costs of nonsurgical and surgical treatment strategies for first carpometacarpal osteoarthritis: A cost-effectiveness analysis using microsimulation. *Submitted to Osteoarthritis and Cartilage*

Hoogendam L., Wouters R. M., Bakx J. A. C., Souer, J. S., Slijper, H. P., Hand Wrist Study Group, Selles, R. W. Generic prediction models for improvement in patient-reported outcomes after treatment for multiple hand or wrist conditions: one model to rule them all? *Submitted to JAMA Surgery*

Hoogendam, L., Koopman, J. E., Selles, R. W., Hundepool, C. A., Zuidam, J. M., Slijper, H. P., Wouters, R. M., The Hand-Wrist Study Group, Terluin, B. What are the Patient Acceptable Symptom State values of commonly used multi-item Patient-Reported Outcome Measures for 35 hand and wrist condition-treatment combinations? *In preparation*

Loos, N. L., Koopman, J. E., Slijper, H. P., Selles, R.W., Wouters, R.M., Hoogendam, L. Prediction models for postoperative outcomes in patients undergoing surgical treatment for upper extremity disorders: A systematic review and critical appraisal. *In preparation*

Other publications

Cohen, A., Hoogendam, L., Reijman, M., Selles, R. W., Hovius, S. E. R., Colaris, J. W. (2021). Patient-reported physical functioning and pain improve after scaphoid nonunion surgery: A Cohort Study. *Injury*, *52*(10), 2952–2958. https://doi.org/10.1016/j.injury.2021.06.016

Mendelaar, N. H. A., Hundepool, C. A., Hoogendam, L., Duraku, L. S., Power, D. M., Walbeehm, E. T., Selles, R. W., Zuidam, J. M., The Hand-Wrist Study Group (2022). Outcome of Simple Decompression of Primary Cubital Tunnel Syndrome Based on Patient-Reported Outcome Measurements. The Journal of hand surgery, 47(3), 247–256.e1. https://doi.org/10.1016/j.jhsa.2021.11.020

Duraku, L. S., Hoogendam, L., Hundepool, C. A., Power, D. M., Rajaratnam, V., Slijper, H. P., Feitz, R., Zuidam, J. M., Selles, R. W., The Hand-Wrist Study Group & Esser Research Group (2022). Collaborative hand surgery clinical research without sharing individual patient data; proof of principle study. *Journal of plastic, reconstructive & aesthetic surgery: JPRAS*, *75*(7), 2242–2250. https://doi.org/10.1016/j.bjps.2022.02.065

Duraku, L. S., Hundepool, C. A., Hoogendam, L., Selles, R. W., van der Heijden, B. E. P. A., Colaris, J. W., Hovius, S. E. R., Zuidam, J. M., The Hand-Wrist Study Group (2022). Two-Corner Fusion or Four-Corner Fusion of the Wrist for Midcarpal Osteoarthritis?

A Multicenter Prospective Comparative Cohort Study. *Plastic and reconstructive surgery*, *149*(6), 1130e–1139e. https://doi.org/10.1097/PRS.0000000000009116

Loos, N. L., Hoogendam, L., Souer, J. S., Slijper, H. P., Andrinopoulou, E. R., Coppieters, M. W., Selles, R. W. (2022). Reply to the Letter to the Editor: Machine Learning Can be Used to Predict Function but Not Pain After Surgery for Thumb Carpometacarpal Osteoarthritis. *Clinical orthopaedics and related research*, *480*(11), 2278–2279. https://doi.org/10.1097/CORR.0000000000002398

Greminger, M., Koopman, J. E., van Kooij, Y. E., Hoogendam, L., Smit, J., Slijper, H. P., Selles, R. W., Calcagni, M., SUpExOr Study Group, The Hand-Wrist Study Group (2023). Factors associated with self-reported pain and hand function following dorsal wrist ganglion excision. *The Journal of hand surgery, European volume*, *48*(6), 551–560. https://doi.org/10.1177/17531934231153029

Mendelaar, N. H. A., Hundepool, C. A., Hoogendam, L., Duraku, L. S., Zöphel, O. T., Selles, R. W., Zuidam, J. M., The Hand-Wrist Study Group (2023). Multiple Compression Syndromes of the Same Upper Extremity: Prevalence, Risk Factors, and Treatment Outcomes of Concomitant Treatment. *The Journal of hand surgery*, *48*(5), 479–488. https://doi.org/10.1016/j.jhsa.2023.01.024

Ring, D., Wouters, R. M., Hoogendam, L., van der Oest, M. J. W., Vermeulen, G. M., Feitz, R., Hovius, S. E. R., Zuidam, J. M., Slijper, H. P., Selles, R. W. (2023). Rethinking Complications in Hand Surgery. *The Journal of hand surgery*, *48*(10), e13–e14. https://doi.org/10.1016/j.jhsa.2023.07.009

van Es, E. M., Dijkhof, M., Souer, J. S., van Ewijk, F. J., Hoogendam, L., Slijper, H. P., Selles, R. W., The Hand-Wrist Study Group (2024). Forearm rotation improves after corrective osteotomy in patients with symptomatic distal radius malunion. *Heliyon*, *10*(9), e29570. https://doi.org/10.1016/j.heliyon.2024.e29570

Notermans, B. J. W., Teunissen, J. S., Hoogendam, L., de Boer, L. H. L., Selles, R. W., van der Heijden, B. E. P. A., The Hand-Wrist Study Group (2024). "Proximal interphalangeal joint arthroplasty for osteoarthritis: Long-term follow-up of patient satisfaction, Michigan Hand outcomes Questionnaire scores, and reoperations.". *Plastic and reconstructive surgery*, 10.1097/PRS.0000000000011599. Advance online publication. https://doi.org/10.1097/PRS.0000000000011599

van Kooij, Y. E., Ter Stege, M. H. P., de Ridder, W. A., Hoogendam, L., Hovius, S. E. R., MacDermid, J. C., Selles, R. W., The Hand-Wrist Study Group, Wouters, R. M. (2024). The Validity and Responsiveness of the Patient-Specific Functional Scale in Patients With First Carpometacarpal Osteoarthritis. *The Journal of hand surgery*, *49*(9), 817–826. https://doi.org/10.1016/j.jhsa.2024.04.006

Loos, N. L., Selles, R. W., ter Stege, M. H. P., Arends, G. R., Hoogendam, L., van Kooij, Y. E., Veltkamp, J., The Outcome-Based Healthcare Research Group, Wouters, R. M. Using outcome information during consultation yields better shared decision-making, better patient experiences, and more positive expectations: a comparative effectiveness study. *In revision at Value in Health*

Loos, N. L., de Ridder, W. A., Hoogendam, L., MacDermid, J. C., Hoogeboom, T. J., Furniss, D., Wouters, R. M., The Hand-Wrist Study Group. The Personal Meaningful Gain is a Better Measure of Treatment Success than the MIC and PASS: a cohort study. *Submitted to the Journal of Bone and Joint Surgery (American volume)*

de Roode, A., de Ridder W. A., Hoogendam, L., Slijper, H. P., Hovius, S. E. R., Zuidam, J. M., Selles, R. W., The Hand-Wrist Study Group, Wouters, R. M. Which factors are independently associated with fulfilling information needs in patients treated for hand or wrist conditions? A prospective cohort study. *Submitted to Archives of Physical Medicine and Rehabilitation*

AUTHORS AND AFFILIATIONS

Contributing authors to the publications in this thesis

Listed in alphabetical order
Affiliations at the time the research was conducted

Eleni-Rosalina Andrinopoulou
Department of Biostatistics
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Jeanne A. C. Bakx
Department of Rehabilitation Medicine and Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Thijs Bink
Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Jakob B. Bjorner
Department of Public Health
University of Copenhagen & QualityMetric,
Copenhagen, Denmark & Johnston, Rhode Island, United States

Joost W. Colaris
Department of Orthopaedic Surgery
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Michiel W. Coppieters
Amsterdam Movement Sciences,
Faculty of Behavioural and Movement Sciences,
Vrije Universiteit Amsterdam,
Amsterdam, the Netherlands

Lisa M. J. Esteban Lopez
Department of Rehabilitation Medicine and Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center & Xpert Clinics Hand Therapy
Rotterdam & Utrecht, The Netherlands

Reinier Feitz
Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center & Xpert Clinics Hand and Wrist Care,
Rotterdam & Eindhoven, The Netherlands

Dominic Furniss
Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford,
Oxford, United Kingdom

Pip Griffiths
Patient Centered Endpoints,
IQVIA,
Reading, United Kingdom

E. P. A. (Brigitte) van der Heijden
Department of Plastic, Reconstructive, and Hand Surgery
Radboud University Medical Center,
Nijmegen, The Netherlands

Steven E. R. Hovius
Department of Plastic, Reconstructive, and Hand Surgery
Radboud University Medical Center & Xpert Clinics Hand and Wrist Care,
Nijmegen & Eindhoven, The Netherlands

Caroline A. Hundepool
Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Yara E. van Kooij
Department of Rehabilitation Medicine and Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center & Xpert Clinics Hand Therapy

Rotterdam & Utrecht, The Netherlands

Jaimy E. Koopman
Department of Plastic, Reconstructive, and Hand Surgery and Department of Rehabilitation Medicine
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Jennifer C. E. Lane
Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford,
Oxford, United Kingdom

Jurriaan de Lange
Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Nina L. Loos
Department of Rehabilitation Medicine and Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Jarry T. Porsius
Department of Rehabilitation Medicine and Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Mark J. W. van der Oest
Department of Plastic, Reconstructive, and Hand Surgery and Department of Rehabilitation Medicine
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Willemijn A. de Ridder
Department of Rehabilitation Medicine and Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center

Rotterdam, The Netherlands

Ruud W. Selles
Department of Rehabilitation Medicine and Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Harm P. Slijper
Equipe Zorgbedrijven,
Eindhoven, The Netherlands

J. Sebastiaan Souer
Xpert Clinics Hand and Wrist Care,
Eindhoven, The Netherlands

Berend Terluin
Department of General Practics & Amsterdam Public Health Research Institute
Amsterdam UMC, location VUmc,
Amsterdam, The Netherlands

Caroline B. Terwee
Department of Epidemiology and Data Science & Amsterdam Public Health Research Institute,
Amsterdam UMC, location VUmc,
Amsterdam, The Netherlands

Jonathan Tsehaie
Department of Plastic, Reconstructive, and Hand Surgery and Department of Rehabilitation Medicine
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Jeroen H. van Uchelen
Xpert Clinics Hand and Wrist Care,
Eindhoven, The Netherlands

Guus M. Vermeulen
Xpert Clinics Hand and Wrist Care,
Eindhoven, The Netherlands

Ana-Maria Vranceanu
Department of Psychiatry,
Massachusetts General Hospital,
Boston, Massachusetts, United States

Robbert M. Wouters
Department of Rehabilitation Medicine and Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center
Rotterdam, The Netherlands

Fadoua Zeroual
Department of Rehabilitation Medicine, Department of Plastic, Reconstructive, and Hand Surgery, and Department of Intensive Care,
Erasmus MC University Medical Center,
Rotterdam, the Netherlands

Chao Zhou
Department of Plastic, Reconstructive and Hand Surgery,
Maastricht University Hospital,
Maastricht, the Netherlands

J. Michiel Zuidam
Department of Plastic, Reconstructive, and Hand Surgery
Erasmus MC University Medical Center
Rotterdam, The Netherlands

PHD PORTFOLIO

PhD portfolio

Summary of PhD training and teaching

Name PhD student: Lisa Hoogendam	PhD period: Sep 2019 – Aug 2024
Erasmus MC Departments: Plastic, Reconstructive, and Hand Surgery & Rehabilitation Medicine	Promotor: Prof. Dr. Ruud. W. Selles Supervisors: Dr. Robbert M. Wouters, Dr. J. Sebastiaan Souer
Research School: NIHES	

1. **PhD training**

	Year	Workload (ECTS)
Courses		
• Systematisch literatuuronderzoek in PubMed	2019	0.4
• EndNote	2019	0.2
• Scientific integrity	2019	0.3
• Microsurgery training	2019	4.0
• CAIRElab summer school	2021	1.0
• Personal leadership and communication	2021	1.0
• Young ESSER masterclass	2021	0.3
• Basis Kwalificatie Onderwijs	2022	5.2
• Basis Kwalificatie Examinering	2022	1.8
• Datacamp “Big Data with R” track	2022	1.0
• Speech Republic – Impactvolle communicatie	2022	0.5
• Oxford summer school – Real world evidence using the Observational Medical Outcomes Partnership (OMOP)	2023	1.0
Participation in seminars and workshops		
• Participant in OMOP-Common Data Model Hackathon (organized by the BSSH)	2021	0.3
• Observational Health Data Sciences and Informatics (OHDSI) Europe pre-conference workshops	2023	0.5
• OHDSI Europe pre-conference workshops	2024	0.5

• Psychological factors are more strongly associated with pain than radiographic severity in non-invasively treated first carpometacarpal osteoarthritis International Federation of Societies for Surgery of the Hand (IFSSH) conference, Berlin	2019	0.5
• Patiënten met hogere verwachtingen zijn vaker tevreden met het behandelresultaat na conservatieve behandeling voor CMC-1 artrose Nederlandse Vereniging voor Plastische Chirurgie (NVPC) najaarscongres, Leeuwarden	2019	0.5
• Patients with higher treatment outcome expectations are more satisfied with the results of non-operative treatment for thumb base osteoarthritis: a cohort study Federation of European Societies for Surgery of the Hand (FESSH) conference, Online (due to COVID-19)	2020	0.5
• Predicting symptom improvement after carpal tunnel release: a machine learning approach FESSH conference, Online (due to COVID-19) European Federation of National Associations of Orthopaedics and Traumatology (EFORT) conference, Online (due to COVID-19)	2021	1.0
• Complications and patient-reported outcomes following trapeziectomy with a Weilby sling: a prospective cohort study FESSH conference, Online (due to COVID-19) EFORT conference, Online (due to COVID-19)	2021	1.0
• Could an unresolved personal injury claim affect outcomes of hand therapy and hand surgery? FESSH conference, Online (due to COVID-19) EFORT conference, Online (due to COVID-19)	2021	1.0
• Evaluation of cost-effectiveness of four treatment strategies for thumb base osteoarthritis: a microsimulation approach British Society for Surgery of the Hand (BSSH) conference, Oxford	2021	0.5

• Evaluation of cost-effectiveness of four treatment strategies for thumb base osteoarthritis: a microsimulation approach IFSSH conference, London	2022	0.5
• A single machine learning model can accurately predict post-treatment pain improvement in 26 hand and wrist condition-treatment combinations FESSH conference, Rimini	2023	0.5
• AI-gedreven data analyse en toepasbaarheid van predictiemodellen in de zorg Limé Medicolegal jaarcongres, Utrecht	2024	0.5
(Inter)national conference attendance		
• IFSSH, Berlin	2019	1.1
• NVPC, Leeuwarden	2019	0.3
• FESSH, Online	2020	1.1
• FESSH, Online	2021	1.1
• EFORT, Online	2021	0.9
• BSSH, Oxford	2021	0.6
• IFSSH, London	2022	1.1
• FESSH, Rimini	2023	1.1
• EFORT, Vienna	2023	0.9
• OHDSI Europe, Rotterdam	2023	0.3
• FESSH, Rotterdam	2024	1.1
• OHDSI Europe, Rotterdam	2024	0.3
2. **Teaching**		
Lecturing		
• Hand and wrist anatomy, Erasmus MC	2019 – 2023	3.0
• Elective “The musculoskeletal system”, Erasmus MC	2020 – 2023	1.0
• Academic development (R practicals), Erasmus MC – TU Delft	2021 – 2024	0.8
• Introduction to Engineering Research, TU Delft	2021 – 2024	4.0

Supervising		
• Master student – Seyed Hosseini *Effects of visualization of expected treatment outcomes for carpal tunnel surgery patients*	2020	0.7
• Master student – Jop Wagemans *Long-term outcomes of thumb base osteoarthritis surgery*	2020	0.7
• Research master student – Nina Loos *Prediction modeling for carpal tunnel syndrome and thumb base osteoarthritis*	2020 – 2022	2.2
• Minor students – Systematic review	2020	0.4
• Second-year students – Systematic review	2020 – 2022	1.0
• Minor students – Data analysis assignment	2021	1.0
• Bachelor student – Camille Blaaker *Factors associated with recurrence of dorsal wrist ganglions*	2023	0.2
• NIHES student – Camille Blaaker *Treatment patterns for thumb base osteoarthritis using the OMOP-CDM*	2023 – 2024	1.4
• NIHES student – Marta Sudilovskaya *Prediction of conversion to surgery for carpal tunnel syndrome using the OMOP-CDM*	2023 – 2024	1.4
Total		54.1

ABOUT THE AUTHOR

Lisa Hoogendam was born in Rotterdam on March 22nd, 1996. She grew up in Hellevoetsluis, where she lived with her parents and her younger sister. During high school, she participated in Junior Med School, which sparked her interest in medical research.

In 2014, she started her bachelor's degree in Medicine at the Erasmus Medical Center. She participated in the minor 'Reconstruction from Head to Hands', which led to her interest in hand surgery.

To gain more knowledge about and experience in medical research, she started the NIHES-research master Clinical Epidemiology in 2017. During this period, she successfully completed her research internship in the Hand-Wrist Study Group on the association between mindset factors and treatment outcomes in patients with thumb base osteoarthritis, under supervision of Mark van der Oest and Ruud Selles. She continued her research within the Hand-Wrist Study Group as a PhD project, in which she focused on the development and validation of prognostic prediction models for treatments of many hand and wrist conditions.

After the completion of her PhD period, which she very much enjoyed, she decided to pursue a career in research and data science, instead of a master's degree in medicine. In 2022, she started as a data scientist at Equipe Zorgbedrijven, contributing to improving data pipelines, standardized reporting, and various projects aiming to improve data-driven healthcare. From 2023 onwards, she combined this with a postdoctoral position at the Erasmus Medical Center, working on a European project about federated prediction modelling in rehabilitation medicine. Since early 2025, she focuses on research, and also contributes to a second European project on federated learning in the stroke care pathway.

Lisa lives in Soesterberg, together with her partner Steven and their two children Juliëtte and Berend. In her spare time, she enjoys being outdoors, cycling, and skiing with her family.