Research methodology in distal radius fracture care

1 step backward, 2 steps forward

Kleinlugtenbelt, Y.V.

Research methodology in distal radius fracture care

1 step backward, 2 steps forward

Ydo Vincent Kleinlugtenbelt

Research methodology in distal radius fracture care

1 step backward, 2 steps forward

Ydo Vincent Kleinlugtenbelt

Digital version of this thesis:

Link: https://www.gildeprint.nl/flippingbook/4076-research-methodology-in-distal-radius-fracture-care/

The publication of this thesis was kindly supported by:

Nederlandse Orthopaedische Vereniging
Nederlandse Vereniging voor Orthopedische Traumatologie
Stichting Deventer Evidence Based Orthopaedics
Wetenschappelijk Fonds Chirurgie AMC
Het Anna Fonds|NOREF
Deventer Ziekenhuis

Colofon

Author: Ydo Vincent Kleinlugtenbelt
Cover: Fotografiepolak
Lay-out: Gildeprint – The Netherlands
Printing: Gildeprint – The Netherlands
ISBN: 9789462337145

Research methodology in distal radius fracture care

1 step backward, 2 steps forward

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 27 september 2017, te 12:00 uur

door

Ydo Vincent Kleinlugtenbelt
geboren te Aalburg

PROMOTIECOMMISSIE:

Promotores:	Prof. dr. J.C. Goslings	AMC-UvA
	Prof. dr. M. Bhandari	McMaster University
Copromotores:	Dr. R.W. Poolman	Onze Lieve Vrouwe Gasthuis
	Dr. V.A.B. Scholtes	Onze Lieve Vrouwe Gasthuis
Overige leden:	Prof. dr. M. Maas	AMC-UvA
	Prof. dr. D. Eygendaal	AMC-UvA
	Dr. S.D. Strackee	AMC-UvA
	Prof. dr. R.J. de Haan	AMC-UvA
	Dr. N.W.L. Schep	Maasstad Ziekenhuis
	Prof. dr. M. Poeze	Maastricht UMC+

Faculteit der Geneeskunde

CONTENTS

GENERAL INTRODUCTION AND OUTLINE OF THE THESIS

The wrist joint consists of the distal radius, distal ulna, and the carpal bones. In the majority of patients, when clinicians speak about wrist fractures, they mean distal radius fractures (DRFs). The most common cause of this type of injury is a fall on an outstretched hand. In many patients, DRFs are clinically evident but they are always confirmed with plain radiographs and sometimes an additional computed tomography (CT). DRFs account for approximately 17% of all fractures and are the most common fracture in adults seen in the Emergency Department.(1) Annual incidence in adults in the Netherlands varies between 3 and 3.6 per 1000 population with the majority occurring in older women (over 50 years of age) after low energy trauma.(2) This variety is also influenced by the severity of the winter and the possibility of skating on natural ice in the Netherlands.(3)

Depending on the severity of the fracture and certain patient characteristics, patients with DRFs can be treated non-operatively with cast or operatively. In the last decade, there has been a shift toward more invasive treatment of DRFs; both from non-operative to operative treatment as well as from closed reduction and external fixation to (volar) plating.(4-6) It is suggested that this shift is mainly industry driven and not supported by strong evidence, but it has significant financial implications.(7) In the Netherlands, 13% of all DRFs in adults are treated operatively. However, operative treatment of DRFs is responsible for about 50% of the hospital costs of treating DRFs.(2)

There is currently a lack of consensus on the best treatment for certain types of DRFs. Recently, multiple randomized controlled studies have been published that did not show a clinically significant difference between different types of treatments for patients with DRFs.(8-11) However, many clinicians continue their surgical treatment since they are convinced that the outcome for their patients is better. Therefore, determining the most effective evidence-based treatments for DRFs is crucial. Randomized controlled trials (RCTs) are increasingly becoming more common to answer this sort of clinical question. To conduct best evidence clinical trials in DRF treatment and to exchange results globally in a standardized way, researchers need to use reliable inclusion criteria and validated outcome measures. These should ideally be determined prior to use, as the quality of such an instrument directly defines the quality of the information obtained with this instrument.(12) If outcomes are not determined prior to use, one risks imprecise or biased results, which might lead to misleading conclusions.

(13) The most common way to measure the quality of individual RCTs is the Cochrane Risk of Bias tool. The Cochrane tool does not include an evaluation of the quality of outcomes nor of measurement instruments used to assess eligibility for a trial. This means that researchers often overlook the possibility of unreliable or unvalidated (or insufficiently validated) measurement tools and how they can affect study results. The focus of this thesis is on an underappreciated aspect of methodology: the quality of patient-reported outcome measures (PROMs) and assessment tools used for diagnosis and trial eligibility in patients with DRFs.

This thesis aims to answer the following questions:

- What is the influence of computed tomography on the reliability and accuracy of classification systems for distal radius fractures? (Chapter 1)
- Can experienced surgeons predict the additional value of a CT scan in patients with displaced intra articular distal radius fractures? (Chapter 2)
- Are validated outcome measures used in distal radius fracture studies truly valid? (Chapter 3 and 4)
- What is the influence of spectrum bias on the intraobserver agreement of distal radius fracture treatment plans? (Chapter 5)
- Are volar locking plates superior to percutaneous K-wires for dorsally displaced distal radius fractures? (Chapter 6)

The best choice of treatment depends — to some extent — on the characteristics of the fracture (e.g. open/closed, non-displaced/displaced, extra-/intra-articular, etc.). The most commonly used classification systems in the literature that can classify any type of DRFs are the Frykman(14), Fernández(15), Universal(16), and AO/OTA classification(17).

The reliability, based on conventional radiographs, is poor for all classification systems.(18-22) Classifying the severity of fractures is especially important in clinical research, as the classified grade or type can be used as part of a study's eligibility criteria. However, in order to apply the treatment recommendations arising from these trials, the applicable classification systems must also be used in daily practice. Given the low degree of reliability using conventional radiography alone for fracture classifications, supplemental information may be warranted for more accurate and reproducible evaluations. A CT scan could possibly add some information about the characteristics of the fracture. However, little is known about the influence of an additional CT scan on the reliability and accuracy of the available classification systems. If the reliability of the classification systems improves when using an additional CT scan, we expect that this will positively affect the usefulness of classification systems used for trial inclusion.

In **Chapter 1** we will determine the intra- and inter-observer reliability of the most commonly used fracture classification systems, using both conventional radiography and conventional radiography with the addition of a CT in a representative clinical setting. Additionally, we determine the accuracy of the classification systems. Direct visualization through operative intervention would theoretically be the gold standard classification, but practically this is unrealistic. Both the volar and dorsal approach, which are used in the treatment of the majority of DRFs, do not provide an adequate view of the dorsal, volar and intra-articular comminuted fracture. We will use the CT scan as a reference standard instead of the "gold standard" to determine the accuracy of the classification systems using conventional radiography.(23)

The liberal use of CT scans in the management of DRFs, specifically for displaced intra-articular fractures, has become widely accepted as an additional imaging tool in pre-operative evaluation and planning.(24) In contrast to the shift to a more invasive treatment of DRFs, this change over the years is supported by previous literature. Furthermore, when a treatment plan is based on both X-ray and CT scan, a surgeon is more likely to treat the DRF patient surgically than when the treatment plan is based on X-ray alone.(25) Guidelines (e.g. AAOS, Dutch guidelines) have been developed to aid surgeons in decision making, but they are not clear about when to use an additional CT scan for the treatment of DRFs.(26) Some surgeons have a low threshold to request a CT scan for displaced intra-articular DRFs while others rarely obtain a CT scan. In **Chapter 2** we will investigate whether surgeons can predict the usefulness of a CT scan in patients with displaced intra-articular DRFs.

Historically, outcome assessment after DRFs has focused on imaging and physical examination (e.g. grip strength and range of motion). These assessments, however, do not represent the patients' perspective as they do not take the patients' feelings/opinion or wellbeing into account, which might be more important for the patient.(27) In the last two decades, the outcome assessment has shifted towards a patient-centered approach. This approach assesses the outcome directly from the opinion of the patient. Outcomes such as pain and functional ability, which are highly relevant for patients, can, for instance, be assessed by Patient Reported Outcome Measures (PROMs).(28)

Currently, a wide variety of PROMs are available and used to assess patient reported functional outcomes for upper limb and wrist disorders.(29-41) It is common to conduct validation and reliability studies for PROMS, however, they are rarely adequately validated in high-quality clinimetric studies. Several (non-)systematic studies reviewed the existing literature in order to present the best available PROMs for assessing wrist and hand function in general.(42-46) These conclusions were drawn based on the results of the available clinimetric studies, but these reviews did not evaluate the methodological quality of these studies.(26) In order to assess the methodological quality of a PROM, standards are needed. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) group set these standards for adequate study design and statistical analysis.(12) In an international Delphi study, the COSMIN group also developed a checklist in which consensus was reached on terminology, definitions, and a taxonomy of measurement properties of PROMs.(13)

In **Chapter 3** we therefore systematically review the methodological quality, using the COSMIN checklist, of the clinimetric studies that evaluated measurement properties of the available PROMs used in patients with DRFs. Based on the shortcomings shown in this review, in **Chapter 4** we further examine which PROM has the best measurement properties for evaluation of functional outcome in patients with a DRF by using a high-quality design following the COSMIN standards. The results of this study might help us to determine which PROM is most appropriate for the evaluation of patients with DRFs.

Another common yet unrecognized methodological issue in DRF studies is a phenomenon called spectrum bias, defined as the bias inherent when investigators choose a population lacking therapeutic uncertainty for evaluation. Spectrum bias is an issue that especially affects orthopedic agreement studies. Current studies on the additional benefit of using a CT scan to evaluate surgeon agreement on fracture treatment plans are inconsistent, varying from no agreement to almost perfect agreement. (5-7) In addition, the agreement on treatment plans did not even always improve when adding a CT scan compared to conventional radiographs alone. One explanation for these apparently inconsistent results in the literature may be attributed to differences in the chosen study population in these agreement studies. Ideally, the test results should be evaluated in a study population which is a perfect resemblance of the population of interest. If not, test results may be biased, as a result of spectrum bias. If a clinically less appropriate population is chosen for a study of a diagnostic test, the results may seriously mislead clinicians.(47) In **Chapter 5** we evaluate the potential influence of spectrum bias, and examine whether or not the agreement on treatment plans, with and without a CT scan, is related to the chosen population.

In **Chapter 6** we conduct a systematic review and meta-analysis of percutaneous Kirschner wire (PKW) fixation and plaster cast compared to volar locking plate (VLP) for displaced DRFs. We aim to compare functional outcome measures, range of motion, and complications between these two common surgical treatment options for DRFs. While answering the clinical question of which treatment is most favourable, we uncover examples of methodological shortcomings evident in the DRF literature. This exemplifies the underappreciated aspects of methodology, which I discuss throughout this thesis.

REFERENCES

(1) Court-Brown CM, Caesar B. Epidemiology of adult fractures: A review. Injury 2006 Aug;37(8):691-697.

(2) Kleinlugtenbelt YV. Polsbreuken worden in Nederland vooral conservatief behandeld. 2014; Available at: http://www.vektis.nl/index.php/nieuws/cijfers-en-feiten/477-polsbreuken-worden-in-nederland-vooral-conservatief-behandeld. Accessed May, 2014.

(3) van Lieshout AP, van Manen CJ, du Pre KJ, Kleinlugtenbelt YV, Poolman RW, Goslings JC, et al. Peak incidence of distal radius fractures due to ice skating on natural ice in The Netherlands. Strategies Trauma Limb Reconstr 2010 Aug;5(2):65-69.

(4) Mellstrand-Navarro C, Pettersson HJ, Tornqvist H, Ponzer S. The operative treatment of fractures of the distal radius is increasing: results from a nationwide Swedish study. Bone Joint J 2014 Jul;96-B(7):963-969.

(5) Mattila VM, Huttunen TT, Sillanpaa P, Niemi S, Pihlajamaki H, Kannus P. Significant change in the surgical treatment of distal radius fractures: a nationwide study between 1998 and 2008 in Finland. J Trauma 2011 Oct;71(4):939-42; discussion 942-3.

(6) Wilcke MK, Hammarberg H, Adolphson PY. Epidemiology and changed surgical treatment methods for fractures of the distal radius: a registry analysis of 42,583 patients in Stockholm County, Sweden, 2004-2010. Acta Orthop 2013 Jun;84(3):292-296.

(7) Martineau PA, Berry GK, Harvey EJ. Plating for distal radius fractures. Hand Clin 2010 Feb;26(1):61-69.

(8) Plate JF, Gaffney DL, Emory CL, Mannava S, Smith BP, Koman LA, et al. Randomized comparison of volar locking plates and intramedullary nails for unstable distal radius fractures. J Hand Surg Am 2015 Jun;40(6):1095-1101.

(9) Mellstrand Navarro C, Ahrengart L, Tornqvist H, Ponzer S. Volar Locking Plate or External Fixation With Optional Addition of K-Wires for Dorsally Displaced Distal Radius Fractures: A Randomized Controlled Study. J Orthop Trauma 2016 Apr;30(4):217-224.

(10) Bartl C, Stengel D, Bruckner T, Gebhard F, ORCHID Study Group. The treatment of displaced intra-articular distal radius fractures in elderly patients. Dtsch Arztebl Int 2014 Nov 14;111(46):779-787.

(11) Boyer M. Volar locking-plate and Kirschner-wire fixation did not differ in terms of functional outcomes after dorsally displaced distal radial fracture. J Bone Joint Surg Am 2015 May 20;97(10):859.

(12) Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010 May;19(4):539-549.

(13) Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-745.

(14) Frykman G. Fracture of the distal radius including sequelae--shoulder-hand-finger syndrome, disturbance in the distal radio-ulnar joint and impairment of nerve function. A clinical and experimental study. Acta Orthop Scand 1967:Suppl 108:3+.

(15) Fernandez DL. Distal radius fracture: the rationale of a classification. Chir Main 2001 Dec;20(6):411-425.

(16) Cooney WP. Fractures of the distal radius. A modern treatment-based classification. Orthop Clin North Am 1993 Apr;24(2):211-216.

(17) Marsh JL, Slongo TF, Agel J, Broderick JS, Creevey W, DeCoster TA, et al. Fracture and dislocation classification compendium - 2007: Orthopaedic Trauma Association classification, database and outcomes committee. J Orthop Trauma 2007 Nov-Dec;21(10 Suppl):S1-133.

(18) Cole RJ, Bindra RR, Evanoff BA, Gilula LA, Yamaguchi K, Gelberman RH. Radiographic evaluation of osseous displacement following intra-articular fractures of the distal radius: reliability of plain radiography versus computed tomography. J Hand Surg Am 1997 Sep;22(5):792-800.

(19) Dahlen HC, Franck WM, Sabauri G, Amlang M, Zwipp H. Incorrect classification of extra-articular distal radius fractures by conventional X-rays. Comparison between biplanar radiologic diagnostics and CT assessment of fracture morphology. Unfallchirurg 2004 Jun;107(6):491-498.

(20) Harness NG, Ring D, Zurakowski D, Harris GJ, Jupiter JB. The influence of three-dimensional computed tomography reconstructions on the characterization and treatment of distal radial fractures. J Bone Joint Surg Am 2006 Jun;88(6):1315-1323.

(21) Pruitt DL, Gilula LA, Manske PR, Vannier MW. Computed tomography scanning with image reconstruction in evaluation of distal radius fractures. J Hand Surg Am 1994 Sep;19(5):720-727.

(22) Johnston GH, Friedman L, Kriegler JC. Computerized tomographic evaluation of acute distal radial fractures. J Hand Surg Am 1992 Jul;17(4):738-744.

(23) Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. J Clin Epidemiol 2003 Nov;56(11):1118-1128.

(24) das Gracas Nascimento V, da Costa AC, Falcochio DF, Lanzarin LD, Checchia SL, Chakkour I. Computed tomography's influence on the classifications and treatment of the distal radius fractures. Hand (N Y) 2015 Dec;10(4):663-669.

(25) Hunt JJ, Lumsdaine W, Attia J, Balogh ZJ. AO type-C distal radius fractures: the influence of computed tomography on surgeon's decision-making. ANZ J Surg 2013 Sep;83(9):676-678.

(26) Murray J, Gross L. Treatment of distal radius fractures. J Am Acad Orthop Surg 2013 Aug;21(8):502-505.

(27) Darzi L. High Quality Care For All – NHS Next Stage Review Final Report. 2008.

(28) Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2(14):i-iv, 1-74.

(29) Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). Am J Ind Med 1996 Jun;29(6):602-608.

(30) MacDermid JC. Development of a scale for patient rating of wrist pain and disability. J Hand Ther 1996 Apr-Jun;9(2):178-183.

(31) Macey AC, Burke FD, Abbott K, Barton NJ, Bradbury E, Bradley A, et al. Outcomes of hand surgery. British Society for Surgery of the Hand. J Hand Surg Br 1995 Dec;20(6):841-855.

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

(33) Ware JE,Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483.

(34) Levine DW, Simmons BP, Koris MJ, Daltroy LH, Hohl GG, Fossel AH, 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 Nov;75(11):1585-1592.

(35) Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. Arthritis Rheum 1992 Jan;35(1):1-10.

(36) Smith HB. Smith hand function evaluation. Am J Occup Ther 1973 Jul-Aug;27(5):244-251.

(37) Lips P, Cooper C, Agnusdei D, Caulin F, Egger P, Johnell O, et al. Quality of life in patients with vertebral fractures: validation of the Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO). Working Party for Quality of Life of the European Foundation for Osteoporosis. Osteoporos Int 1999;10(2):150-160.

(38) Bialocerkowski AE, Grimmer KA, Bain GI. Development of a patient-focused wrist outcome instrument. Hand Clin 2003 Aug;19(3):437-48, ix.

(39) Kori S, Miller R, Todd D. Kinesiophobia: a new view of chronic pain behavior. Pain managment 1990:35-43.

(40) Rosenstiel AK, Keefe FJ. The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. Pain 1983 Sep;17(1):33-44.

(41) Altmaier EM, Russell DW, Kao CF, Lehmann TR, Weinstein JN. Role of self-efficacy in rehabilitation outcome among chronic low back pain patients. Journal of Counseling Psychology 1993;3:335-391.

(42) Hoang-Kim A, Pegreffi F, Moroni A, Ladd A. Measuring wrist and hand function: common scales and checklists. Injury 2011 Mar;42(3):253-258.

(43) Changulani M, Okonkwo U, Keswani T, Kalairajah Y. Outcome evaluation measures for wrist and hand: which one to choose? Int Orthop 2008 Feb;32(1):1-6.

(44) Bialocerkowski AE, Grimmer KA, Bain GI. A systematic review of the content and quality of wrist outcome instruments. Int J Qual Health Care 2000 Apr;12(2):149-157.

(45) Schuind FA, Mouraux D, Robert C, Brassinne E, Remy P, Salvia P, et al. Functional and outcome evaluation of the hand and wrist. Hand Clin 2003 Aug;19(3):361-369.

(46) Goldhahn J, Angst F, Simmen BR. What counts: outcome assessment after distal radius fractures in aged patients. J Orthop Trauma 2008 Sep;22(8 Suppl):S126-30.

(47) Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978 Oct 26;299(17):926-930.

CHAPTER 1

Classification systems for distal radius fractures. Does the reliability improve using additional computerized tomography?

<u>Y.V. Kleinlugtenbelt</u>
S.R. Groen
S.J. Ham
P. Kloen
R. Haverlag
M.P. Simons
V.A.B. Scholtes
M. Bhandari
J.C. Goslings
R.W. Poolman

Published in Acta Orthopaedica, 2017 Jun; 14:1-7.

ABSTRACT

Background

The reliability of conventional radiography when classifying distal radius fractures (DRF) is fair to moderate. We investigated whether reliability increases when additional computerized tomography scans (CT) are used.

Patients and methods

In this prospective study, we performed pre- and post-reduction posterior-anterior and lateral radiographs of 51 patients presenting with a displaced DRF. The case was included when there was a (questionable) indication for surgical treatment and an additional CT was conducted within 5 days. 4 observers assessed the cases using the Frykman, Fernández, Universal, and AO classification systems. The first 2 assessments were performed using conventional radiography alone; the following 2 assessments were performed with an additional CT. We used the intraclass correlation coefficient (ICC) to evaluate reliability. The CT was used as a reference standard to determine the accuracy.

Results

The intra-observer ICC for conventional radiography alone versus radiography and an additional CT was: Frykman 0.57 v. 0.51; Fernández 0.53 v 0.66; Universal 0.57 v. 0.64; AO 0.59 v. 0.71. The inter-observer ICC was: Frykman: 0.45 v 0.28; Fernández: 0.38 v. 0.44; Universal: 0.32 v. 0.43; AO: 0.46 v. 0.40.

Discussion

The intra-observer reliability of the classification systems was fair but improved when an additional CT was used, except for the Frykman classification. The inter-observer reliability ranged from poor to fair and did not improve when using an additional CT. Additional CT scanning has implications on the accuracy of scoring the fracture types, especially for simple fracture types.

BACKGROUND

Distal radius fractures were initially, since 1814, called “Pouteau” fractures and later renamed “Colles” fractures, after the Irish surgeon Abraham Colles.(1) At that time, no further distinctions were made into various subtypes of distal radius fractures. After the introduction of the roentgen and the growing awareness of the diversity of fracture features, the number of subtypes along with fracture eponyms increased. The first classification system was originally based on clinical features only, but additional classification systems have been developed since through the use of conventional radiographs.

The most common classification systems for distal radius fractures include Frykman(2), Fernández(3), Universal(4), and AO classification(5). An overview of reliability studies evaluating these 4 classification systems is presented in Tables 1 and 2. Using these full classification systems, the inter-observer reliability was fair to moderate. Good to excellent agreement was found only when using the AO classification of only the 3 types (Type A: extra-articular, Type B: partial articular and Type C: complete articular). This is comparable for the intra-observer reliability. However, 2 studies found substantial reliability for the universal classification.(6,7) Currently, there is no gold standard for classifying distal radius fractures.

Validated trauma classification systems offer a structured framework to communicate effectively about clinical cases, and support the treatment decision process (i.e., non-surgical vs. surgical management, type of surgical intervention). In addition, classifying the severity of fractures is important in clinical research, as the classified grade or type can be used as part of a study’s eligibility criteria. However, in order to apply the treatment recommendations arising from these trials, the applicable classification systems must also be used in daily practice. Given the low degree of reliability using conventional radiography alone for fracture classifications, supplemental information may be warranted for more accurate and reproducible evaluations.

Prior studies have used computer tomography (CT) to investigate the AO, Fernandez and Universal classification of distal radius fractures. However, these studies have been limited by lack of standardization on expertise of the reviewers (8) and focus on simplified versions of the original classifications (9). Currently, no evaluation of the utility of an additional CT scan on any original full classification system with experienced reviewers has been published. Additionally, prior studies have not evaluated radiographs versus radiographs plus CT scans.

Table 1 – Inter-observer reliability of distal radius fractures in literature (Cohen's kappa coefficient)

Study	Observers (n)	Cases (n)	Frykman	Fernández	Universal	AO (groups)	*AO (sub-groups)*	*AO (type)*	*AO (other)*
Andersen *et al.* (1996)	4	55	0.36 (X)			0.29 (X)	*0.25 (X)*	*0.64 (X)*	
Belloti *et al.* (2008)	5	98	0.24 (X)	0.34 (X)	0.39 (X)		*0.27 (X)*		
Flinkkilä *et al.* (1998)	5	30							*0.18 (X)*[A] *0.16 (XCT)*
Flinkkilä *et al.* (1998)	5	30							*0.23(X)*[B] *0.25 (XCT)*
Flinkkilä *et al.* (1998)	5	30							*0.48 (X)*[C] *0.78 (XCT)*
Illarremendi *et al.* (1998)	6	200							*0.37 (X)*[D]
Jin *et al.* (2007)	5	43	0.36 (X)		0.36 (X)	0.27 (X)		*0.46 (X)*	
Kreder *et al.* (1996)	36	30				0.48 (X)	*0.33 (X)*	*0.68 (X)*	
Kucuk et al. (2013)	10	50	0.40 (X)	0.46 (X)	0.32 (X)	0.34 (X)			
Kural (2010)	9	32	0.22 (X)	0.24 (X)	0.23 (X)	0.10(X)			
Naqvi *et al.* (2009)	6	25		0.27 (X)					
Oliveira Filho *et al.*(2004)	10	40	0.36 (X)		0.33 (X)		*0.19(X)*		
Oskam *et al.* (2001)	2	124						*0.65 (X)* [E] *(0.86 (X)* [F] *)*	
Van Leerdam (2010)	2	585				0.41 (X)	*0.33 (X)*	*0.60 (X)*	
Sirpakarn et al. (2013)	3	98	0.28 (X)	0.41 (X)		0.34 (X)			
YunesFilho *et al.*(2007) [G]	5	21			0.42 (X) 0.37 (CT)	0.32 (X) 0.21 (CT)	*0.21 (X)* *0.11 (CT)*	*0.47 (X)* *0.34 (CT)*	
Arealis et al. (2014)	5	26		0.43 (X) 0.40 (CT)	0.27 (X) 0.19 (CT)	0.30 (X) 0.30 (CT)			
Plant et al. (2015)	3	456				0.29(X)	*0.28(X)*	*0.56(X)*	
Buijtenen et al (2015)	6	54				0.48(X)		*0.49(X)*	
Distribution on X-ray alone			**0.24-0.40**	**0.27-0.46**	**0.33-0.42**	**0.10-0.48**	***0.19-0.33***	***0.46-0.68***	
Distribution CT-scan alone			**–**	**0.40-0.43**	**0.19-0.37**[G]	**0.21-0.30**[G]	***0.11***	***0.34***	
Distribution on X-ray and additional CT-scan			**–**	**–**	**–**	**–**			

The range of all published Cohen's kappa coefficients are shown in bold (X) assessment based on X-rays alone, (CT) assessment based on CT scan alone(XCT) assessment based on X-rays and additional CT-scan
[A] = AO simplified into 14 subgroup (A2.2-3; A3.1-3; C1.1-3; C2.1-3; C31.1-3)
[B] = AO simplified into 5 groups (A2, A3, C1, C2, C3)
[C] = AO simplified into2 types (A, C)
[D] = AO simplified into 5 groups (A, B, C1, C2, C3)
[E] = AO type A,B,C and extra group D (= not to be attributed to any AO type)
[F] =After consensus meeting
[G]= observers were residents

To address this current lack of knowledge, we aimed to determine the intra- and inter-observer reliability of the most commonly used fracture classification systems, using both conventional radiography and conventional radiography with the addition of a CT in a representative clinical setting of cases with a questionable indication for surgery. We evaluated the most commonly used classification systems that have been developed to classify any type of distal radius fracture; the Frykman(2), Fernández(3), Universal(4), and AO classification(5). By using experienced observers, we hypothesized that the intra-observer and inter-observer reliability is higher when using conventional radiography with additional CT. In addition, we determined the accuracy of the classification systems using the CT scan as a reference standard.

PATIENTS AND METHODS

Study design

A prospective database was established between January 1, 2007 and March 2, 2011 of patients with a displaced distal radius fracture seen at the emergency rooms in a hospital in Amsterdam (Onze Lieve Vrouwe Gasthuis).

Experience of the observers

The observers consisted of 4 experienced Dutch surgeons, of whom 2 were trauma surgeons [MS, RH] and 2 were orthopaedic surgeons [JH, PK]. Each had over 10 years of experience in fracture treatment. All of them were responsible for the (distal radius) fracture care within their department.

Study patients

Patients were eligible for inclusion if they 1) were 18 years-of-age or older presenting with a displaced distal radius fracture in the emergency department, 2) had pre- and post-reduction conventional posterior-anterior and lateral radiographs of the wrist, and 3) had an additional CT within 5 days in cases of a (questionable) indication for surgery. Questionable indication for surgery was defined as an inadequate reduction of the fracture as described by the AAOS guidelines (10) or in case of a presumably unstable fracture (11). Patients were excluded if they had a prior fracture or pathology of the distal radius.

Table 2 – Intra-observer reliability of classification systems in literature (Cohen's kappa coefficient)

Study	Observers (n)	Cases (n)	Frykman	Fernández	Universal	AO (groups)	AO (sub-groups)	AO (type)	AO (other)
Andersen *et al.* (1996)	4	55	0.40-0.61			0.35-0.45	0.23-0.38	0.58-0.70	
Belloti *et al.* (2008)	5	98	0.55	0.59	0.61		0.49		
Illarremendi *et al.* (1998)	6	200							0.57[A]
Jin *et al.* (2007)	5	43	0.54			0.36		0.49	
Kreder *et al.* (1996)	36	30					025-0.42	0.67-0.86	
Kucuk et al. (2013)	10	50	0.60	0.60	0.49	0.50			
Kural et al. (2010)	9	32	0.31	0.47	0.62	0.31			
Naqvi *et al.* (2009)	6	25		0.29-0.48					
Oliveira Filho *et al.* (2004)	10	40	0.55		0.54		0.38		
Ploegmakers *et al.*(2009)	45	5	0.26	0.42		0.52			
Siripakarn et al. (2013)	3	98	0.31	0.34		0.29			
Plant et al.(2015)						0.70		0.68	
Buijtenen et al.(2015)						0.53	0.49	0.65	
Distribution on X-ray alone			**0.26-0.60**	**0.42-0.60**	**0.54-0.62**	**0.36-0.70**	***0.23-0.49***	***0.49-0.86***	
Distribution CT-scan alone			–	–	–	–	–	–	
Distribution on X-ray and additional CT-scan			–	–	–	–	–	–	

The range of all published Cohen's kappa coefficients are shown in bold. [A] = AO reduction into 5 groups (A, B, C1, C2, C3)

Scoring procedure

The 4 observers independently classified the radiographic and/or CT images at 4 different time points. Each scoring round was performed with an interval of at least 4 weeks. All images were digitalized and anonymized.

Although increasing the number in either group would yield a more precise reliability estimate, the number of fractures has a greater impact on the precision than the number of observers (12). For this reason we chose a relatively low, but clinically representative, number of 4 observers.

At time points 1 and 2, the pre- and post-reduction conventional radiographs were used to classify the fracture according to the Frykman, Fernández, Universal and AO classification systems. At time points 3 and 4, both the conventional radiographs and all the 2D CT scan images were used (axial, sagittal and coronal planes). The order of the images was randomized at each time point. A short description of the 4 classification systems with additional illustrations was available for each observer.

Classification systems (Table 3)

The subgroups of the AO classification were not used in this study to simplify the evaluation and keep the number of grading criteria comparable to the other classification systems.

Sample size

Based on the methodology proposed by Giraudeau and Mary (2001), we used the expected value of the ICC, along with the number of raters and the desired confidence interval and confidence level to determine the number of subjects to be evaluated in this study. When using an additional CT, we expected a higher ICC than what is shown in previous literature when using conventional radiographs. We therefore estimated an ICC between 0.6 and 0.8. To obtain a 95% confidence interval (CI) with a confidence level of ±0.10 we needed between 30 and 81 patients.

Statistics

Classifications at time points 1 and 2 were used to determine the intra-observer reliability for the conventional radiographs for each observer separately. Classifications at time points 3 and 4 were used to determine the intra-observer reliability for the conventional radiographs with added CT scans for each observer separately.

Table 3 – Overview of included classification systems

Classification	Types (n)	Description types
Frykman (1967)	8 (I – VIII)	I – Extra articular, II – with ulna fracture III – Intra-articular into RC joint - IV with Ulna fracture V – Intra-articular into RU joint – VI with Ulna fracture VII – intra-articular into RC+RU joint – VIII with ulna fracture
Fernández (2001)	5 types	Based on trauma mechanism Type 1 – Bending fracture of metaphysis Type 2 – Shearing fracture of joint surface Type 3 – Compression fracture of joint surface Type 4 – Avulsion fractures or radiocarpal fracture-dislocation Type 5 – Combined fractures associated with high velocity injuries
Universal (1993)	4 types, subdivision in 2x3 groups	Type 1 – Extra articular fracture, without deviation; Type 2 – Extra articular fracture, with deviation 2A – Reducible and stable 2B – Reducible and unstable 2C – Irreducible Type 3 – Intra-articular fracture, without deviation Type 4 – Intra-articular fracture, with deviation 4A – Reducible and stable 4B – Reducible and unstable 4C – Irreducible
AO/ASIF (2007)	3 types, 9 groups	A – Extra articular fractures A1 - ulna #, radius intact A2 – radius fracture, simple and impacted A3 – radius fracture, multifragmentary B – Partial articular fractures B1 – radius #, sagittal B2 – Radius #, frontal, dorsal rim B3 Radius #, frontal, volar rim C – Complete articular fractures C1 – articular simple + metaphyseal simple C2 – articular simple, metaphyseal multifragmentary C3 – articular multifragmentary

Classifications at time points 1 and 3 were used to determine the inter-observer reliability for the conventional radiographs for each pair of observers (observer 1-2, 1-3, 1-4; 2-3, 2-4; 3-4) and we report the mean of these results with the associated CI.

We present descriptive statistics of the study patients, including means (SD) for continuous data. Intra- and inter-observer reliability was evaluated using the intraclass correlation coefficient (ICC). While other reliability studies have chosen a Kappa statistic, the ICC is able to take into account skewed data as well as to give credit for partial agreement. Kappa statistics are less accurate if responses are skewed and only appropriate for categorical data (13). Fleiss and Cohen (1973) showed that weighted kappa and ICC are equivalent in general cases when interval scales are used. To compare our results with previous literature, Cohen's Kappa was determined as well. The values were interpreted as described by Cicchetti (14); ICC values less than 0.40 indicate poor agreement, values between 0.40 – 0.59 indicate fair agreement, values between 0.60 – 0.74 indicate good agreement, and values ranging from 0.75 – 1.00 indicate excellent agreement. To determine the accuracy, direct visualization through operative intervention would theoretically be the gold standard classification diagnosis, but practically this is unrealistic. Both the volar and dorsal approach, which are used in the treatment of the majority of distal radius fractures, do not provide an adequate view of the dorsal, volar and intra-articular comminuted fracture. We used the CT scan as a reference standard instead of the "gold standard" to more accurately classify the fracture (15). With respect to the distribution of fracture types, absolute and percentile frequencies were calculated and differentiated according to radiographs (round 1) and radiographs with an additional CT (round 3). In addition, the percentage of change per fracture type was determined for all 4 classification systems. We compared the distribution of fracture classifications using conventional radiographs only and conventional radiographs with added CT scans for each classification system using chi-square tests. We corrected for multiple testing using a Bonferroni correction.

Ethical approval

Ethics approval was obtained from the medical ethical committee at this hospital (WO 10.086). We conducted this study according to the Collaboration for Outcome Assessment in Surgical Trials (COAST) guidelines.(13)

Funding

No external funding was received for this study.

Conflicts of interest

Authors declare that they have no competing interests.

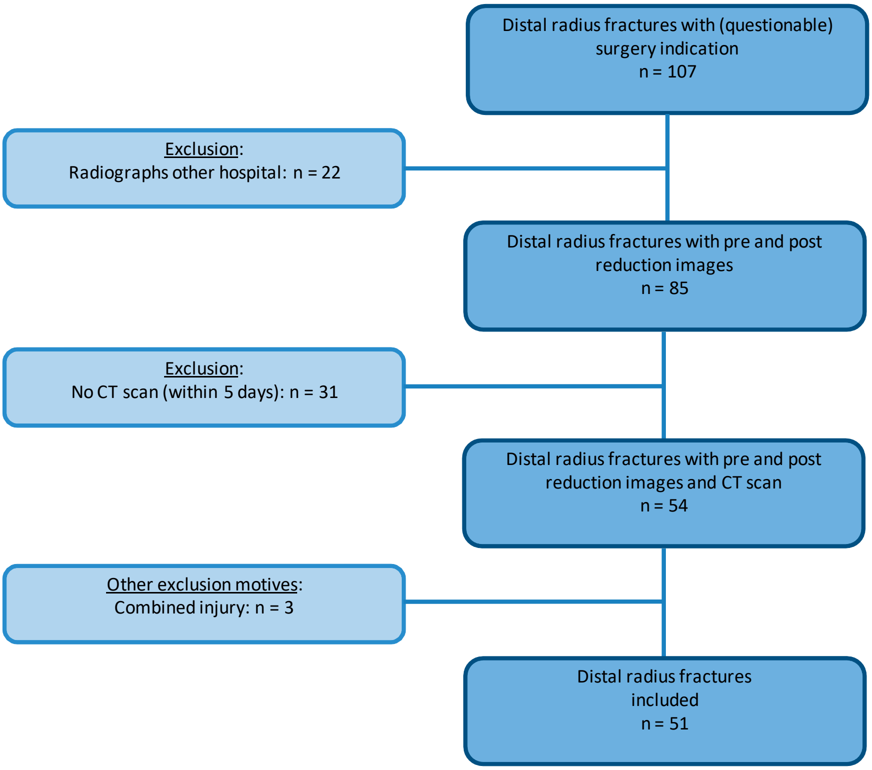

Figure 1; Flow chart of patients in the study

RESULTS

Study participants

From the 107 patients who entered the emergency room during the study period with a distal radius fracture with a (questionable) indication for surgery, 51 patients met the inclusion criteria (Figure 1). The included patients had a mean age of 50 (14) years. 38 patients (75%) were female. The post reduction CT scan was performed after a mean of 2.5 (2.2) days. The number of cases selected for surgical treatment ranged widely from 31%-96% between the 4 observers.

Reliability of classification systems

All ICCs for the intra-observer reliability with the range of the 4 observers are presented in Table 4. All ICCs for the inter-observer reliability and their respective 95% confidence intervals (CI) are presented in Table 5. The calculated Kappa values, to compare with previous literature, are presented in Tables 8 and 9 (See Supplementary data).

Table 4 – **Intra**-observer reliability

	Conventional radiograph		Conventional radiograph + CT scan		
Classification	ICC	Agreement	ICC	Agreement	P-value
Frykman	0.57(0.34-0.77)	Fair	0.51(0.33-0.80)	Fair	0.64
Fernandez	0.53(0.32-0.62)	Fair	0.66(0.53-0.90)	Good	0.13
Universal	0.57(0.43-0.71)	Fair	0.64(0.50-0.78)	Good	0.46
AO groups	0.59(0.51-0.66)	Fair	0.71(0.56-0.91)	Good	0.30

Mean ICC (intra-class correlation coefficient) of the intra-observer reliability with the range of the 4 observers in parentheses.

Table 5 – Inter-observer reliability

	Conventional radiograph		Conventional radiograph + CT scan		
Classification	ICC	Agreement	ICC	Agreement	P-value
Frykman	0.45 (0.31-0.60)	Fair	0.28 (0.14-0.44)	Poor	0.03
Fernandez	0.38 (0.21-0.55)	Poor	0.44 (0.30-0.59)	Fair	0.4
Universal	0.32 (0.18-0.48)	Poor	0.43 (0.20-0.51)	Fair	0.01
AO groups	0.46 (0.31-0.60)	Fair	0.40 (0.26-0.53)	Fair	0.4

Mean ICC (intraclass coefficient) of the inter-observer reliability with 95 % CI in parentheses.

Frykman classification

The mean ICC of the Frykman classification was 0.57 when using conventional radiographs, representing fair intra-observer reliability. The addition of CT showed no statistically significant improvement. The mean reliability was also fair (mean ICC = 0.51).

The mean ICC of the Frykman classification was 0.45 when using conventional radiographs, representing fair inter-observer reliability. The addition of CT scanning was less reliable (p=0.03). The mean intra-observer reliability was poor (mean ICC = 0.28).

Fernández classification

The mean ICC of the Fernández classification was 0.53 when using conventional radiographs, representing fair reliability. The addition of CT scanning showed a trend toward improvement but it was not statistically significant. The mean reliability was good (mean ICC = 0.66).

The mean ICC of the Fernández classification was 0.38 when using conventional radiographs, representing poor inter-observer reliability. The addition of CT scanning showed no statistically significant improvement. The mean reliability was fair (mean ICC = 0.44).

Universal classification

The mean ICC of the Universal classification was 0.57 when using conventional radiographs, representing fair intra-observer reliability. The addition of CT scanning showed no statistically significant improvement. The mean reliability was good (mean ICC = 0.64).

The mean ICC of the Universal classification was 0.32 when using conventional radiographs, representing poor inter-observer reliability. The addition of CT scanning showed significant improvement (p=0.01). The mean reliability was fair (mean ICC = 0.43).

AO classification

The mean ICC of the AO classification was 0.59 when using conventional radiographs, representing fair intra-observer reliability. The addition of CT scanning showed no statistically significant improvement. The mean reliability was good (mean ICC = 0.71).

The mean ICC of the AO classification was 0.46 when using conventional radiographs, representing fair inter-observer reliability. The addition of CT scanning showed no statistically significant change. The mean reliability was just fair (mean ICC = 0.40).

Distribution of fracture types with and without CT scan

The overall distribution of fracture types changed after adding a CT scan using the AO and Fernandez classification systems (p<0.001 and p=0.006 respectively). The overall distribution of fracture types did not significantly change using the Universal and Frykman classification systems (p=0.09 and p=0.06 respectively).

In general, in each classification system approximately half of the extra-articular fractures were classified as an intra-articular fracture when adding CT scanning. For example in the AO classification, the ratio of intra-articular to extra-articular fractures increased from 171:33 (=5) in round 1 (based on conventional radiography), to 189:15 (=13) in round 3 (based on CT scanning). The other 3 classifications showed a similar increase in the number of intra-articular fractures. In addition, when adding CT scanning the extra-articular fracture types were classified differently between 60% (Universal type 1) and 100% (Universal: 2B and AO: A3), as the scoring of the intra-articular fracture types changed between 17% (AO: C3) - 53% (Frykman: III/IV). Besides these features, the other statistically significant changes for each classification system are described below (Tables 6 and 7).

Table 6 – Distribution of fracture types

				Frykman				
	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**
X	9	5	20	25	3	1	18	19
X + CT	5	2	18	16	3	0	24	31
				Fernandez				
	1	**2**	**3**	**4**	**5**			
X	18	11	50	1	20			
X + CT	7	21	59	0	13			
				Universal				
	1	**2A**	**2B**	**2C**	**3**	**4A**	**4B**	**4C**
X	2	8	4	0	4	25	39	17
X + CT	1	5	1	0	9	27	32	24
				AO groups				
	A2	**A3**	**B1**	**B2**	**B3**	**C1**	**C2**	**C3**
X	12	4	1	3	12	31	30	6
X + CT	2	5	3	5	12	19	32	22

Distribution of fracture types in round 1 (conventional radiography = X) and round 3 (conventional radiography with additional CT scan = X+CT) of all four observers given in percentage (%). The coloured boxes show significant changes in fracture distribution in that category.

Table 7 – Percentage of changes in classification after adding a CT scan

			Frykman			
I + II	**III + IV**	**V + VI**	**VII + VIII**			
69	48	78	0			
			Fernandez			
1	**2**	**3**	**4**	**5**		
69	41	7	n/a	0		
			Universal			
1	**2A**	**2B**	**3**	**4A**	**4B**	**4C**
60	71	75	25	29	23	0
			AO groups			
A2	**A3**	**C1**	**C2**	**C3**		
88	89	51	25	0		

Percentage of changes in fracture types after adding a CT scan (Round 1 versus 3). Frykman: The fracture types with and without an ulna fracture are added together. Fernandez: type 4 and Universal type 2c were not taken into account, because these were not classified.

Frykman

Only the number of fractures classified as type VIII (intra-articular radio-ulnar and radio-carpal joint with an ulna fracture) increased. The other changes were not statistically significant.

Fernández

The number of fractures classified as type 2 (shearing fracture of joint surface) increased, while the number of fractures classified as type 1 (bending fractures of metaphysis) decreased.

Universal

None of the changes in distribution were statistically significant.

AO classification

The number of fractures classified as type C3 (intra-articular multifragmentary) increased, while the number of fractures classified as type A2 (extra-articular simple) and type C1 (articular simple+metaphyseal simple) decreased.

DISCUSSION

In contrast to our hypothesis, the results of this study revealed that the increase in reliability when using additional CT scanning was only seen in the intra-observer reliability, with the exception of the Frykman classification. The Frykman classification distinguishes between intra-articular and extra-articular fractures of the distal radioulnar joint. On conventional radiography, the distal radioulnar joint fracture line is not always clearly imaged and therefore generally not taken into account in the classification evaluation. However, on a CT scan a small fracture line is often seen in the region of the distal radioulnar joint, allowing room for interpretation and potentially discrepant results. This could explain the decrease in reliability of the Frykman classification when using additional CT scanning.

3 prior studies also used CT scanning to investigate the inter-observer reliability of the AO classification for distal radius fractures (8,9,16). In our study, the inter-observer reliability (kappa values) was found to be comparable, both using conventional radiography alone and with an additional CT scan. One would expect a higher reliability, since determining the three-dimensional morphology of the fracture might be more difficult when CT scans are not combined with radiographs. Surprisingly, using an additional CT scan and only experienced observers in our study the reliability did not improve in comparison to some other studies that used a CT scan alone and observers of all levels of experience (Yunes Filho et al. 2009, Arealis et al. 2014).

Using the CT scan as the reference standard, we can state that simple fracture types are less accurately classified when using only a radiograph than more severe types. This is contradictory to the clinical practice in which CT scans are especially used in the more severe fracture types to plan treatment. Similar to other published reports (17,18), we found a systematic decrease of about 50% in the ratio of extra- to intra-articular fractures when the CT scan was added to conventional radiographs.

Furthermore, our results confirm earlier statements that the severity of a distal radius fracture may be underestimated in standard radiographs. For instance, Cole et al. (1997) reported an improved reliability of assessing specific displacement features, in particular the measurement of gapping or stepping-off, based on CT compared to conventional radiography. This is best shown by the AO classification as the number of type C3 (articular multifragmentary) fractures, increased after adding CT to the evaluation.

Rozental et al. (2001) and Heo et al. (2012) reported that sigmoid notch involvement is underestimated when using only conventional radiography. This feature is also seen in our study. The Frykman classification distinguishes between intra-articular and extra-articular fractures of the distal radioulnar joint. As shown in Table 3, the number of type VIII (intra-articular distal radioulnar joint and distal radiocarpal joint with ulna fracture) increased when using the additional CT scan as also shown by Goldwyn et al (2012).

We suggest use of the AO classification, as this is currently the classification system most frequently used and the reliability is comparable to the other classification systems. Preferably, a new classification system also based on CT instead of conventional radiography alone should be developed. Such new classification system should mainly focus on giving direction to the type of treatment.

A limitation of our study is that the sample size was underestimated for the inter-observer reliability. The pre-specified estimation of ICC for intra-observer reliability (as described in the methods) was comparable to our estimation so we can be confident that our number of raters is sufficient. The range of intra-observer ICCs was relatively large for some classifications, however, replacing either the best or worst observer would be unlikely to change the conclusion that an additional CT scan improves reliability. It would likely only affect the absolute ICC, not the difference between conventional radiography only and conventional radiography with additional CT scan. By choosing a group of patients with a (questionable) indication for surgery we introduced a selection bias, which possibly influenced our results and therefore risks a lack of generalizability to other patients. However, the optimal treatment of this group of patients lacks consensus and therefore, these patients will likely benefit most from additional evaluation criteria for accurate classification. Although the intra-observer reliability improved from fair to good, the p-values were not significant. However, these T-tests were likely to be underpowered and had a high rate of Type II error. Previous studies have shown better reliability for younger patients when classifying DRFs (19). The relatively low mean age in this study may affect the outcome and could give a higher reliability. Additionally, it is important to note that there is some inferential uncertainty with these results. It may be difficult to apply our results to broader populations, although we took precautions to ensure a representative sample and as close to a real-world clinical setting as possible.

A strength of our study is that we used the COAST criteria to ensure we addressed all components of a reliability study. Another strength is that the number of patients selected for surgical treatment ranged widely between the 4 observers, showing that this group of patients is representative of the group of patients lacking consensus.

Our study results suggest that the additional value of CT scanning over conventional radiographs is limited in regard to reliability. However, it has significant implications for accurate scoring of the fracture types. Using an additional CT scan changes how patients are classified into fracture types, therefore trials using conventional radiography alone to evaluate eligibility will have different patients included compared to trials using additional CT scans. This has implications for external validity (generalizability) and for comparing trials to each other.

Although previous literature showed that CT scans are more reliable than conventional radiographs quantifying articular surface incongruencies, to our knowledge no previous studies have reported the impact of intra-articular involvement without a step or gap on clinical outcome. The outcomes of the current study are not necessarily related to better patient outcomes. Prospective randomized studies - comparing conventional radiographs for patients with displaced DRF to additional CT scans - should be conducted to confirm the additional value of a CT scan for patient outcomes. Also a cost effectiveness analysis should be conducted as national care budgets are limited.

In summary, our study results suggest that the additional value of CT scanning over conventional radiographs is limited in regard to reliability, but has significant implications for accurate scoring of the fracture types. The reliability of the classification system might be decreased due to the fact that the additional information about fracture morphology provided by the additional CT scan leaves increased room for interpretation when classifying a distal radius fracture.

CONCLUSION

To our knowledge this is the first reliability study on 4 classification systems, which determines the intra- and inter-observer reliability using conventional radiography alone and with additional CT scanning. The intra-observer reliability of the classification systems was found to be fair but improves to good agreement if an additional CT is used, with the exception of the Frykman classification. The inter-observer reliability of the investigated classification systems for distal radius fractures was poor to fair and did not improve when using additional CT scanning. Additional CT scanning has significant implications for accurate scoring of the fracture types in AO and Fernandez classifications, especially for the less severe fractures.

REFERENCES

(1) Colles A. Historical paper on the fracture of the carpal extremity of the radius (1814). Injury 1970 Jul;2(1):48-50.

(2) Frykman G. Fracture of the distal radius including sequelae--shoulder-hand-finger syndrome, disturbance in the distal radio-ulnar joint and impairment of nerve function. A clinical and experimental study. Acta Orthop Scand 1967:Suppl 108:3+.

(3) Fernandez DL. Distal radius fracture: the rationale of a classification. Chir Main 2001 Dec;20(6):411-425.

(4) Cooney WP. Fractures of the distal radius. A modern treatment-based classification. Orthop Clin North Am 1993 Apr;24(2):211-216.

(5) Marsh JL, Slongo TF, Agel J, Broderick JS, Creevey W, DeCoster TA, et al. Fracture and dislocation classification compendium - 2007: Orthopaedic Trauma Association classification, database and outcomes committee. J Orthop Trauma 2007 Nov-Dec;21(10 Suppl):S1-133.

(6) Kural C, Sungur I, Kaya I, Ugras A, Erturk A, Cetinus E. Evaluation of the reliability of classification systems used for distal radius fractures. Orthopedics 2010 Nov;33(11):801.

(7) Belloti JC, Tamaoki MJ, Franciozi CE, Santos JB, Balbachevsky D, Chap Chap E, et al. Are distal radius fracture classifications reproducible? Intra and interobserver agreement. Sao Paulo Med J 2008 May 1;126(3):180-185.

(8) Yunes Filho PRM, Pereira Filho MV, Gomes DCP, Medeiros RS, De Paula EJL, Mattar Junior R. Classifying radiusfractures with X-RAY and Tomography Imaging. Acta Orthopedica Brasileira 2009;17(2):9-13.

(9) Flikkila T, Nikkola-Sihto A, Kaarela O, Paakko E, Raatikainen T. Poor interobserver reliability of AO classification of fractures of the distal radius. Additional computed tomography is of minor value. J Bone Joint Surg Br 1998 Jul;80(4):670-672.

(10) Lichtman DM, Bindra RR, Boyer MI, Putnam MD, Ring D, Slutsky DJ, et al. Treatment of distal radius fractures. J Am Acad Orthop Surg 2010 Mar;18(3):180-189.

(11) Mackenney PJ, McQueen MM, Elton R. Prediction of instability in distal radial fractures. J Bone Joint Surg Am 2006 Sep;88(9):1944-1951.

(12) Steiner DL, Norman GR. Health Measurement Scales: A practical guide to their development and use. 3rd ed.: Oxford University Press; 2013.

(13) Karanicolas PJ, Bhandari M, Kreder H, Moroni A, Richardson M, Walter SD, et al. Evaluating agreement: conducting a reliability study. J Bone Joint Surg Am 2009 May;91 Suppl 3:99-106.

(14) Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology.. Psychological Assessment 1994;6(4):284-290.

(15) Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. J Clin Epidemiol 2003 Nov;56(11):1118-1128.

(16) Arealis G, Galanopoulos I, Nikolaou VS, Lacon A, Ashwood N, Kitsis C. Does the CT improve inter- and intra-observer agreement for the AO, Fernandez and Universal classification systems for distal radius fractures? Injury 2014 Oct;45(10):1579-1584.

(17) Dahlen HC, Franck WM, Sabauri G, Amlang M, Zwipp H. Incorrect classification of extra-articular distal radius fractures by conventional X-rays. Comparison between biplanar radiologic diagnostics and CT assessment of fracture morphology. Unfallchirurg 2004 Jun;107(6):491-498.

(18) Johnston GH, Friedman L, Kriegler JC. Computerized tomographic evaluation of acute distal radial fractures. J Hand Surg Am 1992 Jul;17(4):738-744.

(19) Wadsten MA, Sayed-Noor AS, Sjoden GO, Svensson O, Buttazzoni GG. The Buttazzoni classification of distal radial fractures in adults: interobserver and intraobserver reliability. Hand (N Y) 2009 Sep;4(3):283-288.

SUPPLEMENTARY DATA

Table 8 – **Intra**-observer reliability

	Conventional radiograph		Conventional radiograph + CT scan		
Classification	Kappa	Agreement	Kappa	Agreement	P-value
Frykman	0.49 (0.32-0.66)	Moderate	0.50 (0.35-0.74)	Moderate	0.8
Fernandez	0.45 (0.28-0.65)	Moderate	0.54 (0.27-0.90)	Moderate	0.6
Universal	0.40 (0.31-0.44)	Fair	0.32 (0.22-0.46)	Fair	0.4
AO groups	0.40 (0.36-0.44)	Fair	0.41 (0.31-0.50)	Moderate	0.9

Mean Kappa of the intra-observer reliability with the range of the 4 observers in parentheses. This is presented for readers to compare kappa values with the previous literature.

Table 9 – Inter-observer reliability

	Conventional radiograph		Conventional radiograph + CT scan		
Classification	Kappa	Agreement	Kappa	Agreement	P-value
Frykman	0.40(0.32-0.48)	Fair	0.36 (0.29- 0.42)	Fair	0.3
Fernandez	0.28(0.09-0.46)	Fair	0.26 (0.16- 0.36)	Fair	0.9
Universal	0.20(0.15-0.25)	Slight	0.11 (0.05- 0.17)	Slight	0.05
AO groups	0.26(0.20-0.33)	Fair	0.28 (0.076-0.30)	Fair	0.2

Mean Kappa of the inter-observer reliability with 95 % CI (confidence interval) in parentheses. This is presented for readers to compare kappa values with the previous literature.

CHAPTER 2

Can experienced surgeons predict the additional value of a CT scan in patients with displaced intra articular distal radius fractures?

Y.V. Kleinlugtenbelt
K. Madden
S.R. Groen
S.J. Ham
P. Kloen
R. Haverlag
M.P. Simons
M. Bhandari
J.C. Goslings
V.A.B. Scholtes
R.W. Poolman

Published in Strategies in Trauma and Limb Reconstruction, 2017 Aug; 12(2): 91-97

ABSTRACT

Background

There are no clear guidelines when an additional CT-scan needs to be obtained for the treatment of displaced intra-articular distal radius fractures (DRF). The aim of this study was to investigate whether surgeons can predict the usefulness of a CT scan to facilitate choice of treatment plan and/or preoperative planning for these fractures.

Patients and methods

our surgeons evaluated 51 patients with displaced DRF. The choice of treatment (operative or nonoperative) was first based on conventional radiographs. Subsequently, the surgeons were asked if they would have requested an additional CT scan to determine this treatment choice, but also if they required a CT scan for the preoperative planning. After 4 weeks the additional CT scan was directly provided along with the radiographs for all patients. The choice of treatment was again assessed and whether the CT scan was useful for operative planning. Based on these data we calculated the number needed to scan (NNS) and number needed to harm (NNH) for two decision models. <u>Model 1</u>: Only provide a CT scan if the surgeon requested one based on their judgment of the X-rays. <u>Model 2</u>: CT scans for all displaced intra-articular DRF.

Results

For choice of treatment the NNS was lower for model 1 than for model 2 (2.6 versus 4.3) and the NNH is higher for model 1 (3.1 versus 1.3). For preoperative planning the NNS (1.3 versus 1.4) and NNH (3.7 versus 3.4) were comparable for both models.

Discussion

Surgeons are able to predict the usefulness of an additional CT scan for intra-articular displaced DRF regarding whether one treats the patient operatively or non-operatively. However, for pre-operative planning the usefulness of a CT scan is much harder to predict.

BACKGROUND

Multiple studies have shown an increased number of computed tomography (CT) requested by physicians(1-3). This issue has become a subject of concern for patients, health care providers and regulators, and is receiving increasing attention in the medical literature(4). CT scans are more expensive than X-rays and national health care budgets are limited.

Recently the liberal use of CT scan in the management of distal radius fractures (DRF), specifically for displaced intra-articular fractures, became widely accepted as an additional imaging tool in pre-operative evaluation and planning(5). This increased popularity is supported by previous literature. CT scans have been shown to be more reliable than X-rays in quantifying articular surface incongruencies(6-10). Furthermore, when a treatment plan is based on both X-ray and CT scan, a surgeon is more likely to treat the DRF patient surgically than when the treatment plan is based on X-ray alone(11).

Does this mean that we should request a CT scan for all patients with displaced intra articular fractures? Much information required for treatment planning can be obtained from plain X-rays. However, CT scans can give additional information which is not always seen on the X-ray. In DRF management CT scans can be requested by the surgeon for two reasons: Firstly, to decide whether to treat the patient operatively or nonoperatively, and secondly, for pre-operative planning purposes. Guidelines (e.g. AAOS, Dutch guidelines) have been developed to aid surgeons in decision making, but they are not clear about when to use an additional CT scan for the treatment of DRF(12). Some surgeons have a low threshold to request a CT scan for displaced intra articular DRF while others rarely obtain a CT scan.

The aim of this study was to investigate whether surgeons can predict the usefulness of a CT scan in patients with displaced intra-articular DRF. This was done by comparing two decision models for when to request an additional CT scan.

METHODS

Experimental design

In our experiment we compared two different models of decision making.

Model 1: Only provide a CT scan if the surgeon requested one based on their judgment of the X-rays.
Model 2: CT scans for all displaced intra-articular DRF.

Moreover, we investigated this for two reasons for requesting an additional CT scan: 1) to decide whether to treat the patient operatively or nonoperatively (OR indication); and 2) for preoperative planning (OR preparation).

Patient Selection

Consecutive patients with displaced distal radius fractures were selected from our Emergency Department database. The protocol in the recruiting department is to always order a CT scan for patients with a displaced intra-articluar distal radius fracture. The CT scan was obtained by the resident on call. Patients were eligible for inclusion if they 1) presented with a displaced DRF in the Emergency Department between January 1, 2007 and March 2, 2011, 2) were 18 years of age or older, 3) had no prior fracture or pathology of the distal radius, 4) had both pre- and post-reduction plain posterior-anterior and lateral radiographs of the wrist, and 5) had an additional post-reduction CT taken within 5 days after the reduction.

Observers

Four experienced trauma surgeons reviewed the images. They all have over 10 years of experience in fracture treatment. All of them are responsible for the distal radius fracture care within their department.

Time points

All surgeons scored the images at four different time points (T1-T4). Each scoring round was performed with an interval of at least 4 weeks.

T1 and T2: pre- and post-reduction plain radiographs.
T3 and T4: pre- and post-reduction plain radiographs & axial, sagittal and coronal planes CT.

All images were digitized and anonymized, and presented with the relevant clinical data (e.g., age of the patient, gender, dominant hand, profession and specific hobbies).

Scoring form

We used two scoring forms with the following questions

T1 and T2:

1) Type of fracture: intra or extra articular
2) Choice of treatment plan: nonoperative treatment with plaster after closed reduction or operative treatment (OR indication)
3) Would you request a CT scan for OR indication?: Yes or No?
4) If treated operatively, would you request a CT scan for pre-operative planning (OR preparation)?: Yes or No

T3 and T4:

1) Choice of treatment plan: nonoperative treatment with plaster after closed reduction or operative treatment (OR indication).
2) Was the CT scan useful for OR preparation?: Yes or No

Methods to prevent bias

Surgeons were not informed we were testing two decisions models and were blinded to the study hypotheses. We informed them that they were participating in an inter observer reliability study. The order of the images was randomized to differ at all time points. Cases were presented in random order at different time points to prevent recall bias.

Statistical Analysis

We compared across timepoints with X-ray only (T1 and T2) and X-ray+CT scan (T3 and T4) which gives a total of four comparisons: T1-T3, T1-T4, T2-T3 and T2-T4. Therefore, we have 204 observations (4 surgeons times 51 cases) per comparison and 816 observations (204 observations times 4 comparisons) in total for our statistical analysis.

In both models of decision making we assessed the number needed to scan (NNS) and number needed to harm (NNH) separately for the two reasons a CT scan is ordered (OR indication and OR preparation). NNS is akin to number needed to treat (NNT) for treatment studies. It is the number of patients who need a CT scan to achieve one additional good outcome. When requested for OR indication, a change in treatment (Tx change) is defined as a good outcome. When requested for OR preparation, useful for operative planning is defined as a good outcome. NNH is the number of patients on average need to be exposed to a risk factor to cause harm in an average of one patient who would not otherwise have been harmed. We defined two risk factors which could harm the patient. 1) Unnecessary radiation: Requesting a CT scan unnecessarily can harm a patient due to excess radiation. 2) Suboptimal informed: Not requesting a CT scan when in fact the CT scan would have been useful can harm the patient because the surgeon is suboptimally informed about the fracture. The total NNH and the NNH for the two risk factors is determined separately. A low NNS and a high NNH is preferred.

In clinical practice, when a CT scan is ordered and the decision is made to treat the patient operatively based on the CT scan, it is not necessary to predict whether we need a CT scan for operative planning, because the CT scan is already available. We conducted a sensitivity analysis in which only patients who were treated operatively but who did not have a CT scan ordered for OR indication were included in the analysis.
Formulas for NNS/NNH

Model 1:

$$NNS_{\text{OR indication}} = \frac{1}{(Tx\ change\ when\ CT\ requested)/(Total\ CT\ requested)}$$

$$NNH_{\text{ORindication}} = \frac{1}{(No\ tx\ change\ when\ CT\ requested)+(Tx\ change\ when\ CT\ not\ requested)/(All\ observations)}$$

$$NNS_{\text{OR preparation}} = \frac{1}{(CT\ useful\ when\ CT\ requested)/(Total\ CT\ requested)}$$

$$NNH_{\text{OR preparation}} = \frac{1}{(CT\ not\ useful\ when\ CT\ requested)+(CT\ useful\ when\ CT\ not\ requested)/(All\ operative\ observations)}$$

Model 2:

$$NNS_{\text{OR indication}} = \frac{1}{(Tx\ change\ when\ CT\ requested)+(Tx\ change\ when\ CT\ not\ requested)/(All\ observations)}$$

$$NNH_{\text{OR indication}} = \frac{1}{(No\ tx\ change\ when\ CT\ requested)+(No\ tx\ change\ when\ CT\ not\ requested)/(All\ observations)}$$

$$NNS_{\text{OR preparation}} = \frac{1}{(CT\ useful\ when\ CT\ requested)+(CT\ useful\ when\ CT\ not\ requested)/(All\ operative\ observations)}$$

$$NNH_{\text{OR preparation}} = \frac{1}{(CT\ not\ useful\ when\ CT\ requested)+(CT\ not\ useful\ when\ CT\ not\ requested)/(All\ operative\ observations)}$$

We used SPSS version 22 to conduct these analyses.

RESULTS

Patient characteristics

During the study period 85 patients who entered the Emergency Department with a displaced DRF had a post-reduction CT scan. A total of 51 patients met the complete inclusion criteria. Their mean age was 50 years (SD, 14). 75 percent of the patients were female. The CT scan was performed a mean of 2.53 days post-reduction (SD, 2.21). Out of 816 observations, 688 times they were scored as intra-articular fractures (Figure 1 and 2) and were included in this study.

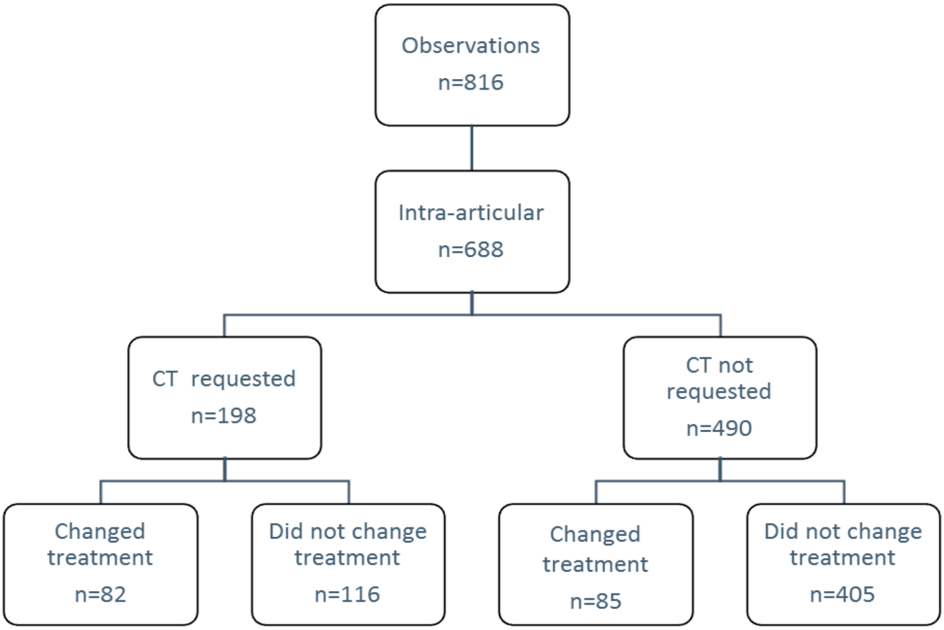

Figure 1: OR indication flowchart. Total score of the 4 observers.

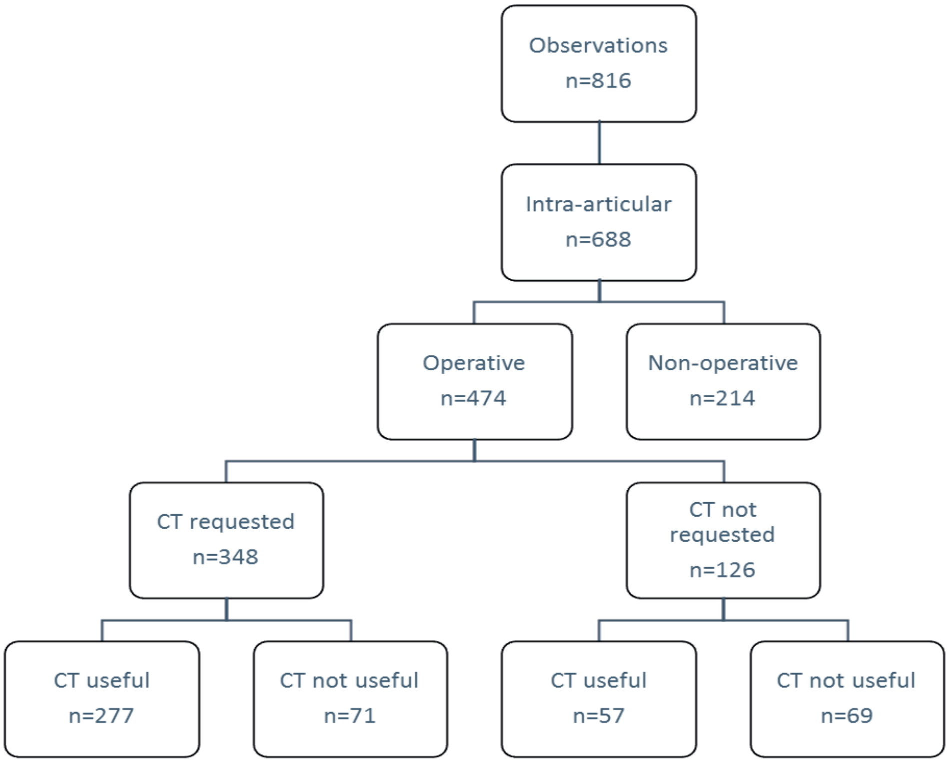

Figure 2: OR preparation flowchart. Total score of the 4 observers.

CT scan for OR indication

Of the 688 intra-articular observations based on X-rays only, the surgeon requested a CT scan for OR indication in 198 cases (198/688, 28.8%) (Figure 1). For the cases where the surgeon requested a CT scan, the surgeon changed their treatment plan 41.4% of the time (82/198) with an additional CT scan. For the cases where the surgeon did not request a CT scan the surgeon changed their treatment plan 17.3% of the time (85/490) with an additional CT scan. The NNS was lower for model 1 than for model 2 (2.6 versus 4.3) and the NNH is higher for model 1 (3.1 versus 1.3) (Table 1).

Table 1. CT scan for OR indication (with 95 % CI)

	NNS	NNH total	*NNH: Suboptimal informed*	*NNH: Unnecessary radiation*
Model 1 (CT on request)	**2.4** (0.42)	**3.4** (0.39)	*8.1 (0.78)*	*5.9 (1.04)*
Model 2 (CT for all)	**4.1** (0.30)	**1.3** (0.03)	*0*	*1.3 (0.03)*

CT scan for OR preparation

In 474 observations the surgeon decided to treat the patient operatively based on X-rays. The surgeon requested a CT scan for OR preparation in 348 cases (348/474, 73%) (Figure 2). For the cases where the surgeon requested a CT scan for OR preparation, the CT scan was useful 80% of the time (277/348). For the cases where the surgeon did not request a CT scan, the CT scan was useful 45% of the time (57/126). The NNS (1.3 versus 1.4) and NNH (3.7 versus 3.4) are comparable for both models (Table 2). In the sensitivity analysis, which more closely approximates a real clinical situation, the NNS (1.2 versus 1.3) and NNH (4.3 versus 4.8) only slightly changed (Table 2).

Table 2. CT scan for OR preparation (with 95 % CI) (Corrected: sensitivity analysis for clinical practice)

	NNS	NNH total	*NNH: Suboptimal informed*	*NNH: Unnecessary radiation*
Model 1 (CT on request)	**1.3** (0.02)	**3.7** (0.26)	*8.3 (1.2)*	*6.7 (0.35)*
Model 2 (CT for all)	**1.4** (0.06)	**3.4** (0.36)	*0*	*3.4 (0.36)*
Model 1 corrected	**1.2** (0.02)	**4.3** (0.42)	10.2 (1.21)	*7.6 (0.67)*
Model 2 corrected	**1.3** (0.03)	**4.8** (0.27)	0	*4.8 (0.27)*

DISCUSSION

Main outcomes

Surgeons were able to predict the usefulness of an additional CT scan for intra-articular displaced DRF regarding whether one will advise the patient to be treated operatively or non-operatively. The NNS for OR indication was clearly lower when the surgeons predict that the additional CT scan will be useful. The NNH was also lower, which means that when surgeons predict the usefulness of an additional CT scan, fewer CT scans are needed and therefore fewer patients are harmed than when surgeons order CT scans for all patients with intra-articular displaced DRF. The only disadvantage is that in one out of eight patients (NNH: 8.1) the surgeon is suboptimally informed compared to if they had had an additional CT scan.

However, for pre-operative planning the usefulness is much harder to predict, so it is more defendable to request CT scans for all intra-articular displaced DRF which are operatively treated. The NNS and NNH in each model are similar, even when corrected for the clinical practice. To choose the appropriate decision making model for OR preparation, surgeons must weigh the pros and cons of each model based on which harm they think is most important to avoid. If surgeons request an additional CT scan for all operatively treated DRFs, there will be no patients for which the surgeon is suboptimally informed, but one in 3.4 patients will be exposed to radiation unnecessarily. If surgeons predict that an additional CT scan would be useful for OR preparation, the surgeon will be suboptimally informed in one out of 10 patients but in only one out of 8 cases the patient will be exposed to radiation unnecessarily. They should also take into account the extra costs, which are about € 250 per scan in the Netherlands, of ordering additional CT scans in all cases.

Strengths and limitations

This study has several strengths including the use of highly experienced observers. These observers are typical surgeons who would make such decisions in hospitals, thereby improving generalizability. Also, the order of images was randomized and the time in between scoring moments was adequate to avoid bias due to memory. Surgeons were not informed we were testing two decisions models and were blinded to the study hypotheses. There is a potential risk of clustering, however since the intraobserver agreement of classifying fractures and for treatment planning is known to be fair to moderate, we are justified in combining results of the 4 comparisons. Additionally, the relatively low confidence intevals of NNS and NNH shows that the results are consistent across comparison groups. We used a database so there is a possibility that in some less severe intra-articular displaced DRF, an additional CT scan was not ordered. This could introduce selection bias. The chance that a CT scan would have changed treatment plan in these cases is very low. However, the protocol in the recruiting hospital is to always order a CT scan for operatively treated cases, so this limitation would not have affected the results for OR preparation.

Although previous literature showed that CT scans are more reliable than X-rays quantifying articular surface incongruencies (7-10,13). To the best of our knowledge no previous studies have reported whether the usefulness of an additional CT scan is predictable.

Future research

The outcomes of the current study are not necessarily related to better patient outcomes. Prospective randomized studies - comparing ordering CT scans for all patients with intra-articular displaced DRF to CT scanning on request of the surgeon- should be conducted to confirm the results of this study and to follow patients to determine which decision making model improves patient outcomes. Also a cost effectiveness analysis should be conducted as national care budgets are limited.

CONCLUSION

Surgeons are able to predict the usefulness of an additional CT scan for intra-articular displaced DRF regarding whether one treats the patient operatively or non-operatively. We recommend letting the surgeon decide which patients require an additional CT scan for treatment planning. However, for pre-operative planning the usefulness is much harder to predict, therefore we cannot give a strong recommendation for or against CT scanning of all patients with a displaced intra articular DRF.

REFERENCES

(1) Mendelson RM, Bairstow PJ. Inappropriate imaging: why it matters, why it happens, what can be done. J Med Imaging Radiat Oncol 2010 Jun;54(3):173-177.

(2) Amis ES,Jr, Butler PF, American College of Radiology. ACR white paper on radiation dose in medicine: three years later. J Am Coll Radiol 2010 Nov;7(11):865-870.

(3) Korley FK, Pham JC, Kirsch TD. Use of advanced radiology during visits to US emergency departments for injury-related conditions, 1998-2007. JAMA 2010 Oct 6;304(13):1465-1471.

(4) Larson DB, Johnson LW, Schnell BM, Goske MJ, Salisbury SR, Forman HP. Rising use of CT in child visits to the emergency department in the United States, 1995-2008. Radiology 2011 Jun;259(3):793-801.

(5) das Gracas Nascimento V, da Costa AC, Falcochio DF, Lanzarin LD, Checchia SL, Chakkour I. Computed tomography's influence on the classifications and treatment of the distal radius fractures. Hand (N Y) 2015 Dec;10(4):663-669.

(6) Cole RJ, Bindra RR, Evanoff BA, Gilula LA, Yamaguchi K, Gelberman RH. Radiographic evaluation of osseous displacement following intra-articular fractures of the distal radius: reliability of plain radiography versus computed tomography. J Hand Surg Am 1997 Sep;22(5):792-800.

(7) Dahlen HC, Franck WM, Sabauri G, Amlang M, Zwipp H. Incorrect classification of extra-articular distal radius fractures by conventional X-rays. Comparison between biplanar radiologic diagnostics and CT assessment of fracture morphology. Unfallchirurg 2004 Jun;107(6):491-498.

(8) Harness NG, Ring D, Zurakowski D, Harris GJ, Jupiter JB. The influence of three-dimensional computed tomography reconstructions on the characterization and treatment of distal radial fractures. J Bone Joint Surg Am 2006 Jun;88(6):1315-1323.

(9) Johnston GH, Friedman L, Kriegler JC. Computerized tomographic evaluation of acute distal radial fractures. J Hand Surg Am 1992 Jul;17(4):738-744.

(10) Pruitt DL, Gilula LA, Manske PR, Vannier MW. Computed tomography scanning with image reconstruction in evaluation of distal radius fractures. J Hand Surg Am 1994 Sep;19(5):720-727.

(11) Hunt JJ, Lumsdaine W, Attia J, Balogh ZJ. AO type-C distal radius fractures: the influence of computed tomography on surgeon's decision-making. ANZ J Surg 2013 Sep;83(9):676-678.

(12) Murray J, Gross L. Treatment of distal radius fractures. J Am Acad Orthop Surg 2013 Aug;21(8):502-505.

(13) Cole RJ, Bindra RR, Evanoff BA, Gilula LA, Yamaguchi K, Gelberman RH. Radiographic evaluation of osseous displacement following intra-articular fractures of the distal radius: reliability of plain radiography versus computed tomography. J Hand Surg Am 1997 Sep;22(5):792-800.

CHAPTER 3

Are validated outcome measures used in distal radius fractures truly valid? A critical assessment using COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist

Y.V. Kleinlugtenbelt
R.W. Nienhuis
M. Bhandari
J.C. Goslings
R.W. Poolman
V.A.B. Scholtes

Published in Bone and Joint Research, 2016 Apr;5(4):153-61.

ABSTRACT

Background

Patient Reported Outcome Measures (PROMs) are often used to evaluate the outcome of treatment in patients with distal radius fractures. Which PROM to select is often based on assessment of the measurement properties, such as good validity and reliability. Measurement properties are assessed in clinimetric studies, and results are often reviewed without considering the methodological quality of these studies. Our aim was to systematically review the methodological quality of clinimetric studies that evaluated measurement properties of PROMs used in patients with distal radius fractures, and to make recommendations for the selection of PROMs based on the level of evidence of each individual measurement property.

Methods

A systematic literature search was performed in Pubmed, Embase, Cinahl and PsycInfo databases to identify relevant clinimetric studies. Two reviewers independently assessed the methodological quality of the studies on measurement properties, using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. Level of evidence (strong / moderate / limited / lacking) for each measurement property per PROM was determined by combining the methodological quality and the results of the different clinimetric studies.

Results

19 out of 1508 unique studies were included, in which 12 PROMs were rated. The Patient-Rated Wrist Evaluation (PRWE) and the Disabilities of Arm, Shoulder and Hand questionnaire (DASH) were evaluated on most measurement properties. The evidence for the PRWE is moderate that its reliability, validity (content and hypothesis testing), and responsiveness are good. The evidence is limited that its internal consistency and cross-cultural validity is good, and its measurement error is acceptable. There is no evidence for its structural and criterion validity . The evidence for the DASH is moderate that its responsiveness is good. The evidence is limited that its reliability and the validity on hypotheses testing is good. There is no evidence for the other measurement properties.

Discussion

According to this systematic review there is at best moderate evidence that the responsiveness of the PRWE and DASH are good, as is the reliability and validity of the PRWE. We recommend these PROMs in clinical studies in patients with distal radius fractures, however, more clinimetric studies of higher methodological quality are needed to adequately determine its other measurement properties.

BACKGROUND

Distal radius fractures account for approximately 17 % of all fractures(1). Of all fractures in the upper extremity, the distal radius is the most common fracture site (2-4). Despite its high incidence, there is no treatment consensus (5). To conduct best evidence clinical trials in DRF treatment, and to properly compare trial results, there must be a consensus on the use of outcome measures. Historically, outcome assessment after distal radius fractures focused on imaging and physical examination (e.g. grip strength and range of motion). These assessments, however, do not represent the patients' perspective as they do not take the patients' feelings/opinion or wellbeing into account, which might be more important for the patient (6).

In the last two decades, the outcome assessment has shifted towards a patient-centered approach. This approach assesses the outcome directly from the opinion of the patient. Outcomes such as pain and functional ability, which are highly relevant for patients, can for instance be assessed by Patient Reported Outcome Measures (PROMs) (7).

Currently, a wide variety of PROMs are available and used to assess patient reported functional outcomes for upper limb and wrist disorders (8-20). Several (non-) systematic studies reviewed the existing literature in order to present available PROMs for assessing wrist and hand function in general. (21-25) Over a period of 25 years, two PROMs were most extensively used for evaluating the treatment outcome of patients with distal radius fractures(26). These were the Disabilities of Arm, Shoulder and Hand (DASH), and the (original or modified) Gartland and Werley scoring system. However, the Patient Related Wrist Evaluation (PRWE) was found to have the best measurement properties, e.g. is found to be the most reliable, valid and responsive instrument for these patients. This conclusion was drawn based on the results of the available clinimetric studies(26). Clinimetrics is a scientific discipline that aims to develop methods of assessing the properties of health measurement instrument, with the aim to improve the quality of outcome measures. And although the measurement properties were found to be good, the author did not incorporate the methodological quality of these clinimetric studies.

It is important for the understanding of this systematic review to distinct between the "methodological quality" of clinimetric studies on PROMs and the "quality" (e.g. the measurement properties) of the PROMs itself. Evidently, a PROM is only as good as the methodological quality of its study. In order to assess the methodological quality of clinimetric studies (i.e. studies on measurement properties) of PROMs, the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) group formulated a set of guidelines. First, the COSMIN group reached consensus on terminology, definitions and a taxonomy of measurement properties of PROMs in an international Delphi study. Then, they developed a checklist containing standards for evaluating the methodological quality of studies on the measurement properties (e.g. reliability) of measurement instruments (e.g. DASH). (27) The best PROM should have high level of evidence (e.g. as evaluated in high quality studies) supporting good quality on all measurement properties. The definitions and a "jargon free" description of the measurement properties are given table 1.

The aim this systematic review was to evaluate the methodological quality (using the COSMIN checklist) of the clinimetric studies that evaluated measurement properties of the available PROMs used in patients with distal radius fractures, and to make recommendations for the selection of PROMs based on the level of evidence of each individual measurement property. The results of this study might help us to determine which PROM is most appropriate for the evaluation of patients with distal radius fractures.

METHODS

Literature search

We performed a literature search on November 13th, 2015 to identify all published studies on the measurement properties of PROM's in the evaluation of treatment of distal radius fractures. The following databases were searched with specific index terms and derivatives of these terms: Pubmed (1990 to 2015), EMbase (1990 to 2015), CINAHL (1990 to 2015), and PsycINFO (1990 to 2015). In PubMed we used a validated search filter for finding studies on measurement properties.(28) We also added the names of all PROM's that are described for wrist disorders.(29) The full search strategy is provided in Appendix 1. We restricted our search to studies published in English, German and Dutch because both reviewers are fluent in these languages. Reference lists were hand searched to identify additional relevant studies.

Selection Criteria

Two reviewers (YK, RN) independently assessed all titles and abstracts. We included studies with a description of the measurement properties of PROM's used in patients with a distal radius fracture. When in doubt about the applicability of a study, the full text article was retrieved and screened for eligibility. Afterwards, the researchers discussed their assessments and consensus was reached. In cases of where consensus couldn't be obtained a third reviewer (VS), was employed to achieve consensus.

Assessment of the Quality of the Studies

The same two reviewers independently rated the methodological quality of the studies using the COSMIN checklist (www.cosmin.nl).(30).

The COSMIN checklist consists of 11 separate checklists, called "boxes". In nine boxes the quality nine measurement properties is addressed : A: internal consistency, B: reliability, C: measurement error, D: content validity, E: structural validity, F: hypotheses testing, G: cross-cultural validity, H: criterion validity and I: responsiveness. The last box "I: interpretability" is no measurement property, but nevertheless a meaningful requirement for the applicability of PROMs in research. The generalisability of the results are determined with a final box. The definitions of the measurement properties and interpretability are given table 1.

In each box, the methodological quality can be evaluated based on a variety of items addressing adequate study design and statistical analysis. Each question in any box must be rated as 'excellent', 'good', 'fair', 'poor' or 'not applicable'. Scoring is then performed using the criteria set by the COSMIN group. To obtain a total score for the methodological quality of one of the boxes "the worst score counts" algorithm was applied as set out by the COSMIN guidelines.(31) Meaning, the methodological quality of that measurement property was only rated 'excellent' if all relevant questions pertaining to that box (e.g. measurement property) were scored as excellent. In all boxes, a small sample size was considered poor methodological quality. As a rule of thumb, a sample size of ≥100 received a rating of 'excellent', 50-100 received 'good', 30-50 was rated 'fair', and less than 30 was rated as poor.(31)

Table 1 Definitions of the measurement properties. In *italic* the jargon free description.

Internal consistency	The degree of the interrelatedness among the items
	“Does different questions of a PROM that suppose to measure the same general construct produce similar scores?”
Reliability	The proportion of the total variance in the measurements which is because of “true” differences among patients
	“How close are repeated measurements?”
Measurement error	The systematic and random error of a patient’s score that is not attributed to true changes in the construct to be measured
	“What amount of change in a score can not be considered a real or true change?”
Content validity	The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured
	“Are all items relevant for the specific population and are important activities missed?”
Structural validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured
	“Do all items in a PROM reflect a single or multiple constructs?”
Hypotheses testing	The degree to which the scores of an HR-PRO instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validly measures the construct to be measured
	“What is the expected relationship with other PROMs assessing comparable constructs?”
Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
	“Is the PROM correctly translated en retested in another language and cultural setting?”
Criterion validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a “gold standard”
	“Is the PROM tested against the “gold standard” PROM?”
Responsiveness	The ability of an HR-PRO instrument to detect change over time in the construct to be measured
	“If patients improve or worsen over time does this change in the PROM accordingly”
Interpretability*	The degree to which one can assign qualitative meaning—that is, clinical or commonly understood connotations—to an instrument’s quantitative scores or change in scores.
	“What does the scores or change in scores of a PROM mean”

Level of evidence of the measurement properties per PROM

For each PROM, we determined the level of evidence by combining the results of the different studies for each measurement property, as described by Terwee et al.(31) These factors were taken into account: the number of studies (one or multiple), the methodological quality of the studies (excellent/good/fair/poor/not available), and consistency of the results (positive/negative). Based on these factors each measurement property per PROM could be ranked as strong, moderate, limited or conflicting evidence. Only when the methodological quality of the clinimetric study/studies was poor, the level of evidence was rated as unknown.

Source of Funding

No external funding was received for this study.

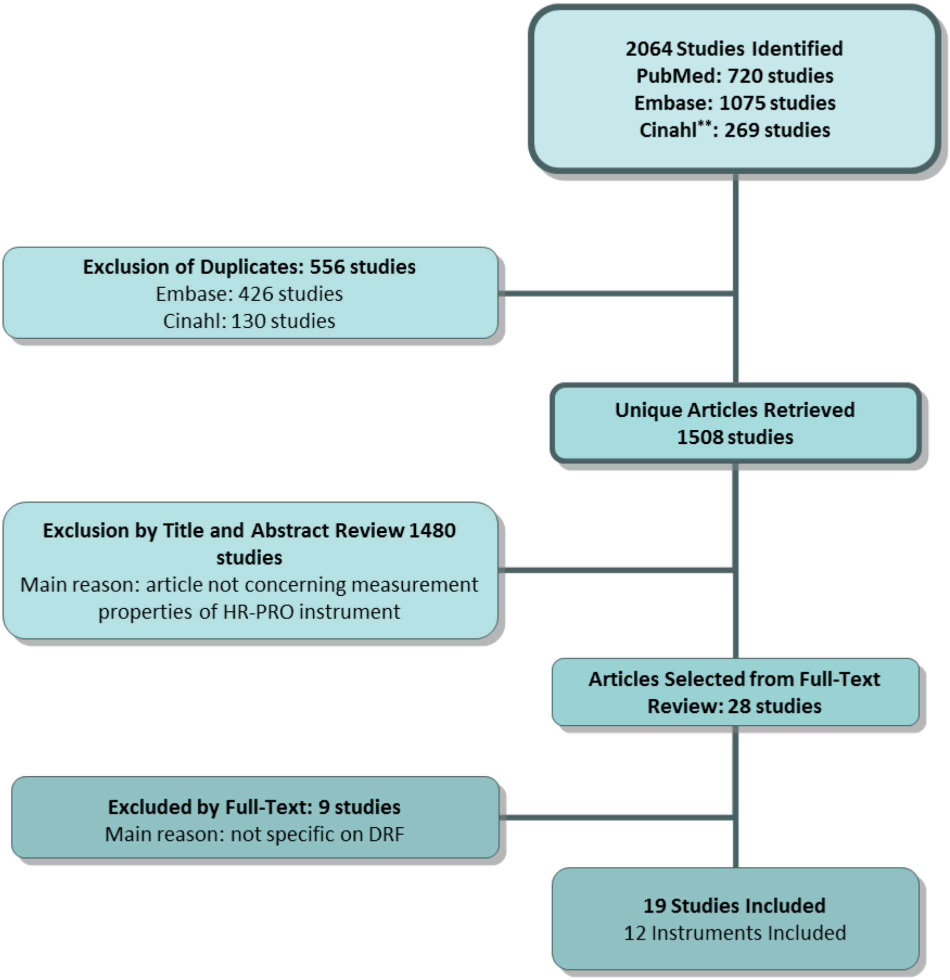

Figure 1; Search strategy and Selection of Articles *Nov 13, 2015; ** Cinahl search includes Psy-cInfo database

RESULTS

Included studies

A total of 2064 studies was retrieved by the electronic search, performed in Pubmed (n = 720), Embase (n = 1075) and Cinahl/ PsycINFO (n = 269) (Figure 1). After removing duplicates, 1508 unique studies were identified. The titles and abstracts were screened by two researchers independently, after which 27 studies were deemed potentially eligible. After retrieving and reading the full-text, 19 studies were included. Reference evaluation of these 19 articles did not yield any additional relevant studies.

Table 2 – Patient Related Outcome Instruments included in the review

Abbreviation	Full name	Original author
PRWE	Patient-Rated Wrist Evaluation	MacDermid (9)
DASH	Disabilities of Arm, Shoulder and Hand	Hudak (8)
MHQ	Michigan Hand Questionnaire	Chung (11)
SF-36	Short Form-36	Ware (12)
PEM	Patient Evaluation Measure	Macey (10)
AIMS2	Arthritis Impact Measurement Scale	Meenan (14)
BWH-CTQ	Brigham and Women's Hospital Carpal Tunnel Questionnaire	Levine (13)
IOF-WFQ	International Osteoporosis Foundation Wrist Fracture Questionnaire	Lips (16)
PFW	Patient Focused Wrist Outcome Instrument	Bialocerkowski (17)
TSK	Tampa Scale of Kinesophobia	Kori (18)
CAT	Catastrophizing Subscale of the Coping Strategies Questionnaire	Rosenstiel (19)
SES	Self-Efficacy Scale	Altmaier (20)

Overall results

In the 19 included studies, a total of 12 PROM's were evaluated (Table 2). In three papers, multiple PROM's were evaluated, three(32), three(33) and five(34) respectively. Most studies (80%) evaluated more than one measurement property. None of the studies evaluated structural validity. Criterion validity was also never evaluated. However, this was as expected, since there are no measurement instruments that can be used as a 'gold standard', which is a prerequisite of this measurement property. A complete overview of the study characteristics is shown in Table 3.

Of all PROMs, the PRWE has been studied most extensively, followed by the DASH. The eight studies evaluating the PRWE assessed almost all measurement properties: seven out of the nine (Table 4a).However, the methodological quality of these studies was overall low, varying from poor to fair for internal consistency, reliability, measurement error, cross cultural validity and responsiveness; and varying from poor to good for content validity and hypothesis testing. Interpretability was also assessed, but these studies were of poor methodological quality.

The four studies evaluating the DASH (32,34-36) assessed less than half of the measurement properties: four out of nine. The methodological quality of these studies was generally low, varying from consistently poor for internal consistency, poor to fair for reliability, to consistently fair for responsiveness. Measurement error, content validity, hypothesis testing, cross cultural validity and interpretability were never assessed (Table 4a).

Table 3 – Study characteristics * It can be deduced

Measurement instrument	Study	n	Mean age (range or SD)	Gender Male (%)	Country	Language
PRWE	Gabl(41)	133	62 (19-92)	27	Austria	German*
	Hemelaers(42)	44	56 (15)	36	Switzerland	German
	MacDermid(43)	36 / 101	45 (10) / 50 (16)	33 / 31	Canada	English*
	MacDermid(32)	59	53 (18)	37	Canada	English*
	Wilcke(35)	99	58 (18)	20	Sweden	Swedish
	Lovgren(34)	16	52 (12)	19	Sweden	Swedish
	Mehta(44)	50	46 (14)	56	India	Hindi
	Kim(45)	63	56 (19-83)	27	Rep. Korea	Korean
	Schonnemann(46)	60/29	55 (19-86)	27	Denmark	Danish
	Walenkamp(47)	102	59 (48-66)	30	Netherlands	Dutch
DASH	Macdermid(32)	59	53 (18)	37	Canada	English*
	Westphal(36)	107	59 (17-84)	27	Germany	German
	Westphal(48)	72	60 (16)	29	Germany	German
	Lovgren(34)	16	52 (12)	19	Sweden	Swedish
MHQ	Kotsis(49)	47 / 37	48 (17) / 51(16)	32 / 38	USA	English
	Shauver(50)	51	50 (19-83)	37	USA	English
	Waljee(51)	128	61 (9)	27	USA/UK	English*
SF-36	Amadio(33)	21	57 (14-84)	14	USA	English*
	MacDermid(32)	59	53 (18)	37	Canada	English*
PEM	Forward(52)	200	54 (24-80)	36	UK	English*
AIMS2	Amadio(33)	21	57 (14-84)	14	USA	English*
BWH-CTQ	Amadio(33)	21	57 (14-84)	14	USA	English*
IOF-WFQ	Lips(16)	105	63 (8)	12	UK/NL/Ita/Bel	English/Dutch/Italian*
PFW	Bialocerkowski(53)	26	62 (22-84)	15	Australia	English
TSK	Lovgren(34)	16	52 (12)	19	Sweden	Swedish
CAT	Lovgren(34)	16	52 (12)	19	Sweden	Swedish
SES	Lovgren(34)	16	52 (12)	19	Sweden	Swedish

Table 4a– Summary of methodological quality of the studies on measurement properties of the PRWE and DASH. A full overview of all the scores are shown in appendix 2

	PRWE (41)	PRWE (42)	PRWE (43)	PRWE (32)	PRWE (35)	PRWE (34)	PRWE (44)	PRWE (46)	PRWE (45)	PRWE (47)	DASH (32)	DASH (36)	DASH (48)	DASH (34)
Generalisability	**Fair**	**Fair**	**Fair**	**Poor**	**Fair**	**Excel**	**Poor**	**Fair**	**Fair**	**Fair**	**Poor**	**Fair**	**Good**	**Excel**
InternalConsistency	**Poor**	**Poor**			**Fair**	**Poor**	**Poor**	**Poor**	**Poor**	**Poor**		**Poor**	**Poor**	**Poor**
Reliability		**Fair**	**Poor**		**Fair**	**Poor**	**Fair**	**Poor**	**Fair**				**Fair**	**Poor**
Measurement Error									**Fair**	**Poor**				
Content validity			**Fair**					**Good**						
Structural validity														
Hypotheses testing		**Fair**			**Good**		**Fair**	**Fair**	**Poor**			**Fair**		
Cross cultural					**Fair**		**Poor**	**Poor**	**Poor**					
Criterion validity														
Responsiveness			**Fair**	**Fair**	**Fair**		**Fair**	**Fair**	**Poor**		**Fair**		**Fair**	
Interpretability					**Poor**				**Poor**		**Poor**	**Poor**		

3

Table 4b– Summary of methodological quality of the studies on measurement properties of the other measurement instruments. A full overview of all the scores are shown in appendix 2

	MHQ (49)	MHQ (50)	MHQ (51)	SF36 (32)	SF36 (33)	PEM (52)	IOF (16)	PFW (53)	AIMS2 (33)	BWH (33)	TSK (34)	CAT (34)	SES (34)
Generalisability	**Fair**	**Fair**	**Poor**	**Poor**	**Fair**	**Poor**	**Fair**	**Fair**	**Fair**	**Fair**	**Excel**	**Excel**	**Excel**
Internal Consistency						**Poor**	**Poor**				**Poor**	**Poor**	**Poor**
Reliability							**Poor**				**Poor**	**Poor**	**Poor**
Measurement Error													
Content validity													
Structural validity													
Hypotheses testing						**Poor**		**Poor**					
Cross cultural													
Criterion validity													
Responsiveness	**Fair**	**Fair**	**Fair**	**Fair**	**Poor**		**Fair**	**Poor**	**Poor**	**Poor**			
Interpretability		**Fair**											

Of the other ten PROMs, one to three measurement properties were assessed. These concerned mostly internal consistency, reliability and responsiveness. Overall, the methodological quality of these clinimetric studies was at best poor to fair (Tabel 4b). This is mainly due to the low sample size in the majority of these studies. Secondly, the high amount of items that were scored as "not applicable". Finally, the lack of description surrounding the statistical methods that were used also contributed to the poor rating. A full overview of all the ratings is shown in Appendix 2

Level of evidence of the measurement properties per PROM

The synthesis of results per PROM and their accompanying level of evidence are presented in Table 5.

The highest levels of evidence were found for the measurement properties of the PRWE. Nevertheless, the evidence is at best limited to moderate. For instance, reliability (assessed in 78% of the studies) ranged from 0.81-0.97 (Table 5). Three studies were of poor methodological quality, and four were of fair quality (Table 4). Therefore, the synthesis of these results is that there is moderate evidence supporting good reliability. There is also moderate evidence that the validity (content and hypothesis testing), and responsiveness are good. The evidence is limited that its internal consistency and cross-cultural validity is good, and its measurement error is acceptable. There is no evidence for its structural and criterion validity. The evidence for the DASH is moderate that its responsiveness is good. The evidence is limited that its reliability and the validity on hypotheses testing is good. There is no evidence for the other measurement properties The evidence for the other ten PROMs is mainly unknown, since the quality of the studies that evaluated some of its measurement properties (mainly internal consistency, reliability and/or responsiveness) were mainly of poor methodological quality.

Table 5 – Ratings of measurement properties and interpretability of measurement instruments with level of evidence

	PRWE (32,34,35,41-47)	**DASH** (32,34,36,48)	**MHQ** (49-51)	**SF-36** (32,33)	**PEM** (52)	**AIMS2** (33)	**BWH** (33)	**IOF** (33)	**PFW** (53)	**TSK** (34)	**CAT** (34)	**SES** (34)
RELIABILITY												
Internal consistency	+	?			?			?		?	?	?
Cronbach's alpha	0,89-0,97	0,93-0,98			0,94			0,96		0,68-0,82	0,88-0,97	0,79-0,95
Reliability	++	+						?		?	?	?
ICC	0,81-0,97	0,78-0,95						na		0,81-0,84	0,85-0,89	0,57-0,86
Measurement error	+											
SDC	4.4-11.0											
VALIDITY												
Content validity	++											
Structural validity												
Hypotheses testing	++	+			?				+			
Comparator instrument	DASH	Gartland			na				na			
Cross-cultural	+											
Criterion validity												
RESPONSIVENESS												
Responsiveness	++	++	++	+		?	?	+	?			
SRM	na	Na	na	na		na	na	na	na			
INTERPRETABILITY												
Interpretability	?		-									
MIC	11.5											

+++ or – – – multiple studies of good quality OR 1 study of excellent quality: strong evidence positive/negative result
++ or – – multiple studies of fair quality OR 1 study of good quality: moderate evidence positive/negative result
+ or – 1 study of fair quality: limited evidence positive/negative result
+/ – conflicting findings
? only studies of poor quality: unknown , due to poor methodological quality
na not available (not performed or described)

DISCUSSION

The aim this systematic review was to evaluate the methodological quality of the clinimetric studies that evaluated measurement properties of the available PROMs used in patients with distal radius fractures, and to make recommendations for the selection of PROMs based on the level of evidence of each individual measurement property

Key findings

The two PROMs that were most extensively evaluated were the PRWE (with 7 out of 9 measurement properties investigated) and the DASH (with 4 out of 9 investigated). The methodological quality of these studies ranged at best from poor to good. So, after synthesis of the scores and incorporating the levels of evidence, the quality of these two PROMS is not supported with strong levels of evidence on any of the measurement properties. For the PRWE, there is at best moderate evidence supporting a good reliability, content validity, hypotheses testing and responsiveness. The evidence is only limited that the measurement error is acceptable and the cross-cultural validity and internal consistency are good. Structural validity and criterion validity were never evaluated, so these lack in evidence. The evidence for interpretability, which is not a measurement property, is unknown, since this was only evaluated in three studies with poor methodological quality. The DASH showed at best moderate evidence for good responsiveness and limited evidence for good hypotheses testing and reliability. All other measurement properties were found to be lacking in evidence.

These findings do not mean that these and other PROMs have bad measurement properties, an thus are of poor quality. Since we found that overall, the measurement properties were good, but the methodological quality of these studies was overall low, it does mean that these results may be biased. Therefore, the results of our review do imply that studies of higher methodological quality are needed to properly assess their measurement properties. For instance, many PROMs are translated into multiple languages. The PRWE has been correctly translated in 14 languages, following the translation process described by Beaton et al.(37) Nevertheless we only found cross cultural validity studies for the Swedish, Hindi, Korean and Danish version, because the other translated versions were not adequately evaluated on their cross cultural validity. However, our search was limited to English, German and Dutch, so it can be assumed that the cross-cultural validity was evaluated but the results were not published in any of these languages.

Comparison of results with previous literature

Previous reviews described a variety of PROMS measuring wrist and/or hand disorders in general, but not PROMs specific to distal radius fractures. Goldhahn et al(25) advise using a combination of a disease specific PROM (PRWE), an extremity specific PROM (DASH) and a generic PROM (SF-36). Changulani et al(22) compared the measurement properties of four PROMs for wrist and hand disorders. They concluded that the PRWE is the most responsive instrument for evaluating outcomes in patients with a distal radius fracture. These conclusions were drawn before the COSMIN checklist was available. Since the methodological quality of the clinimetric studies was not taken into account, and these results might therefore be biased. Especially since in the current review we found that the methodological quality of these studies was at best fair. Therefore, we can only conclude that both this good responsiveness of the DASH and PRWE is supported by moderate evidence.

Hoang-Kim et al.(21) assessed the quality of reviews published on currently used PROM's for assessing function of the hand and wrist joints. Although they used COSMIN's taxonomy, terminology and definitions to define the different measurement properties, they did not systematically review the methodological quality of these studies. Nevertheless they concluded that the PRWE has good construct validity and responsiveness, and found this to be only slightly better than the DASH for assessing patients with wrist injuries. Based on the results of our review we agree that the PRWE is slightly better investigated than the DASH, but disagree with their rating of "good" on some measurement properties. This difference may occur because we did incorporate the methodological quality of these studies by using the COSMIN checklist instead of only using the COSMIN taxonomy.

Study strengths

To our knowledge, this is the first study that has used the COSMIN checklist to systematically review the methodological quality of studies on the measurement properties of PROMs in the evaluation of treatment of distal radius fractures. Furthermore, the quality of each study was assessed by two independent reviewers, as recommended by the COSMIN group and a third reviewer in cases of disagreement. Using these methods, we were able to minimize subjective judgment having an influence on the outcome. We searched for relevant articles from 1990 onwards, so consider it unlikely that any relevant PROMs were missed. This is especially true since most PROMs were developed after 1990. Since we found 19 studies eligible from a possible 1508, this shows that our search strategy was very broad and inclusive. Yet, it also demonstrates that the literature on this topic is somewhat lacking. Our search was not just limited to the English language, as both reviewers have good knowledge of the German and Dutch language as well.

Study weaknesses

There were some limitations to this review. As in all reviews, publication bias from unpublished studies may threaten the internal validity as unpublished studies are more likely to report negative or unfavourable results. Another limitation of this study was that it was not always clear to the reviewers if specific methodological aspects were not reported or not performed, making it impossible to distinguish between poor study reporting and poor methodological quality. We did not contact the authors of the studies to clarify these issues. It can be assumed that some studies have been executed properly but are not sufficiently well described according to the COSMIN criteria. This may have affected the quality ratings.

The shortcomings of outcome measurement research in distal radius fractures exposed by this review can not just be generalized to all clinimetric research in orthopedic surgery. However, it is known that strong evidence supporting good quality of multiple PROMs for various pathology is lacking.(38-40) So we advise the reader to be cautious when choosing a PROM based on the results of clinimetric studies without considering the methodological quality of these studies.

For future research, we believe that it is especially important to further evaluate the measurement properties and interpretability of the PRWE and DASH outcome measures in higher quality studies. Based on the results of the available clinimetric studies there is no evidence that these PROMs are not useful to evaluate the treatment of distal radius fractures. So we do not see a reason to develop a new instrument for the evaluation of treatment of distal radius fractures. Therefore, for now, based on the best available evidence, we recommend to use the PRWE or DASH to evaluate the outcome of treatment in patients with distal radius fractures. But we cannot stress strongly enough that more clinimetric studies of higher methodological quality are needed to select the PROMs more carefully.

CONCLUSION

According to this systematic review, strong evidence supporting good quality of any of the current available PROMs in patients with distal radius fractures is lacking. The evidence that the responsiveness of the PRWE and DASH are good is moderate, as is the evidence for good validity and reliability of the PRWE. We therefore recommend these PROMs in clinical studies in patients with distal radius fractures, however, more clinimetric studies of higher methodological quality are needed to adequately determine its other measurement properties. If the methodological quality of clinimetric studies continues to increase, PROMs can be selected more carefully.

REFERENCES

(1) Court-Brown CM, Caesar B. Epidemiology of adult fractures: A review. Injury 2006 Aug;37(8):691-697.

(2) ALFFRAM PA, BAUER GC. Epidemiology of fractures of the forearm. A biomechanical investigation of bone strength. J Bone Joint Surg Am 1962 Jan;44-A:105-114.

(3) Doczi J, Renner A. Epidemiology of distal radius fractures in Budapest. A retrospective study of 2,241 cases in 1989. Acta Orthop Scand 1994 Aug;65(4):432-433.

(4) Owen RA, Melton LJ,3rd, Johnson KA, Ilstrup DM, Riggs BL. Incidence of Colles' fracture in a North American community. Am J Public Health 1982 Jun;72(6):605-607.

(5) Wei DH, Poolman RW, Bhandari M, Wolfe VM, Rosenwasser MP. External fixation versus internal fixation for unstable distal radius fractures: a systematic review and meta-analysis of comparative clinical trials. J Orthop Trauma 2012 Jul;26(7):386-394.

(6) Darzi L. High Quality Care For All – NHS Next Stage Review Final Report. 2008.

(7) Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2(14):i-iv, 1-74.

(8) Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). Am J Ind Med 1996 Jun;29(6):602-608.

(9) MacDermid JC. Development of a scale for patient rating of wrist pain and disability. J Hand Ther 1996 Apr-Jun;9(2):178-183.

(10) Macey AC, Burke FD, Abbott K, Barton NJ, Bradbury E, Bradley A, et al. Outcomes of hand surgery. British Society for Surgery of the Hand. J Hand Surg Br 1995 Dec;20(6):841-855.

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

(12) Ware JE,Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483.

(13) Levine DW, Simmons BP, Koris MJ, Daltroy LH, Hohl GG, Fossel AH, 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 Nov;75(11):1585-1592.

(14) Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. Arthritis Rheum 1992 Jan;35(1):1-10.

(15) Smith HB. Smith hand function evaluation. Am J Occup Ther 1973 Jul-Aug;27(5):244-251.

(16) Lips P, Cooper C, Agnusdei D, Caulin F, Egger P, Johnell O, et al. Quality of life in patients with vertebral fractures: validation of the Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO). Working Party for Quality of Life of the European Foundation for Osteoporosis. Osteoporos Int 1999;10(2):150-160.

(17) Bialocerkowski AE, Grimmer KA, Bain GI. Development of a patient-focused wrist outcome instrument. Hand Clin 2003 Aug;19(3):437-48, ix.

(18) Kori S, Miller R, Todd D. Kinesiophobia: a new view of chronic pain behavior. Pain managment 1990:35-43.

(19) Rosenstiel AK, Keefe FJ. The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. Pain 1983 Sep;17(1):33-44.

(20) Altmaier EM, Russell DW, Kao CF, Lehmann TR, Weinstein JN. Role of self-efficacy in rehabilitation outcome among chronic low back pain patients. Journal of Counseling Psychology 1993;3:335-391.

3

(21) Hoang-Kim A, Pegreffi F, Moroni A, Ladd A. Measuring wrist and hand function: common scales and checklists. Injury 2011 Mar;42(3):253-258.

(22) Changulani M, Okonkwo U, Keswani T, Kalairajah Y. Outcome evaluation measures for wrist and hand: which one to choose? Int Orthop 2008 Feb;32(1):1-6.

(23) Bialocerkowski AE, Grimmer KA, Bain GI. A systematic review of the content and quality of wrist outcome instruments. Int J Qual Health Care 2000 Apr;12(2):149-157.

(24) Schuind FA, Mouraux D, Robert C, Brassinne E, Remy P, Salvia P, et al. Functional and outcome evaluation of the hand and wrist. Hand Clin 2003 Aug;19(3):361-369.

(25) Goldhahn J, Angst F, Simmen BR. What counts: outcome assessment after distal radius fractures in aged patients. J Orthop Trauma 2008 Sep;22(8 Suppl):S126-30.

(26) Goldhahn J, Beaton D, Ladd A, Macdermid J, Hoang-Kim A. Recommendation for measuring clinical outcome in distal radius fractures: a core set of domains for standardized reporting in clinical practice and research. Arch Orthop Trauma Surg 2014 Feb;134(2):197-205.

(27) Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010 May;19(4):539-549.

(28) Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009 Oct;18(8):1115-1123.

(29) Suk M, Hanson B, Norvell D, Helfet D. Musculoskeletal Outcomes Measures and Instruments, AO Handbook. first edition ed.: Thieme; 2005.

(30) Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-745.

(31) Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012 May;21(4):651-657.

(32) MacDermid JC, Richards RS, Donner A, Bellamy N, Roth JH. Responsiveness of the short form-36, disability of the arm, shoulder, and hand questionnaire, patient-rated wrist evaluation, and physical impairment measurements in evaluating recovery after a distal radius fracture. J Hand Surg Am 2000 Mar;25(2):330-340.

(33) Amadio PC, Silverstein MD, Ilstrup DM, Schleck CD, Jensen LM. Outcome after Colles fracture: the relative responsiveness of three questionnaires and physical examination measures. J Hand Surg Am 1996 Sep;21(5):781-787.

(34) Lovgren A, Hellstrom K. Reliability and validity of measurement and associations between disability and behavioural factors in patients with Colles' fracture. Physiother Theory Pract 2012 Apr;28(3):188-197.

(35) Wilcke MT, Abbaszadegan H, Adolphson PY. Evaluation of a Swedish version of the patient-rated wrist evaluation outcome questionnaire: good responsiveness, validity, and reliability, in 99 patients recovering from a fracture of the distal radius. Scand J Plast Reconstr Surg Hand Surg 2009;43(2):94-101.

(36) Westphal T, Piatek S, Schubert S, Schuschke T, Winckler S. Reliability and validity of the upper limb DASH questionnaire in patients with distal radius fractures. Z Orthop Ihre Grenzgeb 2002 Jul-Aug;140(4):447-451.

(37) Goldhahn J, Shisha T, Macdermid JC, Goldhahn S. Multilingual cross-cultural adaptation of the patient-rated wrist evaluation (PRWE) into Czech, French, Hungarian, Italian, Por-

tuguese (Brazil), Russian and Ukrainian. Arch Orthop Trauma Surg 2013 May;133(5):589-593.

(38) Green A, Liles C, Rushton A, Kyte DG. Measurement properties of patient-reported outcome measures (PROMS) in Patellofemoral Pain Syndrome: a systematic review. Man Ther 2014 Dec;19(6):517-526.

(39) Grevnerts HT, Terwee CB, Kvist J. The measurement properties of the IKDC-subjective knee form. Knee Surg Sports Traumatol Arthrosc 2014 Sep 6.

(40) Kroman SL, Roos EM, Bennell KL, Hinman RS, Dobson F. Measurement properties of performance-based outcome measures to assess physical function in young and middle-aged people known to be at high risk of hip and/or knee osteoarthritis: a systematic review. Osteoarthritis Cartilage 2014 Jan;22(1):26-39.

(41) Gabl M, Krappinger D, Arora R, Zimmermann R, Angermann P, Pechlaner S. Acceptance of patient-related evaluation of wrist function following distal radius fracture (DRF). Handchir Mikrochir Plast Chir 2007 Feb;39(1):68-72.

(42) Hemelaers L, Angst F, Drerup S, Simmen BR, Wood-Dauphinee S. Reliability and validity of the German version of "the Patient-rated Wrist Evaluation (PRWE)" as an outcome measure of wrist pain and disability in patients with acute distal radius fractures. J Hand Ther 2008 Oct-Dec;21(4):366-376.

(43) 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 Nov-Dec;12(8):577-586.

(44) Mehta SP, Mhatre B, MacDermid JC, Mehta A. Cross-cultural adaptation and psychometric testing of the Hindi version of the patient-rated wrist evaluation. J Hand Ther 2012 Jan-Mar;25(1):65-77; quiz 78.

(45) Kim JK, Kang JS. Evaluation of the Korean version of the patient-rated wrist evaluation. J Hand Ther 2013 Jul-Sep;26(3):238-43; quiz 244.

(46) Schonnemann JO, Hansen TB, Soballe K. Translation and validation of the Danish version of the Patient Rated Wrist Evaluation questionnaire. J Plast Surg Hand Surg 2013 Dec;47(6):489-492.

(47) Walenkamp MM, Vos LM, de Muinck Keizer RJ, Rosenwasser MP, Goslings JC, Schep NW. The Minimum Clinically Important Difference of the Patient-rated Wrist Evaluation Score for Patients With Distal Radius Fractures. Clin Orthop Relat Res 2015 Oct;473(10):3235-3241.

(48) Westphal T. Reliability and responsiveness of the German version of the Disabilities of the Arm, Shoulder and Hand questionnaire (DASH). Unfallchirurg 2007 Jun;110(6):548-552.

(49) Kotsis SV, Lau FH, Chung KC. Responsiveness of the Michigan Hand Outcomes Questionnaire and physical measurements in outcome studies of distal radius fracture treatment. J Hand Surg Am 2007 Jan;32(1):84-90.

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

(51) Waljee JF, Kim HM, Burns PB, Chung KC. Development of a brief, 12-item version of the Michigan Hand Questionnaire. Plast Reconstr Surg 2011 Jul;128(1):208-220.

(52) Forward DP, Sithole JS, Davis TR. The internal consistency and validity of the Patient Evaluation Measure for outcomes assessment in distal radius fractures. J Hand Surg Eur Vol 2007 Jun;32(3):262-267.

(53) Bialocerkowski AE, Grimmer KA, Bain GI. Validity of the patient-focused wrist outcome instrument: do impairments represent functional ability? Hand Clin 2003 Aug;19(3):449-55, ix.

APPENDIX 1

Search PUBMED systematic review distale radius#
(#1 AND #2 AND #3 NOT #4) AND (english[Language] OR German[Language] OR Dutch[Language]) AND ("1990"[Publication Date] : "3000"[Publication Date])

#1
activities of daily living[MeSH Terms] OR "activities of daily living"[Text Word] OR "activities of daily life"[Text Word] OR "physical activity"[Text Word] OR "physical function "[Text Word] OR "functional ability"[Text Word] OR "everyday functioning"[Text Word] OR "functional status"[Text Word] OR "function"[Text Word] OR "physical impairment"[Text Word] OR "after hand therapy"[Text Word] OR "Boston questionnaire" [Text Word] OR "Brigham and Women's carpal tunnel questionnaire"[Text Word] OR "DASH"[Text Word] OR "disabilities of arm, shoulder and hand"[Text Word] OR "quick DASH"[Text Word] OR "forearm symptom severity scale"[Text Word] OR "functional index"[Text Word] OR "Gartland and Werley scoring system"[Text Word] OR "Green and O'Brien"[Text Word] OR "Michigan Hand outcomes Questionnaire"[Text Word] OR "MHQ"[Text Word] OR "New York Orthopedic Hospital wrist rating system"[Text Word] OR "patient focused wrist outcome"[Text Word] OR "patient outcomes of surgery-hand/arm"[Text Word] OR "POS-hand/arm"[Text Word] OR "patient rated wrist evaluation"[Text Word] OR " PRWE"[Text Word] OR "sequential occupational dexterity assessment"[Text Word] OR "SODA"[Text Word]

#2
radius fractures[MeSH Terms] OR colles' fracture[MeSH Terms] OR "radius fractures"[Text Word] OR "colles's fracture"[Text Word] OR "colles' fracture"[Text Word] OR "colles fracture"[Text Word] OR "distal radius fracture"[Text Word] OR "wrist fracture"[Text Word] OR "antebrachial fracture"[Text Word] OR "distal radial fracture"[Text Word] OR "radial fracture"[Text Word] OR "forearm fracture"[Text Word] OR "fore-arm fracture"[Text Word] OR "distal forearm fracture"[Text Word] OR "smith fracture"[Text Word] OR "smith's fracture"[Text Word] OR "smiths fracture"[Text Word] OR "barton's fracture"[Text Word] OR "barton fracture"[Text Word] OR "bartons fracture"[Text Word] OR "chauffeur fracture"[Text Word] OR "chauffeurs fracture"[Text Word] OR "chauffeur's fracture"[Text Word]

#3
(instrumentation[sh] OR methods[sh] OR "Validation Studies"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR "outcome measure*"[tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR "coefficient of variation"[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab]

OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "inters-cale correlation*"[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR "in-terval variability"[tiab] OR "rate variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR "minimal detectable concentration"[tiab] OR interpretab*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adap-tive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab])

#4 (NOT)
("addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient educati-on handout"[Publication Type] OR "popular works"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development con-ference, nih"[Publication Type] OR "practice guideline"[Publication Type]) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms])
Search Embase systematic review distale radius# (597 hits)
Limits: Publication Date (1990-present)
Language: (English, Dutch, German)

#1
exp recreation/ OR exp daily life activity/ OR exp physical capacity/ OR activities of daily living OR activities of daily life OR physical activity OR physical function OR functional ability OR everyday functioning OR functional status OR function OR physical impairment OR after hand therapy OR Boston questionnaire OR Brigham Womens carpal tunnel questionnaire OR DASH OR quick DASH OR forearm symptom severity scale OR functional index OR Gartland Werley scoring system OR Green Brien OR Michigan Hand outcomes Questionnaire OR MHQ OR New York Orthopedic Hos-pital wrist rating system OR patient focused wrist outcome OR patient outcomes of surgery-hand/arm OR POS-hand/arm OR patient rated wrist evaluation OR PRWE OR sequential occupational dexterity assessment OR SODA

#2
exp radius fracture/ OR exp colles fracture/ OR exp forearm fracture/ OR exp wrist fracture/ OR radius fracture* OR colles fracture* OR distal radius fracture OR wrist fracture* OR antebrachial fracture* OR radial fracture* OR forearm fracture* OR smith fracture* OR smiths fracture* OR barton fracture* OR bartons fracture* OR chauffeur fracture* OR chauffeurs fracture*

#3

(instrumentation OR methods OR Validation Studies OR Comparative Study OR psychometrics OR psychometr* OR clinimetr* OR clinometr* OR outcome assessment OR “outcome assessment” OR outcome measure* OR “observer variation” OR observer variation OR “Health Status Indicators” OR “reproducibility of results” OR reproducib* OR “discriminant analysis” OR reliab* OR unreliab* OR valid* OR coefficient OR homogeneity OR homogeneous OR “internal consistency” OR (cronbach* AND (alpha OR alphas)) OR (item AND (correlation* OR selection* OR reduction*)) OR agreement OR precision OR imprecision OR “precise values” OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intraobserver OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR “known group” OR factor analysis OR factor analyses OR dimension* OR subscale* OR (multitrait AND scaling AND (analysis OR analyses)) OR item discriminant OR interscale correlation* OR error OR errors OR “individual variability” OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR “standard error of measurement” OR sensitiv* OR responsive* OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (changeOR difference)) OR (small* AND (real OR detectable) AND (change OR difference)) OR meaningful change OR “ceiling effect” OR “floor effect” OR “Item response model” OR IRT OR Rasch OR “Differential item functioning” OR DIF OR “computer adaptive testing” OR “item bank” OR “cross-cultural equivalence”)

#4 (NOT)

“Not applicable”

Search CINAHL + Psycinfo systematic review distal radius#

Limits: Publication date 1990-present

English/dutch/german

#1

(“activities of daily living*” OR “activities of daily life*” OR “physical activity*” OR “physical function*” OR “functional ability*” OR “everyday functioning*” OR “functional status*” OR “function*” OR “physical impairment*” OR “after hand therapy*” OR “Boston questionnaire*” OR “Brigham and Womens* carpal tunnel questionnaire*” OR “DASH” OR “disabilities of arm, shoulder and hand” OR “quick DASH” OR “forearm symptom severity scale*” OR “functional index*” OR “Gartland and Werley scoring system” OR “Green and Brien” OR “Michigan Hand outcomes Questionnaire” OR “MHQ” OR “New York Orthopedic Hospital wrist rating system” OR “patient focused wrist outcome” OR “patient outcomes of surgery-hand/arm” OR “POS-hand/arm” OR “patient rated wrist evaluation” OR “ PRWE” OR “sequential occupational dexterity assessment” OR “SODA”)

#2

(“radius fracture*” OR “colles fracture*” OR “distal radius fracture” OR “wrist fracture*” OR “antebrachial fracture*” OR “distal radial fracture” OR “radial fracture” OR “forearm fracture*” OR “smith fracture*” OR “smiths fracture*” OR “barton fracture*” OR “bartons fracture*” OR “chauffeur fracture*” OR “chauffeurs fracture*”)

#3
((instrumentation OR methods OR Validation Studies OR Comparative Study OR psychometrics OR psychometr* OR clinimetr* OR clinometr* OR outcome assessment OR “outcome assessment” OR outcome measure* OR “observer variation” OR observer variation OR “Health Status Indicators” OR “reproducibility of results” OR reproducib* OR “discriminant analysis” OR reliab* OR unreliab* OR valid* OR coefficient OR homogeneity OR homogeneous OR “internal consistency” OR (cronbach* AND (alpha OR alphas)) OR (item AND (correlation* OR selection* OR reduction*)) OR agreement OR precision OR imprecision OR “precise values” OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intraobserver OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR “known group” OR factor analysis OR factor analyses OR dimension* OR subscale* OR (multitrait AND scaling AND (analysis OR analyses)) OR item discriminant OR interscale correlation* OR error OR errors OR “individual variability” OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR “standard error of measurement” OR sensitiv* OR responsive* OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (changeOR difference)) OR (small* AND (real OR detectable) AND (change OR difference)) OR meaningful change OR “ceiling effect” OR “floor effect” OR “Item response model” OR IRT OR Rasch OR “Differential item functioning” OR DIF OR “computer adaptive testing” OR “item bank” OR “cross-cultural equivalence”))

#4 (NOT)
((addresses OR biography OR “case reports” OR comment OR directory OR editorial OR festschrift OR interview OR lectures OR “legal cases” OR legislation OR letter OR news OR “newspaper article” OR “patient education handout” OR “popular works” OR congresses OR “consensus development conference” OR “consensus development conference”, nih OR practice guideline) NOT (animals NOT humans))

APPENDIX 2

Table 4a – Methodological quality of the studies on measurement properties. 4=excellent, 3=good, 2=fair, 1=poor, n/a=not applicable, Excel = excellent

	PRWE(41)	**PRWE**(42)	**PRWE**(43)	**PRWE**(32)	**PRWE**(35)	**PRWE**(34)	**PRWE**(44)	**PRWE**(46)	**PRWE**(45)	**PRWE**(47)	**DASH**(32)	**DASH**(36)	**DASH**(48)	**DASH**(34)
Generalisability	**Fair**	**Fair**	**Fair**	**Poor**	**Fair**	**Excel**	**Poor**	**Fair**	**Fair**	**Fair**	**Poor**	**Fair**	**Good**	**Excel**
Median or mean age	4	4	3	4	4	4	4	4	4	4	4	4	4	4
Distribution of sex	4	4	2	4	4	4	4	4	4	4	4	4	4	4
Disease character	2	2	2	4	4	4	2	2	4	4	4	4	4	4
Setting described	2	4	4	4	4	4	4	4	4	4	4	2	4	4
Countries performed	4	4	4	3	4	4	4	4	4	4	3	3	3	4
Language	3	4	3	3	4	4	4	4	4	4	3	4	4	4
Selecting patients	2	2	4	1	2	4	1	4	4	4	1	4	4	4
Response rate	2	2	3	2	2	4	2	2	2	2	2	4	4	4
InternalConsistency	**Poor**	**Poor**			**Fair**	**Poor**	**Poor**	**Poor**	**Poor**	**Poor**		**Poor**	**Poor**	**Poor**
% missing items	4	3			4	3	3	3	3	4		4	3	3
Handling items	2	2			4	2	2	4	2	4		3	2	2
Sample size	4	2			3	1	3	3	3	4		4	3	1
Factor analysis	1	1			2	1	1	1	1	1		1	1	1
Sample factoranalysis	n/a	n/a			n/a	n/a	n/a	n/a	n/a	3		n/a	n/a	n/a
Statistic subscales	n/a	n/a			n/a	n/a	n/a	n/a	n/a	3		n/a	n/a	n/a
Flaws in design	4	4			4	4	4	4	4	4		4	4	4
Cronbach alpha	1	4			4	4	4	4	4	4		4	4	4
Dichotomous scores	n/a	n/a			n/a	n/a	n/a	n/a	n/a	n/a		n/a	n/a	n/a
χ^2reliabilitycoefficient	n/a	n/a			n/a	n/a	n/a	n/a	n/a	n/a		n/a	n/a	n/a
Reliability		**Fair**	**Poor**		**Fair**	**Poor**	**Fair**	**Poor**	**Fair**				**Fair**	**Poor**
% missing items		3	3		4	3	3	3	3	4			3	3
Handling items		2	2		4	2	2	4	2	4			2	2

Table 4a – Methodological quality of the studies on measurement properties. (*continued*)

	PRWE(41)	**PRWE**(42)	**PRWE**(43)	**PRWE**(32)	**PRWE**(35)	**PRWE**(34)	**PRWE**(44)	**PRWE**(46)	**PRWE**(45)	**PRWE**(47)	**DASH**(32)	**DASH**(36)	**DASH**(48)	**DASH**(34)
Sample size		2	1		2	1	3	1	3	4			3	1
Two measurements		4	4		4	4	4	4	4	4			4	4
Independent admin		3	4		2	4	3	4	4	4			4	4
Time interval stated		4	4		4	4	4	4	4	4			4	4
Patients stable		2	4		2	3	3	3	3	1			3	3
Time interval		4	4		2	4	4	4	4	1			4	4
Conditions similar		3	2		3	4	4	1	2	3			3	4
Flaws in design		4	4		4	4	4	4	4	4			4	4
ICC		4	4		3	4	4	4	4	n/a			4	4
Kappa *		n/a	n/a		n/a	n/a	n/a	n/a	n/a	n/a			n/a	n/a
MeasurementError#									**Fair**	**Poor**				
SEM/SDC									4	1				
Content val (box D)			**Fair**					**Good**						
Relevance construct			4					4						
Relevancepopulation			2					3						
Relevance purpose			4					4						
comprehensive			4					4						
Flaws in design			4					4						
Hypotheses testing		**Fair**			**Good**		**Fair**	**Fair**	**Poor**			**Fair**		
% missing items		3			4		3	3	3			4		
Handling items		2			4		2	4	2			2		
Sample size		2			3		3	3	3			4		
Hypotheses a priori		4			4		4	4	1			4		
Direction correlation		4			4		4	4	3			4		
Magnitude correlatio		3			3		4	3	3			4		

Table 4a – Methodological quality of the studies on measurement properties. (*continued*)

	PRWE(41)	**PRWE**(42)	**PRWE**(43)	**PRWE**(32)	**PRWE**(35)	**PRWE**(34)	**PRWE**(44)	**PRWE**(46)	**PRWE**(45)	**PRWE**(47)	**DASH**(32)	**DASH**(36)	**DASH**(48)	**DASH**(34)
Comparator instrum		4			4		4	4	4			4		
Properties comparat		3			4		4	2	2			2		
Flaws in design		4			4		4	4	4			4		
Statistical methods		3			3		4	3	3			4		
Cross cultural					**Fair**		**Poor**	**Poor**	**Poor**					
% missing items					4		3	3	3					
Handling items					4		2	4	2					
Sample size					3		1	1	1					
Both languages					4		4	4	4					
Expertise adequate					3		4	4	4					
Translators indepen					3		4	1	4					
Forward/backward					4		4	1	4					
Difference resolving					4		4	4	3					
Review committee					4		4	3	4					
Instrum pre-tested					4		2	4	4					
Sample pre-test					2		4	4	4					
Samples similar					3		4	4	3					
Flaws in design					4		4	4	4					
Confirm analysis					n/a		1	1	1					
Differential function					n/a		n/a	n/a	n/a					
Responsiveness			**Fair**	**Fair**	**Fair**		**Fair**	**Fair**	**Poor**		**Fair**		**Fair**	
% missing items			3	3	4		3	3	3		3		3	
Handling items			2	2	4		2	4	2		2		2	
Sample size			4	3	3		3	3	3		3		3	
Two measurements			4	4	4		4	4	4		4		4	
Time interval stated			4	4	4		4	4	4		4		4	

Table 4a – Methodological quality of the studies on measurement properties. (*continued*)

	PRWE(41)	**PRWE**(42)	**PRWE**(43)	**PRWE**(32)	**PRWE**(35)	**PRWE**(34)	**PRWE**(44)	**PRWE**(46)	**PRWE**(45)	**PRWE**(47)	**DASH**(32)	**DASH**(36)	**DASH**(48)	**DASH**(34)
Anything occurred			3	3	2		2	3	2		3		2	
Patient changed			3	3	2		3	3	3		3		3	
Hypotheses a priori			4	n/a	n/a		4	4	1		n/a		n/a	
Direction correlation			4	n/a	n/a		4	4	3		n/a		n/a	
Magnitude corr			3	n/a	n/a		4	3	3		n/a		n/a	
Comparator instr			2	n/a	n/a		4	4	4		n/a		n/a	
Properties comparat			4	n/a	n/a		4	2	4		n/a		n/a	
Flaws in design			4	n/a	n/a		4	4	4		n/a		n/a	
Statistical methods			4	n/a	n/a		2	2	1		n/a		n/a	
Gold standard			n/a	n/a	n/a		n/a	n/a	n/a		n/a		n/a	
Flaws in design			n/a	n/a	n/a		n/a	n/a	n/a		n/a		n/a	
AUC calculated			n/a	n/a	n/a		n/a	n/a	n/a		n/a		n/a	
Sensitivity/specificity			n/a	n/a	n/a		n/a	n/a	n/a		n/a		n/a	
Interpretability					**Poor**				**Poor**		**Poor**	**Poor**		
% missing items					4				3		3	4		
Handling items					4				2		4	4		
Sample size					3				3		3	4		
Distrib total score					4				4		2	4		
Floor effect					4				4		4	1		
Ceiling effect					4				4		4	1		
Change scores					4				4		2	4		
MIC					1				1		1	4		
Flaws in design					4				4		4	4		

Table 4b – Methodological quality of the studies on measurement properties. 4=excellent, 3=good, 2=fair, 1=poor, n/a=not applicable, Excel = excellent

	MHQ(49)	**MHQ**(50)	**MHQ** (51)	**SF36**(32)	**SF36**(33)	**PEM**(52)	**IOF**(16)	**PFW**(53)	**AIMS2**(33)	**BWH**(33)	**TSK**(34)	**CAT** (34)	**SES**(34)
Flaws in design													
Hypotheses testing						**Poor**		**Poor**					
% missing items						3		3					
Handling items						2		2					
Sample size						4		1					
Hypotheses a priori						1		4					
Direction correlation						3		4					
Magnitude correlatio						3		4					
Comparator instrum						1		4					
Properties comparat						1		3					
Flaws in design						4		4					
Statistical methods						3		3					
Cross cultural													
% missing items													
Handling items													
Sample size													
Both languages													
Expertise adequate													
Translators indepen													
Forward/backward													
Difference resolving													
Review committee													
Instrum pre-tested													
Sample pre-test													

Table 4b – Methodological quality of the studies on measurement properties. (*continued*)

	MHQ(49)	**MHQ**(50)	**MHQ** (51)	**SF36**(32)	**SF36**(33)	**PEM**(52)	**IOF**(16)	**PFW**(53)	**AIMS2**(33)	**BWH**(33)	**TSK**(34)	**CAT** (34)	**SES**(34)
Samples similar													
Flaws in design													
Confirm analysis													
Differential function													
Responsiveness	**Fair**	**Fair**	**Fair**	**Fair**	**Poor**		**Fair**	**Poor**	**Poor**	**Poor**			
% missing items	3	3	3	3	3		3	3	3	3			
Handling items	2	2	2	2	2		4	2	2	2			
Sample size	2	3	3	3	1		3	1	1	1			
Two measurements	4	4	4	4	4		4	4	4	4			
Time interval stated	4	4	4	4	4		4	4	4	4			
Anything occurred	3	2	2	3	2		2	3	2	2			
Patient changed	3	3	4	3	3		3	3	3	3			
Hypotheses a priori	n/a	n/a	4	n/a	n/a		n/a	4	n/a	n/a			
Direction correlation	n/a	n/a	4	n/a	n/a		n/a	4	n/a	n/a			
Magnitude corr	n/a	n/a	3	n/a	n/a		n/a	4	n/a	n/a			
Comparator instr	n/a	n/a	4	n/a	n/a		n/a	4	n/a	n/a			
Properties comparat	n/a	n/a	4	n/a	n/a		n/a	4	n/a	n/a			

Table 4b – Methodological quality of the studies on measurement properties. (*continued*)

	MHQ(49)	**MHQ**(50)	**MHQ** (51)	**SF36**(32)	**SF36**(33)	**PEM**(52)	**IOF**(16)	**PFW**(53)	**AIMS2**(33)	**BWH**(33)	**TSK**(34)	**CAT** (34)	**SES**(34)
Flaws in design	n/a	n/a	4	n/a	n/a		n/a	4	n/a	n/a			
Statistical methods	n/a	n/a	3	n/a	n/a		n/a	4	n/a	n/a			
Gold standard	n/a	n/a	n/a	n/a	n/a		n/a	n/a	n/a	n/a			
Flaws in design	n/a	n/a	n/a	n/a	n/a		n/a	n/a	n/a	n/a			
AUC calculated	n/a	n/a	n/a	n/a	n/a		n/a	n/a	n/a	n/a			
Sensitivity/specificity	n/a	n/a	n/a	n/a	n/a		n/a	n/a	n/a	n/a			
Interpretability													
% missing items		3											
Handling items		2											
Sample size		3											
Distrib total score		4											
Floor effect		3											
Ceiling effect		4											
Change scores		4											
MIC		4*											
Flaws in design		2											

*** Could not be determined due to high ceiling effect; # The first ten questions of the measurement property reliability and measurement error are the same**

CHAPTER 4

Are the Patient-Rated-Wrist-Evaluation (PRWE) and the Disability of the Arm, Shoulder and Hand questionnaire (DASH) used in distal radius fractures truly valid and reliable?

<u>Y.V. Kleinlugtenbelt</u>
R. Krol
M. Bhandari
J.C. Goslings
R.W. Poolman
V.A.B. Scholtes

Accepted in Bone and Joint Research 2017

ABSTRACT

Background

The Patient-Rated-Wrist-Evaluation (PRWE) and the Disability of the Arm, Shoulder and Hand questionnaire (DASH) are patient-reported outcome measures (PROM) used for clinical and research purposes. Methodological high quality clinimetric studies determining the measurement properties of these PROMs when used in patients with distal radius fractures are lacking. This study aimed to validate the PRWE and DASH in Dutch patients with displaced distal radius fractures.

Patients and methods

Internal consistency was determined using Cronbach's α for the dimensions found in the factor analysis. The intra-class correlation coefficient (ICC) was used for (test-retest) reliability. The measurement error was expressed bij the smallest detectable change (SDC). A semi-structured interview was conducted to assess the content validity.

Results

19 patients (mean age 58 years (SD 15)), 74% female, completed PROMs at a mean time of 6 months (SD 1) post-fracture. One overall meaningful dimension was found for the PRWE and the DASH. Internal consistency was excellent for both PROMs (Cronbach's α 0.96 (PRWE) and 0.97 (DASH)). Test-retest reliability was good (ICC 0.87) for the PRWE and excellent for the DASH (ICC 0.91). The SDC was 20 for the PRWE and 14 for the DASH. No floor or ceiling effects were found. The content validity was good for both questionnaires.

Conclusion

The Dutch PRWE and DASH are valid and reliable PROMs in assessing function and disability in patients with displaced distal radius fractures. However, due to the high SDC the PRWE and DASH are less useful for individual patients with distal radius fractures in clinical practice.

BACKGROUND

In order to conduct high quality clinical studies in the treatment of patients with distal radius fractures and to exchange results globally in a standardised way, there must be consensus on the use of outcomes. Instruments like Patient Related Outcome Measurements (PROMs) are gaining importance in clinical trials of fracture treatment.(1) The methodological quality of these instruments is important; they should be valid and reliable. This should ideally be determined prior to use, as the quality of such instrument directly defines the quality of the information obtained with this instrument.(2) If not, one risks imprecise or biased results, potentially leading to wrong conclusions.(3) To assess the methodological quality of a PROM, standards are needed. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) group set these standards for adequate study design and statistical analysis.(2) They also developed a checklist in an international Delphi study in which consensus was reached on terminology, definitions, and a taxonomy of measurement properties of PROMs.(3)

Recently, we performed a systematic review in which we used this COSMIN checklist to determine the methodological quality of studies that evaluated measurement properties of various PROMs used to evaluate outcome in patients with distal radius fractures.(4) The Patient-Rated Wrist Evaluation (PRWE) and the Disabilities of the Arm, Shoulder and Hand (DASH) were most extensively evaluated in terms of measurement properties. However, strong evidence supporting good quality of any of the current available PROMs in patients with distal radius fractures is lacking. We found that overall, the measurement properties are good, but the methodological quality of these studies is overall low.(5-16). So, based on this review, we currently risk imprecise or biased results when using these PROMs in for instance clinical studies, and might base our knowledge on wrong conclusions. The review has shown that studies of higher methodological quality are needed to adequately determine its measurement properties. If the methodological quality of clinimetric studies continues to increase, PROMs can be selected more carefully.(4)

In the current study, we aim to further examine which PROMs have the best measurement properties for evaluation of functional outcome in patients with distal radius fractures. We will therefor determine the content validity, the (test-retest) reliability, internal consistency, measurement error, and floor and ceiling effects of the Dutch

PRWE and DASH in patients with distal radius fractures. The measurement properties will be assessed according to the recently formulated COSMIN standards.(2)

PATIENTS AND METHODS

Study design

A multi center prospective cross-sectional clinimetric-study was performed between July 2012 and April 2013 at the orthopaedic and surgery departments of three participating hospitals. Ethics approval was obtained from the local medical ethical committee at all three hospitals (WO 12.064).

Study patients

Patients were eligible for inclusion if they were 18 years-of-age or older presenting with a displaced distal radius fracture in the emergency department requiring reduction. Both conservatively and surgically treated patients were included. Patients were excluded if they 1) had a prior fracture or pathology of the distal radius, 2) had multiple fractures 3) had cognitive impairment or 4) were unable to understand the Dutch language.

We aimed to include at least 20 patients to assess the content validity, which is double the number that is required.(17) This group was retrieved at 8-12 weeks post fracture.

In addition, we aimed to include at least 100 patients to assess the reliability domain, as required to obtain a high quality study according to the COSMIN guidelines.(17) These patients were retrieved at 4-8 months post fracture and did not participate in the content validation. The The type of fracture was scored on radiographs according to AO classification(18).

The PRWE is a self-administered, patient specific questionnaire, consisting 15 items. The PRWE was designed to measure wrist pain and disability in activities of daily living. The PRWE consists of two subscales: pain and function. The pain subscale consists of five items and the subscale function is divided in six specific activities and four usual activities. Both subscales are summed and scored on a ten point ordinal scale.(7,19) The score of the subscale 'pain' is the sum of the five items. The score of the 'function' subscale is calculated by the sum of the ten items divided by two. The total score of the PRWE is the sum of the scores of both subscales. A score of 100 represent the worst functional score, whereas 0 represents no disability.

In 2004 the PRWE was modified into the PRWHE (Patient Rated Wrist/Hand Evaluation). The PRWHE consists of the same items and scoring system as the PRWE, with minor changes.(20) In the PRWHE the term 'wrist' was replaced by 'wrist/hand'. Also, two aesthetic items, that are not part of the scoring system, were added. Therefore measurement properties of these two items were not assessed in this study. The PRWHE-Dutch-Language-Version (PRWHE-DLV) was used in this study.(21)

The DASH is a self-administered questionnaire, developed to evaluate symptoms and physical function of the whole upper extremity. It is scored in two components: the main disability/symptom section and two sections. The main component of the DASH is a 30-item scale concerning the patient's health status during the preceding week: 21 items about the degree of difficulty in performing certain physical activities, five items about the severity of pain, activity-related pain, tingling, weakness and stiffness and four items concerning the effect of the upper extremity problems on social activities, work, sleep and self-image. Each item is scored on a five-point ordinal scale. To calculate the main DASH score, all completed responses are summed and averaged. This value is subtracted by one and multiplied by 25, giving a total score ranging from best to worst on a 0–100 scale.(22) At least 27 of the 30 items must be completed to calculate a score.

Both optional sections, high performance Sport/Music and Work consist of four items, scored on a five-point ordinal scale and calculated similarly. However all 5 items must be answered, as the percentage of missing items must not exceed 10%. The DASH-Dutch-Language-Version (DASH-DLV) was used in this study.(23)

Assessment of measurement properties

Validity

Content validity examines the degree to which the content of a Health Related-Patient Reported Outcome Measurement (HR-PROM) is an adequate reflection of the construct to be measured. For all measurement instruments, it is important that the content validity is assessed by experts. For PROMs, patients, particularly representatives of the target population, are the experts. They are the most appropriate assessors of the relevance of the items of the questionnaire.(24)

Content validity was assessed by phone. Eight weeks after incurring a distal radius fracture, patients were asked by phone to participate in the study. When patients agreed to participate, a semi-structured interview was conducted, in which the patient was asked about function. After the interview, a list was composed of the functional problems named during the interview. This list was compared with the items used in the PRWE and the DASH questionnaire. Content validity is based on judgement and no statistical testing is involved.(24) We considered the content validity to be good if 75% of the items of the PRWE and DASH matched the problems mentioned in the interviews.

Reliability

All patients who were eligible for inclusion received an information letter four to eight months after sustaining a distal radius fracture in which they were asked to participate in this study. Participants completed a web based questionnaire containing the PRWE and the DASH at home. If participants did not have access to the internet they could alternatively receive a paper version. Two weeks after completing the questionnaire for the first time patients received an email or letter in which they were asked to complete the questionnaire for the second time at home. In this two week interval, no major changes in the health status were expected and recall would be prevented.

The digital and paper versions of the DASH and PRWE were identical. All items on the web-based questionnaire were obligated to be answered, thus no missing items were expected. If a patient, who received a paper version had not answered all items he/she was contacted by phone and the patient was asked to answer the open items. If the patient did not want to answer any of these items they were excluded from the study.

The internal consistency is the degree of interrelatedness among items.(3) If items in a scale are summarized into a total score, it should be ascertained that the items are sufficiently correlated. This correlation is established by the internal consistency and indicates whether the items seem to measure the same construct.(25) If one item measures something else, it will have a lower item-total correlation than the other items. The internal consistency was assessed by using the first measurement of the test-retest reliability.

Firstly, exploratory factor analysis was performed to the determine whether the PROM forms only one overall dimension or consisted of more than one dimension. Factor analysis was assessed by calculating eigenvalues. An eigenvalue of one or higher indicated a dimension. The eigenvalues were presented in a scree-plot. The relative contribution of different dimension was judged based on the 'elbow' in the scree-plot and the percentage of variance.(24) If the PROM consisted of two or more dimensions, factor loading was assessed. Factor loading represents the correlation between the items in the PROM and the underlying dimensions. We considered factor loadings meaningful of at least 0.50.(26)

Internal consistency was determined by calculating the Cronbach's α for the dimensions found in the factor analysis. If the Cronbach's α has a value of > 0.70 items are considered sufficiently correlated.(27) However, values > 0.95 can indicate that the instrument contains too many items assessing the same underlying construct.(28)

The test-retest reliability is the proportion of the total variance in the measurements, due to true differences between patients over time.(3) This refers to the degree to which the measurement instrument is free from measurement error and estimates the extent to which scores for patients who have not changed are the same for repeated measurements at different time points.(3,29) High reliability is important for discriminative purposes if one wants to distinguish among patients. The test-retest reliability was assessed by calculating the intra-class correlation coefficient (ICC) with a 95% confidence interval (CI). The ICC is a relative parameter and will always have a value between 0 and 1. Higher values represent higher reliability. An ICC of >0.70 is considered acceptable, >0.80 good and >0.90 is considered excellent.(29) Only patients who completed the optional modules of the DASH both times were included for the test-retest assessment of these modules.

The measurement error is the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured.(3) When the measurement error is low or zero, the difference measured is due to true differences.

To express the measurement error, the standard error of the measurement (SEM) and the smallest detectable change (SDC) can be used. The SEM represents the standard deviation of repeated measures of one individual. The SDC represents the minimal change that must occur on the scale to affirm that the change occurred is a real change and not a measurement error. The SEM was calculated from the square root of the variance between the measurements and the error variance of the ICC. For a conventional confidence level of 95%, the SDC was calculated as 1.96 x √2 x SEM.(24)

The presence of floor or ceiling effects may have a negative effect on the quality of the instrument. If patients score primarily in the extremes, the responsiveness may be limited. Floor or ceiling effects were considered to be present if more than 15% of the respondents achieved the minimum or maximum possible score.(30) When taking the SDC into account we should consider floor and ceiling effects more broadly. If a score is closer to the maximum or minimum score than the SDC, a change beyond the measurement error cannot be measured. We also assessed the percentage patients within the SDC range from both extremes. Floor and ceiling effects were assessed by using the first measurement of the test-retest reliability.

Statistical analyses

Statistical analysis was conducted using the Statistical Package for the Social Sciences (SPSS) version 18.

RESULTS

Validity

35 patients met the inclusion criteria and received a phone call (Figure 1). Two patients refused participation in the study and 13 patients did not respond our phone call. Twenty patients were included in the study to assess the content validity and were interviewed.

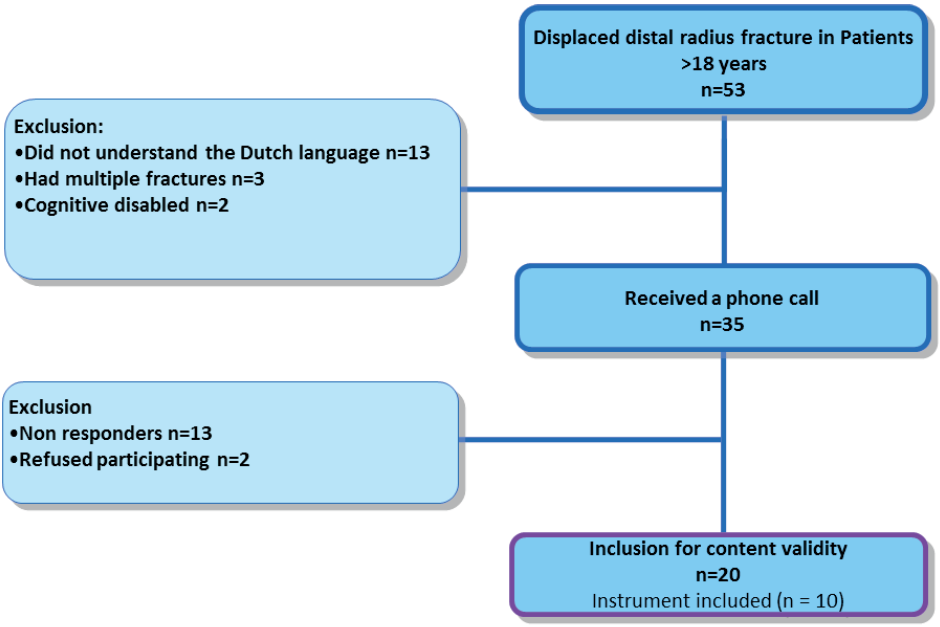

Figure 1: Patient recruitment and exclusion for the assessment of the content validity

The mean age was 59.30 (SD 13.61). Four times more females than men were included. Half of the patients had an AO subtype C1 fracture and most of the patients (65%) received conservative treatment (Table 1). The mean time between the fracture and the semi-structured interview was 9.85 weeks (SD 1.98).

Table 1. Study Characteristic

	Internal consistency	**Test-Retest**	Work module		Sport module		**Content Validity**
			Int cons	Test-Retest	Int cons	Test-Retest	
Number of patients	**n = 119**	**n = 109**	n = 84	n = 77	n = 70	n = 59	**n = 20**
Female, count (%)	88 (74%)	82 (75%)	60 (71.4%)	55 (71.4%)	52 (74.3%)	46 (78.0%)	16 (80%)
Mean age, years (SD)	58.40 (15.32)	58.76 (15.12)	53.74 (13.76)	54.35 (13.65)	55.37 (15.24)	55.83 (14.89)	59.30 (13.61)
AO subtype , count (%)							
A2	25 (21.0%)	23 (21.1%)	18 (21.4%)	17 (22.1%)	14 (20%)	12 (20.3%)	4 (20%)
A3	15 (12.6 %)	13 (11.9%)	10 (11.9%)	8 (10.4%)	10 (14.3%)	7 (11.9%)	2 (10%)
B1	1 (0.8%)	1 (0.9%)	1 (1.2%)	1 (1.3%)	1 (1.4%)	1 (1.7%)	0 (0%)
B2	0 (0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (5%)
B3	6 (5.0 %)	5 (4.6%)	5 (6.0%)	4 (5.2%)	3 (4.3%)	2 (3.4%)	2 (10%)
C1	35 (29.4%)	34 (31.2%)	23 (27.4%)	23 (29.9%)	17 (24.3%)	16 (27.1%)	10 (50%)
C2	30 (25.2%)	27 (24.8%)	23 (27.4%)	21 (27.3%)	22 (31.4%)	19 (32.2%)	1 (5%)
C3	7 (5.9%)	6 (5.5%)	4 (4.8%)	3 (3.9%)	3 (4.3%)	2 (3.4%)	0 (0%)
Treatment, count (%)							
Conservative	73 (61.3%)	67 (61.5%)	51 (60.7%)	48 (62.3%)	43 (61.4%)	36 (61.0%)	13 (65%)
K-wire fixation	1 (0.8%)	1 (0.9%)	1 (1.2%)	1 (1.3%)	1 (1.4%)	1 (1.7%)	0 (0%)
ORIF (volar plate)	39 (32.8%)	36 (33.0%)	30 (35.7%)	27 (35.%)	23(32.9%)	20 (33.9%)	6 (30%)
ORIF (dorsal plate)	0 (0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0%)
External fixation	6 (5.0%)	5 (4.6%)	2 (2.4%)	1 (1.3%)	3 (4.3%)	2 (3.4%)	1 (5%)

Characteristics of included patients for Internal consistency (Int cons), Test-retest reliability and Content validity. SD: Standard deviation

A total of 74 problems were mentioned in the semi-structural interviews. All 15 items (100%) of the PRWHE were named in the interviews. Therefore we considered the content validity of the PRWHE to be good. Only one DASH item, ‘gardening’, was not mentioned in the interviews. 96.7% of the questions of the DASH were mentioned in the interviews, which we considered to be good content validity. Both work and sport were mentioned in the semi-structured interview. Therefore we consider the content validity of both optional modules of the DASH to be good.

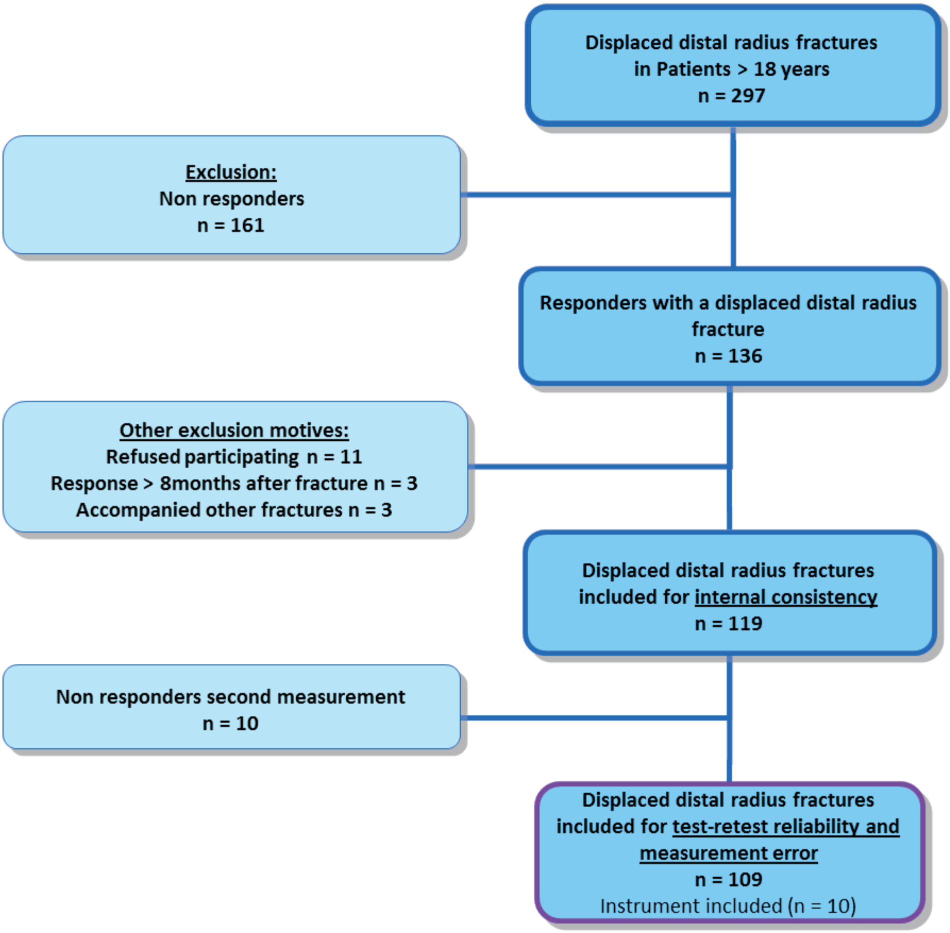

Figure 2: Patient recruitment and exclusion for the assessment of internal consistency, test-retest reliability and measurement error

Reliability

Of the 297 patients who entered the emergency room at one of the three participating hospitals during the study period with a displaced distal radius fracture requiring reduction, a total of 119 patients met the inclusion criteria and completed the questionnaire the first time after a mean time of 6.15 months (SD 1.00). Approximately three times more women were included and about two third had an intra-articular fracture. Ten patients did not respond to the second measurement. Therefore, 119 patients were included for the assessment of the internal consistency and 109 patients were included to assess the test-retest reliability and measurement error (Figure 2).The mean time between the first and second assessment was 18.66 days (SD 7.27) No major change in health status took place in any of the patients in between the measurement points, so no patients were excluded for that reason. Since not all patients were employed and/or played a sport or instrument, the optional work and sport modules of the DASH were assessed with fewer patients. The exact number of patients with their characteristics for each assessment are described in table 1.

Table 2. Results Internal consistency, reliability and measurement error

PROM	Cr α	Mean T1 (SD)		Mean T2 (SD)	ICC (95% CI)	SEM	SDC
		n=119	n=109	n=109			
PRWE	0.96	26.92 (21.16)	25.97 (20.37)	24.95 (20.73)	0.87 (0.82-0.91)	7.38	20.47
DASH	0.97	19.55 (17.70)	18.83 (16.59)	19.36 (17.93)	0.91 (087-0.94)	5.10	14.12
DASH_Work	0.94	15.40 (20.61)	14.98 (18,74)	14.06 (19.93)	0.87 (0.80-0.92)	5.08	14.08
DASH_Sport	0.96	33.66 (32.26)	33.39 (32.57)	28.92 (30.01)	0.87 (0.79-0.92)	11.18	30.99

Cronbach's alpha (Cr α), Mean score at timepoint 1(T1) and 2(T2). Intra Class Correlation (ICC). Standard Error Measurment (SEM) and Smallest Detectable Change (SDC) of the DASH and PRWE

Table 3. Floor and ceiling effects

PROM	Absolute (%)		SDC range		pt within SDC range(%)	
	Floor	Ceiling	Floor	Ceiling	Floor	Ceiling
PRWHE	5.9	0.8	0-20.47	79.53-100	45.4	1.7
DASH	4.2	0	0-14.12	85.88-100	46.6	0.8
DASH_Work	44	1.2	0-14.08	85.92-100	58.3	2.4
DASH_Sport	20	8.6	0-30.99	69.01-100	65.7	18.6

No missing items were seen in the online completed questionnaires. Twenty-three patients completed the questionnaires on a paper version. No missing items were seen in the PRWE's. Twice a missing item was seen in the DASH. Even after a phone consult the answer could not be retrieved (the patients did not want to answer the sexuality item of the DASH). Both patients were excluded for further assessment.

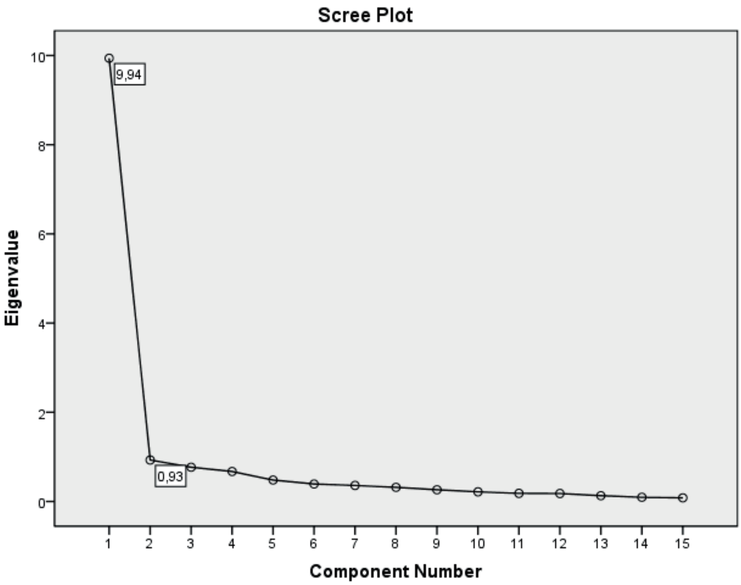

Figure 3: Screeplot PRWE. Only one dimension was extracted of the PRWE, which explained 66.26% of the total variance. The “elbow” in the scree plot was seen at the second component

The factor analysis was performed for both the PRWE and the DASH. Only one dimension was extracted of the PRWE, which explained 66.26% of the total variance. The “elbow” in the scree plot was seen at the second component. (Figure 3) The Cronbach’s α of the total PRWE was 0.96 (Table 2), indicating excellent internal consistency and redundancy. Removing items of the questionnaire did not result in a higher Cronbach’s α.

Five dimensions were extracted of the DASH. Item loading (Table 4) showed that the first dimension consisted of items, typically asking about the strength. The second dimension is more specific on function. The third dimension consisted of items about pain and disabilities. The fourth and fifth dimension consisted of two or more specific items. Dimension one explained 55.71% of the total variance and the scree-plot showed an ‘elbow’ at component two, insinuating only one overall meaningful dimension. (Figure 4) Therefore, despite extracting five dimensions we assessed the Cronbach’s α only for the total DASH and the optional modules.

Table 4. Factor analysis DASH

Question	Component				
	1	2	3	4	5
1	0,76	0,24	0,26	0,14	0,15
2	0,14	0,81	0,15	0,06	0,11
3	0,33	0,48	0,39	0,35	0,26
4	0,35	0,46	0,43	0,31	0,34
5	0,62	0,34	0,34	0,28	0,05
6	0,53	0,51	0,17	0,21	0,08
7	0,61	0,35	0,40	0,31	0,11
8	0,62	0,36	0,24	0,35	0,29
9	0,45	0,58	0,22	0,31	0,20
10	0,73	0,10	0,21	0,34	0,24
11	0,86	0,07	0,15	0,20	0,23
12	0,41	0,53	0,18	0,48	0,13
13	0,20	0,64	0,09	0,52	0,23
14	0,14	0,57	0,25	0,42	0,17
15	0,28	0,42	0,21	0,58	0,39
16	0,27	0,68	0,24	0,37	0,20
17	0,20	0,45	0,31	0,59	0,25
18	0,57	0,49	0,32	0,07	0,19
19	0,61	0,40	0,32	0,20	0,19
20	0,29	0,15	0,19	0,75	0,12
21	0,27	0,27	0,27	0,65	0,24
22	0,20	0,12	0,77	0,22	-0,14
23	0,28	0,34	0,68	0,16	0,07
24	0,27	0,42	0,57	0,05	0,53
25	0,51	0,27	0,53	-0,03	0,38
26	0,17	0,08	0,04	0,29	0,70
27	0,53	0,14	0,44	0,23	0,40
28	0,19	0,19	0,59	0,24	0,28
29	0,30	0,31	0,12	0,23	0,68
30	0,32	0,03	0,58	0,28	0,22

Factor loadings > 0.50 are appropriate (marked).

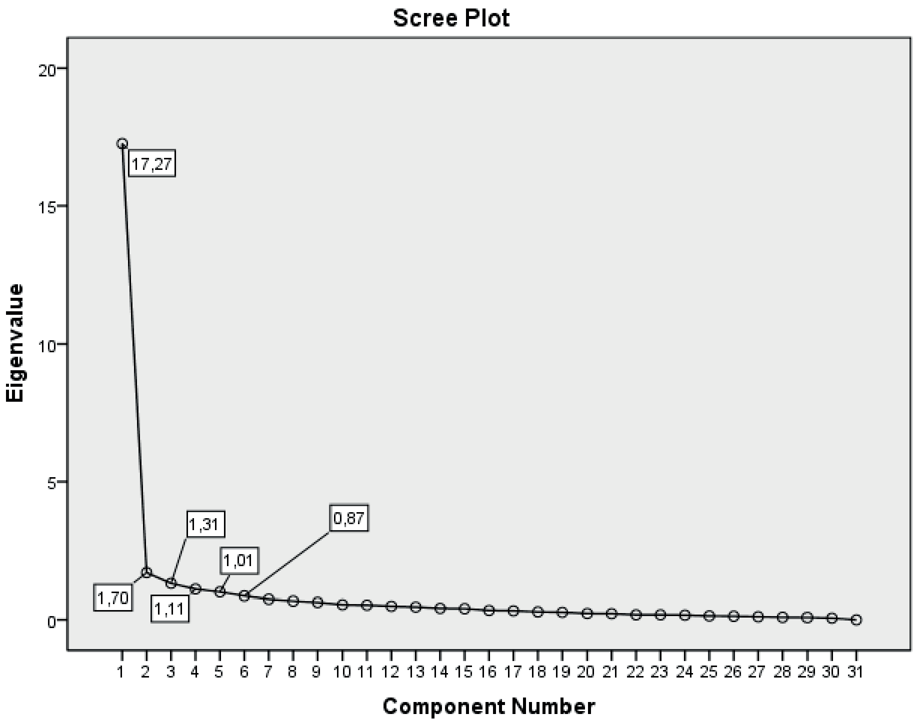

Figure 4: Screeplot DASH. Five dimensions were extracted of the DASH. Dimension one explained 55.71% of the total variance and the "elbow" in the scree plot was seen at the second component, insinuating only one overall meaningful dimension.

The Cronbach's α of the total DASH, the optional work and sport/music section were 0.97, 0.94 and 0.96, respectively, indicating an excellent internal consistency. Removing items of the questionnaire did not result in a higher Cronbach's α.

Table 2 shows the distribution of the data at the first and second measurement. The ICC of the PRWE was 0.87, which indicates good reliability. The SEM was 7.40 with a SDC of 20.51. The ICC of the total DASH, the optional work module and the sport module were 0.91, 0.87 and 0.87, indicating excellent reliability for the total DASH and good reliability for the optional modules. The SEM were 5.10, 5.08 and 11.18 with a SDC of 14.12, 14.08 and 30.99 respectively.

No floor or ceiling effects were found in the PRWE and total DASH questionnaire. Only floor effects were found in both optional modules of the DASH questionnaire. (Table 3). When taking the SDC in consideration, more than 15% of the scores of both the PRWE and the total DASH were within the SDC from the lowest possible score (45%

and 46%), showing a clear floor effect. Only the score for the optional sport module of the DASH questionnaire was within the SDC from the highest possible score.

DISCUSSION

Although the PRWE and the DASH are the most thoroughly studied PROMs, the quality of these two PROMS was not supported with strong levels of evidence on any of the measurement properties to evaluate patients with distal radius fractures.(4) This study, in which the COSMIN standards were followed to ensure high methodological quality, provides strong evidence that both the PRWE and the DASH questionnaires have good content validity and are reliable and internally consistent instruments for the assessment of patients with distal radius fractures.

Using semi-structured interviews with patients with distal radius fractures we found good content validity for both the PRWE and the DASH. In a previous study, only the developers of the PRWE assessed the content validity of the PRWE for patients with distal radius fractures.(7) They assessed the content validity by interviewing experts in relevant field of medicine instead of patients with distal radius fractures. However, the most appropriate assessors of the relevance of items on a questionnaire are the representatives of the target population.(24) To our knowledge this is the first study in which the assessment of the content validity of the DASH in patients with distal radius fractures was performed. A frequent remark was that the affected distal radius fracture was not the dominant side. Another remark was about cutting food. Most people cut food with their right hand. Five (female) patients stated that they experienced problems with putting on their bra. Overall, functional problems experienced by patients with a distal radius fracture consisted of proceedings for which both hands are needed.

To our knowledge this study is the first study in which a factor analysis of the PRWE and the DASH was performed specifically in patients with distal radius fractures. The PRWE was developed as a one-dimensional questionnaire. However, that dimension consists of two subscales (pain and function). Exploratory factor analysis extracted only one dimension of the PRWE. Therefore the Cronbach's α was assessed only for the total 15 items on the PRWE. The DASH was also developed as a one-dimensional questionnaire, with two optional modules. However, we distracted five dimensions. Despite extracting five dimensions we assessed the Cronbach's α only for the total

DASH, as the total score is calculated by using all 30 questions. Component one explained 57.71% of the total variance and the scree-plot showed an ‘elbow’ at component two, insinuating only one overall meaningful dimension.

For both PRWE and DASH a high Cronbach’s α was found, respectively 0.96 and 0.97, as comparable to previous validation studies.(5,6,9-15) However a Cronbach’s α>0.95 could indicate item redundancy. This suggests that some items can be removed when using one of these measurement instruments in patients with distal radius fractures.

We determined a good (ICC=0.87) and excellent (ICC=0.91) reliability for the PRWE and DASH questionnaires respectively, which is comparable to other studies. This study is the third to report on measurement error of the PRWE and the first for the DASH in patients with distal radius fractures. In our study, the SDC of the DASH (14.1) was clearly lower than the SDC of the PRWE (20.5) which could imply that it is more useful, especially in clinical practice.

Kim(12) reported a SDC of 4.4 for the PRWE in 63 patients with distal radius fractures. We found a much higher SDC value of 20.5, indicating that a patient has to improve at least 20% of the total score to ensure an improvement beyond measurement error. This high difference can be partly explained by the fact that they used a confidence interval of 90% instead of 95%. In their study, outcomes were more homogeneous than in our study, which normally led to a lower ICC. However, surprisingly they calculated a higher ICC than we did in our study. Based on their data we could not find an explanation for this difference. As a result their SDC was correspondingly very low.

Walenkamp et all found a SDC of 11. This difference can also partly be explained by the fact that they a confidence interval of 90% instead of 95%. However, the main reason for this difference is probably that they used Cronbach´s α instead of test-. retest parameters (eg. ICC) to calculate the SDC. Cronbach´s α is assessed at a single point in time and does not reflect the variation in scores when the measurement is assessed at different time points. Therefore it is not sufficient to base the SDC on Cronbach´s α.(24)

John(31) reported a SDC of 22.5 in 51 patients with resection interposition arthroplasty for carpometacarpal osteoarthritis. Although these are different patients, their methodology, and therefore their results are more comparable.

For both the PRWE and the total DASH no (substantial) floor and ceiling effects were found. However, when the SDC is taken into account, respectively 45.4% and 46.6% of the patients were within the SDC-range of a floor effect. No real health improvement beyond measurement error could be detected in this group of patients. Besides the aforementioned high SDC, this floor effect makes the PRWE and DASH less useful for individual patients with distal radius fracture in clinical practice. However, when measuring groups of patients (e.g. a randomized controlled trial), the SDC is reduced by a factor √n, when n patients is studied.(24)

A strength of this study is that we used the COSMIN standards for adequate study design and statistical analysis and our large population of patients with distal radius fractures, with only 10 patients (8%) lost to follow up for the test-retest reliability. Furthermore, we had no missing items in the assessment of the reliability for the PRWE and only twice for the total DASH. This was an advantage of the online questionnaires with required questions.

A limitation of this study is that we could not determine the responsiveness and minimal important change (MIC) of the PRWE and DASH in patients with distal radius fractures. Responsiveness is defined as the ability to detect clinically important changes over time(3), MIC (minimal important change) is part of the measurement property responsiveness. The MIC of PRWE and DASH was determined in patients who were treated nonoperatively for isolated tendinitis, arthritis or nerve compression syndromes from forearm to hand.(32) These data cannot be generalized to patients with distal radius fractures. Walenkamp et al determined the MIC for the PRWE in patients with distal radius fractures.(16) However, based on the COSMIN guidelines this study lacks high methodological quality.(4) Determining the MIC of the PRWE and DASH in patients with distal radius fractures should be an important part of future research.

CONCLUSION

The Dutch version of the PRWE and the DASH are valid and reliable PROMs in assessing function and disability in patients with displaced distal radius fractures. Due to the high SDC, the PRWE and DASH are less useful for individual patients with distal radius fractures in clinical practice.

REFERENCES

(1) Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2(14):i-iv, 1-74.

(2) Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010 May;19(4):539-549.

(3) Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-745.

(4) Kleinlugtenbelt YV, Nienhuis RW, Bhandari M, Goslings JC, Poolman RW, Scholtes VA. Are validated outcome measures used in distal radial fractures truly valid? A critical assessment using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. Bone Joint Res 2016 Apr;5(4):153-161.

(5) Gabl M, Krappinger D, Arora R, Zimmermann R, Angermann P, Pechlaner S. Acceptance of patient-related evaluation of wrist function following distal radius fracture (DRF). Handchir Mikrochir Plast Chir 2007 Feb;39(1):68-72.

(6) Hemelaers L, Angst F, Drerup S, Simmen BR, Wood-Dauphinee S. Reliability and validity of the German version of “the Patient-rated Wrist Evaluation (PRWE)” as an outcome measure of wrist pain and disability in patients with acute distal radius fractures. J Hand Ther 2008 Oct-Dec;21(4):366-376.

(7) 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 Nov-Dec;12(8):577-586.

(8) MacDermid JC, Richards RS, Donner A, Bellamy N, Roth JH. Responsiveness of the short form-36, disability of the arm, shoulder, and hand questionnaire, patient-rated wrist evaluation, and physical impairment measurements in evaluating recovery after a distal radius fracture. J Hand Surg Am 2000 Mar;25(2):330-340.

(9) Wilcke MT, Abbaszadegan H, Adolphson PY. Evaluation of a Swedish version of the patient-rated wrist evaluation outcome questionnaire: good responsiveness, validity, and reliability, in 99 patients recovering from a fracture of the distal radius. Scand J Plast Reconstr Surg Hand Surg 2009;43(2):94-101.

(10) Lovgren A, Hellstrom K. Reliability and validity of measurement and associations between disability and behavioural factors in patients with Colles’ fracture. Physiother Theory Pract 2012 Apr;28(3):188-197.

(11) Mehta SP, Mhatre B, MacDermid JC, Mehta A. Cross-cultural adaptation and psychometric testing of the Hindi version of the patient-rated wrist evaluation. J Hand Ther 2012 Jan-Mar;25(1):65-77; quiz 78.

(12) Kim JK, Kang JS. Evaluation of the Korean version of the patient-rated wrist evaluation. J Hand Ther 2013 Jul-Sep;26(3):238-43; quiz 244.

(13) Schonnemann JO, Hansen TB, Soballe K. Translation and validation of the Danish version of the Patient Rated Wrist Evaluation questionnaire. J Plast Surg Hand Surg 2013 Dec;47(6):489-492.

(14) Westphal T, Piatek S, Schubert S, Schuschke T, Winckler S. Reliability and validity of the upper limb DASH questionnaire in patients with distal radius fractures. Z Orthop Ihre Grenzgeb 2002 Jul-Aug;140(4):447-451.

(15) Westphal T. Reliability and responsiveness of the German version of the Disabilities of the Arm, Shoulder and Hand questionnaire (DASH). Unfallchirurg 2007 Jun;110(6):548-552.

(16) Walenkamp MM, Vos LM, de Muinck Keizer RJ, Rosenwasser MP, Goslings JC, Schep NW. The Minimum Clinically Important Difference of the Patient-rated Wrist Evaluation Score for Patients With Distal Radius Fractures. Clin Orthop Relat Res 2015 Oct;473(10):3235-3241.

(17) Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012 May;21(4):651-657.

(18) Marsh JL, Slongo TF, Agel J, Broderick JS, Creevey W, DeCoster TA, et al. Fracture and dislocation classification compendium - 2007: Orthopaedic Trauma Association classification, database and outcomes committee. J Orthop Trauma 2007 Nov-Dec;21(10 Suppl):S1-133.

(19) MacDermid JC. Development of a scale for patient rating of wrist pain and disability. J Hand Ther 1996 Apr-Jun;9(2):178-183.

(20) MacDermid JC, Tottenham V. Responsiveness of the disability of the arm, shoulder, and hand (DASH) and patient-rated wrist/hand evaluation (PRWHE) in evaluating change after hand therapy. J Hand Ther 2004 Jan-Mar;17(1):18-23.

(21) Videler AJ ST. De PRWHE-DLV: De NEderlandse versie van de Patient rated Wrist/Hand Evaluation. Nederlands tijdschrift voor Handtherapie 2008;17(2):8-11.

(22) Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). Am J Ind Med 1996 Jun;29(6):602-608.

(23) Veehof MM, Sleegers EJ, van Veldhoven NH, Schuurman AH, van Meeteren NL. Psychometric qualities of the Dutch language version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH-DLV). J Hand Ther 2002 Oct-Dec;15(4):347-354.

(24) Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. 1st ed. Cambridge: Cambridge University Press; 2011.

(25) Karanicolas PJ, Bhandari M, Kreder H, Moroni A, Richardson M, Walter SD, et al. Evaluating agreement: conducting a reliability study. J Bone Joint Surg Am 2009 May;91 Suppl 3:99-106.

(26) Nunally JC, Bernstein IH. Psychometric theory. 3rd ed. New York: McGraw-Hill; 1994.

(27) Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. Clin Ther 1996 Sep-Oct;18(5):979-992.

(28) Pynsent PB. Choosing an outcome measure. J Bone Joint Surg Br 2001 Aug;83(6):792-794.

(29) Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury 2011 Mar;42(3):236-240.

(30) McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995 Aug;4(4):293-307.

(31) John M, Angst F, Awiszus F, Pap G, Macdermid JC, Simmen BR. The patient-rated wrist evaluation (PRWE): cross-cultural adaptation into German and evaluation of its psychometric properties. Clin Exp Rheumatol 2008 Nov-Dec;26(6):1047-1058.

(32) 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 Apr;38(4):641-649.

CHAPTER 5

Spectrum bias: a common unrecognized issue in orthopedic agreement studies. Do CT scans really influence the agreement on treatment plans in distal radius fractures?

Y.V. Kleinlugtenbelt
M. Hoekstra
S.J. Ham
P. Kloen
R. Haverlag
M.P. Simons
M. Bhandari
J.C. Goslings
R.W. Poolman
V.A.B. Scholtes

Published in Bone and Joint Research. 2015 Dec;4(12):190-4.

ABSTRACT

Background

Current studies on the additional benefit of using computerized tomography (CT) to evaluate surgeon agreement on fracture treatment plans are inconsistent. This inconsistency can be explained by a methodological phenomenon called “spectrum bias”, defined as the bias inherent when investigators choose a population lacking therapeutic uncertainty for evaluation. The aim of the study is to determine the influence of spectrum bias on the intraobserver agreement of distal radius fracture (DRF) treatment plans.

Patients and methods

Four surgeons evaluated 51 patients with displaced (DRF) at four time points: T1 and T2: conventional radiographs (X-ray); T3 and T4: radiographs and additional CT scan (X-ray+CT). Choice of treatment plan (operative or nonoperative) and therapeutic certainty (5 point scale: very uncertain to very certain) were rated. To determine the influence of spectrum bias the intra observer agreement was analyzed, using Kappa statistics, for each degree of therapeutic certainty.

Results

In cases with high therapeutic certainty, intraobserver agreement based on X-ray was almost perfect (range 0.86-0.90), but decreased to moderate based on X-ray+CT (range0.47-0.60). In cases with high therapeutic uncertainty intraobserver agreement was slightly at best (range-0.12-0.19), but increased to a moderate based on X-ray+CT (range 0.56-0.57).

Discussion

Spectrum bias influenced the outcome of this agreement study on treatment plans. An additional CT scan improves the intraobserver agreement on DRF treatment plans only when there is therapeutic uncertainty. Reporting and analyzing intraobserver agreement based on the surgeon’s level of certainty is an appropriate method to minimize spectrum bias.

INTRODUCTION

Various treatment methods are available for distal radius fractures (DRF), mostly guided by fracture characteristics and surgeons' expertise.(1) Historically, plain radiographs (X-ray) have played a large role in characterizing these different type of fractures. However, it is known that plain radiographs are not the best modality for accurate assessment of the distal part of the radius. Especially when a surgeon cannot assess the exact morphology of DRFs from X-rays alone, the use of Computed Tomography (CT) has currently become a popular additional imaging modality to evaluate this more accurately.(2-4)

The increased popularity of using both X-ray and CT may be supported by previous study-results, which show that, as compared to plain radiographs alone, the addition of a CT is a more accurate modality to assess certain fracture characteristics (e.g. the amount of comminution, involvement of the distal radio-ulnar joint and the extent of articular surface depression).(2,3)

5

So, adding a CT improves the accuracy of assessing fracture characteristics of DRFs. But does it also improve the agreement on treatment planning? Studies have found that the treatment plan (conservative or surgical) may shift after the addition of a CT. More specifically: when a treatment plan is based on both X-ray and CT, a surgeon is more likely to treat the patient with a DRF surgically than when the treatment plan is based on X-ray alone. However, the level of agreement in these treatments plans seems to be very inconsistent and varies as much as from no agreement to almost perfect agreement. (5-7). In addition, the agreement on treatment plans neither always improved when compared to X-rays alone. (5-7) One explanation for these apparently inconsistent results may be attributed to differences in the chosen study population in these agreement studies.

Ideally, the test results should be evaluated in a study population which is a perfect resemblance of the population of interest. If not, test results may be biased, as a result of so called "spectrum bias". If a clinically less appropriate population is chosen for a study of a diagnostic test, the results may seriously mislead clinicians.(8)

For example, when only cases with grossly dislocated extra articular fractures, with inadequate positions after closed reduction, are selected in these studies, the intra observer agreement will probably be very high either based on X-rays or based on X-rays and CT. This is because the therapeutic uncertainty will be low: surgeons will most likely plan to operate, based on an X-ray alone, and one would not expect them to change their treatment plan when they reassess a case with the addition of a CT. Therefore, this group of patients would not be an appropriate study population. If chosen, it would gives rise to spectrum bias, as this study population contains many cases without therapeutic uncertainty, and one would already expect that adding a CT scan will only minimally improve the intra observer agreement on treatment planning as when compared to using X-rays alone.

On the other hand, when only cases are selected in which the X-ray leaves room for interpretation, e.g. unclear presence or absence of intra-articular fracture lines, a possible step or gap deformity, the intra observer agreement will probably be low based on X-rays, because of the therapeutic uncertainty. Surgeons are more likely to obtain a CT scan for treatment planning, which is expected to improve the therapeutic certainty. Consequently, the intra observer agreement of the additional CT is expected to be higher in these cases. In fact, in clinical practice surgeons tend to use the additional CT scan for treatment planning especially in these cases where they lack therapeutic certainty.

Therefore, we wanted to evaluate the potential influence of spectrum bias, and examine whether or not the agreement on treatment plans is related to the surgeon's level of therapeutic certainty on their treatment plan. .

To address the potential influence of spectrum bias, we will determine the influence of the surgeon's level of therapeutic certainty on the intraobserver agreement in treatment plan in patients with displaced DRFs using X-ray alone or X-ray and CT. We hypothesised that 1) the intraobserver agreement is positively related to the surgeons' therapeutic certainty, both on X-ray and X-ray plus CT, 2) the level of certainty is most strongly related to the intraobserver agreement based on X-ray, 3) the intraobserver agreement only improves by the addition of CT in therapeutic uncertain cases.

PATIENTS AND METHODS

Study design

This retrospective cohort study was conducted according to the Collaboration for Outcome Assessment in Surgical Trials (COAST) guidelines.(9) Ethics approval was obtained from the medical ethical committee at the Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands (WO 10.086).

Study patients

Between January 1st , 2007 and March 2nd , 2011 a database was established of patients with a displaced DRF seen at the Emergency Department in a busy teaching hospital in Amsterdam, The Netherlands (Onze Lieve Vrouwe Gasthuis).

Patients were eligible for inclusion if they 1) presented with a displaced DRF in the emergency department, 2) were 18 years of age or older, 3) had no prior fracture or pathology of the distal radius, 4) had both pre- and post-reduction plain posterior-anterior and lateral radiographs of the wrist, and 5) had an additional post-reduction CT, made within 5 days after the reduction in case of any doubt of the characteristics of the fracture, or when there was a possible indication for surgery.

Observers

The panel consisted of four experienced Dutch surgeons, of which two were trauma surgeons [MS, RH] and two were orthopaedic surgeons [JH,PK]. They all have over 10 years of experience in fracture treatment. All of them are responsible for the (distal radius) fracture care within their department.

Time points

All surgeons scored the images at four different time points (T1-T4). The order of the images was randomized to differ at all time points. Each scoring round was performed with an interval of at least 4 weeks.

T1: pre- and post-reduction plain radiographs (T1 X-ray).
T2: pre- and post-reduction plain radiographs (T2 X-ray).
T3: pre- and post-reduction plain radiographs & axial, sagittal and coronal planes CT (T3 X-ray+CT).
T4: pre- and post-reduction plain radiographs & axial, sagittal and coronal planes CT (T4 X-ray+CT).

All images were digitalized and anonymized, and presented with the relevant clinical data (e.g., age of the patient, gender, dominant hand, profession and specific hobbies).

Scoring form

Scoring included: 1) choice of treatment plan (nonoperative treatment with plaster after closed reduction, or operative treatment). 2) therapeutic certainty on the treatment plan (1) very uncertain; 2) uncertain; 3) somewhat uncertain, 4) certain, 5) very certain).

Therapeutic certainty was defined as how confident the surgeon was about his treatment plan. For example if the surgeon was totally sure that he would treat a patient operatively he scored a 5 on the level of certainty. If he was unsure about the type of treatment he scored a 1 or 2 on the level of certainty.

Statistical analysis

We determined the intraobserver agreement in two different ways. Firstly, we determined the intraobserver agreement on treatment plans for each surgeon separately and calculated the mean agreement for the four surgeons.

Secondly, we analyzed the intraobserver agreement by the surgeon's therapeutic certainty, scored at T1. Because of having relatively small numbers in the "very uncertain" group, during the final data analysis we subsequently combined "very uncertain" and "uncertain" into one group.

T1 and T2 were used to determine the intraobserver reliability for X-ray.

T3 and T4 were used to determine the intraobserver reliability for X-ray+CT.

The agreement was determined using Kappa's statistic. The Kappa-values will be interpreted according to Landis and Koch(10): a score <0 indicates no agreement, 0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and >0.81indicates almost perfect agreement.

Source of Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

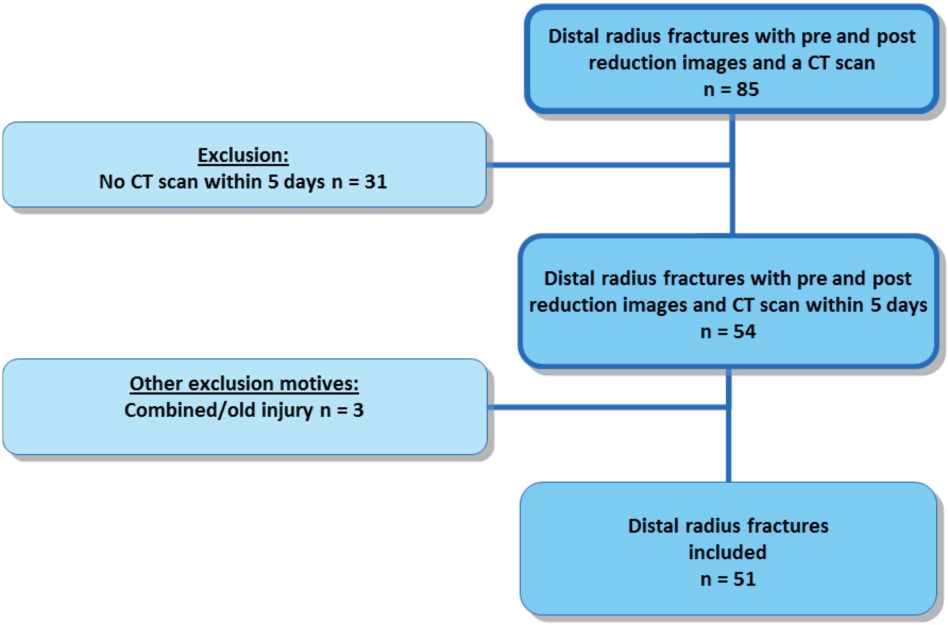

Figure 1: Flow chart

5

RESULTS

Study participants

During the study period, in 85 patients who entered the emergency room with a displaced DRF, a post-reduction CT scan was made. A total of 51 patients met the complete inclusion criteria (Figure 1). Their mean age was 50 years (SD, 14). 75 per cent of the patients were female. The CT scan was performed a mean of 2.53 days post-reduction (SD, 2.21).

Table 1 – Intra observer agreement for all cases

Observer	**X-ray (T1 – T2)**	**X-ray+CT scan (T3-T4)**
Observer 1	0,48 (0,23-0,72)	0,60 (0,39-0,81)
Observer 2	0,50 (0,14-0,86)	0,40 (0,14-0,66)
Observer 3	0,61 (0,34-0,87)	0,79(0,39-1,00)
Observer 4	0,83 (0,67-0,99)	0,44(0,23-0,65)
Mean	**0,69 (0,58-0,79)**	**0,57(0,45-0,69)**

Kappa statistics of the four observers and the mean with 95% confidence interval (CI) in parentheses for treatment plan on plain radiographs (X-ray) and plain radiographs with an additional CT scan (X-ray+CT scan)

Agreement treatment plans

The mean intraobserver agreement, regardless the level of therapeutic certainty, on treatment plan based on X-ray is substantial (0.69). Adding a CT scan resulted in moderate agreement (0.57) (Table 1).

Table 2 presents the agreement when the level of therapeutic certainty is taken into account. Based on X-ray alone, the intraobserver agreement is found to be positively related to the level of therapeutic certainty: it increased from no agreement (-0.12) in therapeutic uncertain cases to almost perfect (0.86) in therapeutic certain cases . Based on X-ray and CT, the degree of intraobserver agreement is found to be unrelated to the level of therapeutic certainty: it was moderate (range 0.47 – 0.60) for all therapeutic cases.

Table 2 – Intra observer agreement, based on surgeon's level of certainty

	X-ray (T1 – T2)		X-ray+CT scan (T3-T4)	
(very) uncertain	**-0,12**(-0,62-0,38)	No	**0,57**(0,18-0,95)	Moderate
somewhat uncertain	**0,19** (-0,11-0,48)	Slight	**0,56**(0,26-0,87)	Moderate
certain	**0,90** (0,76-1,00)	Almost perfect	**0,47**(0,19-0,75)	Moderate
very certain	**0,86** (0,76-0,96)	Almost perfect	**0,60**(0,44-0,76)	Moderate

Kappa statistics based on the surgeon's level of certainty with 95% confidence interval (CI) in parentheses for treatment plan on plain radiographs (X-ray) and plain radiographs with an additional CT scan (X-ray+CT scan)

For those cases where there was therapeutic uncertainty on the treatment plan based on X-ray, adding of a CT improved the intraobserver agreement. It improved from none to slight agreement based on X-ray, to moderate agreement based on X-ray+CT.

For those cases where there was therapeutic certainty on the treatment plan based on X-ray, adding of a CT worsened the intraobserver agreement. It decreased from almost perfect agreement (range 0.86-0.90) based on X-ray, to moderate agreement (range 0.47-0.60) based on X-ray+CT.

DISCUSSION

Using X-ray's alone, the level of therapeutic certainty is positively related to the intraobserver agreement, and even leading to no agreement when the surgeon is uncertain on the treatment plan. This influence is not seen on the intraobserver agreement based on X-ray+CT scan.

In therapeutic uncertain cases the intra observer agreement on treatment plan improves when an additional CT scan is used. In therapeutic certain cases the agreement is already perfect, so there is no room for improvement. In those cases we showed that an additional CT scan even diminished the agreement.

This clearly shows that the CT scan can indeed improve the intraobserver agreement on treatment plan, but only when there is therapeutic uncertainty.

The results based on our entire study population, without taking the surgeon's level of certainty into account, may let us conclude differently and are therefor misleading for clinicians. These results showed us that the intraobserver agreement on treatment plan did not increase when using additional CT scanning for decision making in treatment plans for DRFs. Moreover, it was even less reliable (X-ray alone: Kappa 0.69; X-ray and CT: (Kappa 0.57).

These differences in interpretation of our study results show the relevance of correcting for spectrum bias.

Previous literature

Our results could possibly explain the controversy in the additional value of CT scans for treatment planning in the existing literature. Clinicians do not need diagnostic tests when there is no therapeutic uncertainty. By adding the surgeons' level of therapeutic certainty to our analysis we minimized spectrum bias, and so we were able to determine the intraobserver agreement in a population with and without therapeutic uncertainty.

The controversy seen in DRF literature on agreement in treatment plan is also seen in other fracture types, eg proximal humerus fractures(11-13) and tibia plateau fractures(14-17). Although the CT scan has been shown to be more accurate to assess fracture characteristics, the studies which evaluated the agreement on treatment plans are inconsistent. Spectrum bias could not be excluded in these studies as well. Adding the surgeons' level of therapeutic certainty could possibly overcome this issue.

Strength and limitations

The strength of our study is that all observers were experienced in judging DRF imaging and treatment. As seen in many agreement studies the average intraobserver agreement will probably be slighter lower when you have less experienced surgeons. However, we would still expect a similar pattern: that the agreement based on X-ray is highly influenced by the surgeon's level of certainty on the treatment plan. Furthermore, all observers were blinded to the design and hypothesis of the study. Also, the order of images was randomized and the time in between scoring moments was adequate to avoid bias due to memory. Another strength of this study is that the COAST criteria were used to ensure we addressed all components of an agreement study.

A limitation in this study is the skewed distribution over the different groups of certainty. To maintain power we had to combine “very uncertain” and “uncertain” in one group.

Implications for future research

In summary, our study results show that there is an additional value of CT scanning over conventional radiographs in cases where there is therapeutic uncertainty in displaced DRFs. However, this does not mean that it has influence on the outcome of the patient. Prospective randomized studies should show whether the use of an additional CT scan and their resulting management, in cases of therapeutic uncertainty, influences outcomes in patients with displaced DRFs.

To the best of our knowledge, no previous agreement studies implemented the surgeon’s level of certainty in their analysis to minimize the effect of spectrum bias. This study shows that this is an appropriate method to determine the added value of a diagnostic tool to patients for whom the test would be clinically indicated. To address the current controversies in the additional value of CT scans for agreement in treatment plans in fracture care we suggest to use this method to minimize spectrum bias.

CONCLUSION

Our study shows that spectrum bias may influence the outcome of agreement studies on treatment plans. An additional CT scan improves the intraobserver agreement on distal radius treatment plans only when there is therapeutic uncertainty. Reporting and analyzing intraobserver agreement based on the surgeon’s level of certainty is an appropriate method to minimize spectrum bias.

REFERENCES

(1) Koval K, Haidukewych GJ, Service B, Zirgibel BJ. Controversies in the Management of Distal Radius Fractures. J Am Acad Orthop Surg 2014 Sep;22(9):566-575.

(2) Cole RJ, Bindra RR, Evanoff BA, Gilula LA, Yamaguchi K, Gelberman RH. Radiographic evaluation of osseous displacement following intra-articular fractures of the distal radius: reliability of plain radiography versus computed tomography. J Hand Surg Am 1997 Sep;22(5):792-800.

(3) Pruitt DL, Gilula LA, Manske PR, Vannier MW. Computed tomography scanning with image reconstruction in evaluation of distal radius fractures. J Hand Surg Am 1994 Sep;19(5):720-727.

(4) Dahlen HC, Franck WM, Sabauri G, Amlang M, Zwipp H. Incorrect classification of extra-articular distal radius fractures by conventional X-rays. Comparison between biplanar radiologic diagnostics and CT assessment of fracture morphology. Unfallchirurg 2004 Jun;107(6):491-498.

(5) Arora S, Grover SB, Batra S, Sharma VK. Comparative evaluation of postreduction intra-articular distal radial fractures by radiographs and multidetector computed tomography. J Bone Joint Surg Am 2010 Nov 3;92(15):2523-2532.

(6) Katz MA, Beredjiklian PK, Bozentka DJ, Steinberg DR. Computed tomography scanning of intra-articular distal radius fractures: does it influence treatment? J Hand Surg Am 2001 May;26(3):415-421.

(7) Hunt JJ, Lumsdaine W, Attia J, Balogh ZJ. AO type-C distal radius fractures: the influence of computed tomography on surgeon's decision-making. ANZ J Surg 2013 Sep;83(9):676-678.

(8) Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978 Oct 26;299(17):926-930.

(9) Karanicolas PJ, Bhandari M, Kreder H, Moroni A, Richardson M, Walter SD, et al. Evaluating agreement: conducting a reliability study. J Bone Joint Surg Am 2009 May;91 Suppl 3:99-106.

(10) Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174.

(11) Foroohar A, Tosti R, Richmond JM, Gaughan JP, Ilyas AM. Classification and treatment of proximal humerus fractures: inter-observer reliability and agreement across imaging modalities and experience. J Orthop Surg Res 2011 Jul 29;6:38-799X-6-38.

(12) Ramappa AJ, Patel V, Goswami K, Zurakowski D, Yablon C, Rodriguez EK, et al. Using computed tomography to assess proximal humerus fractures. Am J Orthop (Belle Mead NJ) 2014 Mar;43(3):E43-7.

(13) Brouwer KM, Lindenhovius AL, Dyer GS, Zurakowski D, Mudgal CS, Ring D. Diagnostic accuracy of 2- and 3-dimensional imaging and modeling of distal humerus fractures. J Shoulder Elbow Surg 2012 Jun;21(6):772-776.

(14) Brunner A, Horisberger M, Ulmar B, Hoffmann A, Babst R. Classification systems for tibial plateau fractures; does computed tomography scanning improve their reliability? Injury 2010 Feb;41(2):173-178.

(15) Chan PS, Klimkiewicz JJ, Luchetti WT, Esterhai JL, Kneeland JB, Dalinka MK, et al. Impact of CT scan on treatment plan and fracture classification of tibial plateau fractures. J Orthop Trauma 1997 Oct;11(7):484-489.

(16) te Stroet MA, Holla M, Biert J, van Kampen A. The value of a CT scan compared to plain radiographs for the classification and treatment plan in tibial plateau fractures. Emerg Radiol 2011 Aug;18(4):279-283.

5

(17) Yacoubian SV, Nevins RT, Sallis JG, Potter HG, Lorich DG. Impact of MRI on treatment plan and fracture classification of tibial plateau fractures. J Orthop Trauma 2002 Oct;16(9):632-637.

CHAPTER 6

Are Volar Locking Plates Superior to Percutaneous Kirschner Wires for Distal Radius Fractures? A Meta-analysis

H. Chaudhry
Y.V. Kleinlugtenbelt
R. Mundi
B. Ristevski
J.C. Goslings
M. Bhandari

Published in Clinical Orthopaedics and Related Research, 2015 Sep;473(9):3017-27.

ABSTRACT

Background

Distal radius fractures are common, costly, and increasing in incidence. Percutaneous Kirschner wire fixation (K-wires) and volar locking plates are two of the most commonly used surgical treatments for unstable dorsally displaced distal radius fractures. However, there is uncertainty regarding which of these treatments is superior. We performed a meta-analysis of randomized controlled trials to determine whether patients treated with volar locking plates (1) achieved better function, (2) attained better wrist motion, (3) had better radiographic outcomes, and (4) had fewer complications develop than did patients treated with K-wires for dorsally displaced distal radius fractures.

Methods

We performed a comprehensive search of MEDLINE (inception to 2014, October Week 2), EMBASE (inception to 2014, Week 42), and the Cochrane Central Register of Controlled Trials to identify relevant randomized controlled trials; we supplemented these searches with manual searches. We included studies of both extraarticular and intraarticular distal radius fractures. Adjunctive external fixation was acceptable as long as the intent was to use only K-wires where possible and external fixation was used in less than 25% of the procedures. We considered a difference in the DASH scores of 10 as the minimal clinically important difference. We performed quality assessment with the Cochrane Risk of Bias tool and evaluated the strength of recommendations using the GRADE approach. Seven randomized trials with a total of 875 participants were included in the meta-analysis.

Results

Patients treated with volar locking plates had slightly better function than did patients treated with K-wires as measured by their DASH scores at both 3 months (mean difference [MD], 7.5; 95% confidence interval [CI], 4.4-10.6; $p < 0.001$) and 12 months (MD, 3.8; 95% CI, 1.2-6.3; $p = 0.004$). Neither of these differences exceeded the *a priori*-determined threshold for clinical importance (10 points). There was a small early advantage in flexion and supination in the volar locking plate group (3.7° [95% CI 0.3°-7.1°, $p = 0.04$] and 4.1° [95% CI 0.6°-7.6°, $p = 0.02$]

greater, respectively) at 3-months, but not at later followup (6 or 12 months). There were no differences in radiographic outcomes (volar tilt, radial inclination, and radial height) between the two interventions. Superficial wound infection was more common in patients treated with K-wires (8.2% versus 3.2%, RR = 2.6, p = 0.001), but otherwise no difference in complication rates was found.

Discussion

Despite the small number of studies and the limitations inherent in a meta-analysis, we found that volar locking plate demonstrates better DASH scores at 3- and 12-month followups as compared with K-wires for displaced distal radius fractures in adults; however, these differences were small and unlikely to be clinically important. Further research is required to better delineate if there are specific radiographic, injury, or patient characteristics that may benefit from volar locking plates in the short term and whether there are any differences in long-term outcomes and complications.

BACKGROUND

Distal radius fractures are common injuries with over 600,000 occurring annually in the North American population (4). The distributive pattern of these injuries is bimodal, affecting both young (predominantly male) adults through high-energy mechanisms and elderly (predominantly female) adults through low-energy falls and osteoporosis (32). Economic costs of distal radius fractures also are substantial—direct costs of care are more than USD 480 million in the United States annually; more than USD 170 million of these costs are borne by publically funded Medicare (16, 39). As the population continues to age, the burden of distal radius fractures and the costs of care are expected to increase (33). Unfortunately, the treatment for these injuries is controversial (21). Therefore, determining effective evidence-based treatment of distal radius fractures is crucial.

There are multiple treatment options for patients with distal radius fractures, including closed reduction and cast immobilization, percutaneous K-wire fixation, fixation with volar or dorsal plates (locking or nonlocking), bridge plating, use of an external fixator, or a combination of these techniques. Although the best choice depends to some extent on the characteristics of the fracture (open/closed, nondisplaced/displaced, extra-/intraarticular), there is very little high-quality evidence to inform this decision-making. For instance, clinical practice guidelines for distal radius fracture published by the American Academic of Orthopaedic Surgeons (AAOS) made 29 recommendations; however, none of these recommendations was given a “strong” rating owing to limited strength of the evidence (25).

Most randomized trials, and all meta-analyses conducted to date have focused on comparisons between external fixators and internal plate fixation (17, 26, 44). However, it is becoming less common for the majority of distal radius fractures to be treated with an external fixator because these devices can be bulky and inconvenient for patients and typically are reserved for more severe fracture types (32). According to US Medicare data, internal fixation is the most common surgical intervention for distal radius fracture in the United States, followed closely by percutaneous pinning with K-wires (3). To our knowledge, there have been no meta-analyses comparing these two common interventions despite multiple trials on the topic having been published (5, 12, 18, 20, 29, 30, 37).

The objective of this study therefore was to perform a systematic review and meta-analysis of randomized trials comparing K-wire fixation to volar locking plates for displaced distal radius fractures. The specific goals of this meta-analysis were to determine whether patients treated with volar locking plates 1) achieved better function, 2) attained better wrist motion, 3) had better radiographic outcomes, and 4) had fewer complications than did patients treated with K-wires for dorsally displaced distal radius fractures.

MATERIALS AND METHODS

Search strategy and eEligibility

Our systematic review was conducted and reported in accordance with PRISMA guidelines (31). We performed a comprehensive search of three electronic medical databases: MEDLINE (inception to 2014, October Week 2), EMBASE (inception to 2014, Week 42), and the Cochrane Central Register of Controlled Trials (inception to Issue 9 of 12, September 2014) to identify relevant trials. We also supplemented our search with manual review of recent conference abstracts (Orthopaedic Trauma Association 2012-2014 and AAOS annual meetings 2012-2014) and reference lists. Reference Manager Software Version 12 (Thomson Reuters, Philadelphia, PA, USA) was used to manage the search. Our inclusion criteria were randomized controlled trials that compared volar locking plates with K-wires for distal radius fractures. We did not distinguish between the type of K-wire technique used (such as Kapandji, interfragmentary, mixed, or other). We defined a volar locking plate as any plate applied to the volar aspect of the radius with screws that locked into plate forming a fixed angle construct, with or without adjunctive use of non-locking screws. We included studies of both extraarticular and intraarticular distal radius fractures. We attempted to collect outcome data for only K-wires used alone; however, if not reported independently, we accepted adjunctive external fixation as long as the intent was to use only K-wires where possible and external fixation was used in less than 25% of the total cases.

Study selection and data extraction

Two reviewers (HC and YVK) screened all titles and abstracts for eligibility and conducted full-text reviews in duplicate. Discrepancies were resolved by consensus after discussion between the two reviewers. Data were collected using standardized data collection forms. We collected information pertaining to study characteristics, including publication year, study design, duration, location, number of centers, number of participants, mean age of participants, types of fractures included (AO type), and outcomes reported.

Data collection included functional outcome measures — specifically the Disabilities of the Arm, Shoulder and Hand (DASH) and Patient Rated Wrist Evaluation (PRWE) questionnaires, which are the best available patient-reported outcome measurement instruments for distal radius fractures and have been recommended for functional outcome measurement (13). We also collected reported data on grip strength, wrist ROM (flexion, extension, supination, pronation, ulnar deviation, radial deviation), complications, and radiographic outcomes. In cases in which wrist ROM was reported only as a percentage of the contralateral (normal) wrist, we converted percentages to a degree measurement based on normal physiologic ROM (normal values used: 85° flexion, 80° extension, 85° supination, 80° pronation, 35° ulnar deviation, and 20° radial deviation) (7, 28). Means and standard deviations (SDs) were collected when reported; medians were used in place of means when the latter was not reported, because these provide an acceptable alternate measurement for centrality (35). Where data were only reported in graphical format, GraphClick software (14) was used to extract the relevant values.

Quality assessment

We assessed quality of each included study in duplicate using the Cochrane Risk of Bias tool and reported this in chart format. In particular, this tool captures information on adequacy of randomization, allocation concealment, blinding, completeness of data collection, selective reporting, and other biases. Strength of recommendation for the functional outcome comparison was determined and reported using the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) approach (15).

Statistical analysis

Statistical analysis was performed using SPSS Software Version 21 (IBM Corp, Armonk, NY, USA) and Review Manager (RevMan) Version 5.3 software (36). Mean differences were pooled for common outcomes scores reported across studies or standardized mean differences (SMDs) for outcome scores that differed across studies. We calculated heterogeneity between studies using both the chi square test and the I^2 statistic. We considered either a chi square value of p less than 0.1 or I^2 statistic greater than 35% to represent significant heterogeneity. Outcomes with significant heterogeneity were pooled using a random-effects model; outcomes with low heterogeneity were pooled using a fixed-effects model.

Standard deviations were calculated for the medians from ranges using described methods (19). Where SDs or confidence intervals were not reported, we imputed SDs using a trial-and-error process to reproduce reported p values. Differences in complication rates were compared using the chi square statistical test. We considered a difference in DASH or quickDASH scores of 10 as the minimal clinically important difference (MCID) based on previously published studies and taking into consideration that reported values have not been evaluated specifically in distal radius fracture patients (10, 38). Given that normative data for these scoring instruments are sufficiently similar, and both are reported on a scale of 100, we pooled mean differences of these scores across all trials. We also performed a secondary analysis using SMDs, using 0.5 SDs as the MCID, as has been described as an appropriate threshold (34), to further corroborate our results. A p value of less than 0.05 was used to infer statistical significance.

Literature search

The search yielded 1202 citations (281 Medline, 361 EMBASE, 559 Cochrane Library, one from other sources), of which we excluded 488 duplicates, leaving 714 for title and abstract screening. Fourteen articles met criteria for full-text review, and seven of these met inclusion criteria for our meta-analysis (Fig. 1).

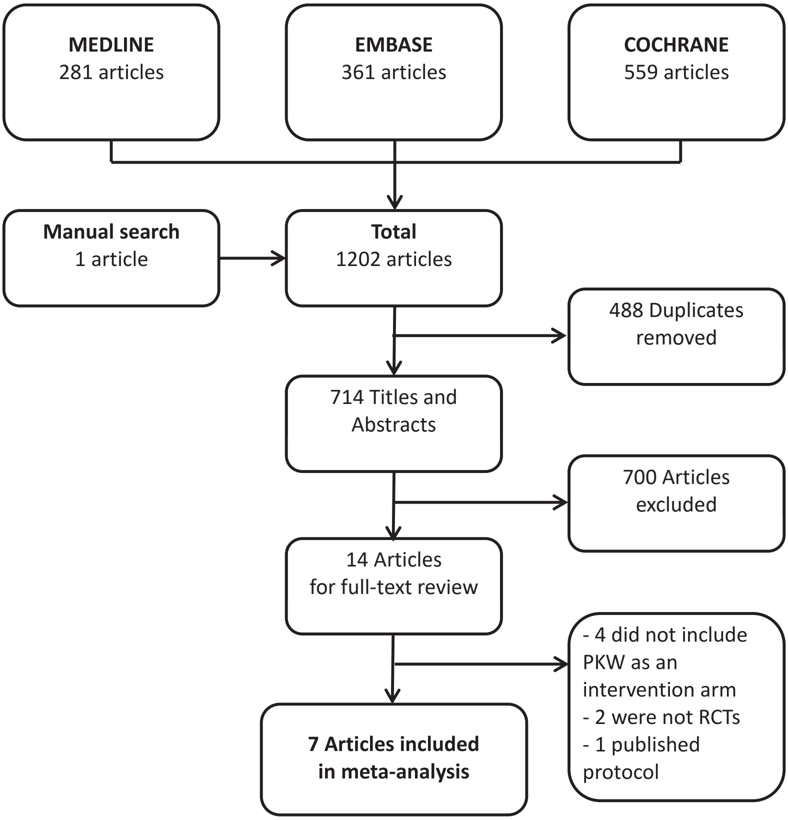

Figure 1: The flowchart shows the search and screening process for article inclusion. PKW = percutaneous K-wire fixation; RCTs = randomized controlled trials.

Publication bias

To assess publication (or positive-outcome bias), we constructed a funnel plot. Although the number of studies was small, we did not appreciate any asymmetry which would suggest publication bias (Fig. 2). A small group of positive industry-funded studies also can suggest publication bias (24). If this were the case, we would expect results to positively favor the volar locking plate, as this represents a newer technology. We therefore assessed funding sources for each study. None of the studies were funded by industry, which further reinforced the lack of a publication bias.

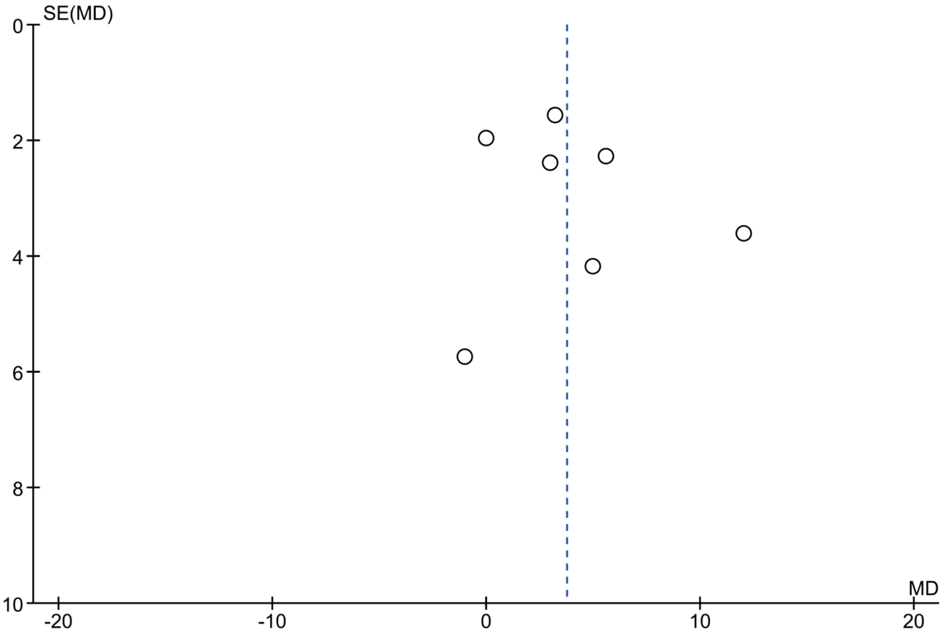

Figure 2: There is no excessive asymmetry to suggest publication bias in the funnel plot. SE = standard error; MD = mean difference.

Study characteristics

All seven studies included in this systematic review were parallel-group randomized controlled trials (Table 1). Six of these trials were conducted in Europe and one in North America. There was only one multicenter study, which was conducted in the United Kingdom and included 18 centers. Five studies reported final followup at 12 months and two studies reported final followup at 6 months. Six studies also reported intermediary followup data at 3 months or earlier. All trials reported less than 20% loss to follow-up, and in six of seven trials there was less than 10% loss to follow-up.

All trials included only dorsally displaced distal radius fractures. Six trials included patients with both extraarticular and intraarticular distal radius fractures; one trial included patients with only extraarticular fractures. Two trials included patients who received supplemental external fixation because of residual instability after K-wire (9% of patients receiving K-wire in one trial, 17% of patients in the second trial; this represented less than 3% of all patients analyzed in the K-wire group). All studies except one excluded patients with polytrauma or multiple injuries. Costa et al (5) did not explicitly exclude these high energy injuries; however, "fall" was reported as the mechanism of injury in 98% (451 of 461) of distal radius fractures included in their study.

Study	Method	Location	Surgeon Experience	Participants	Mean age (years; SD)
Rozental et al, 2009 (37)	Parallel-group, randomized controlled trial	USA	Fellowship-trained hand and upper extremity surgeons	45 total; 22 PKW, 21 VLP 2/22 (9%) required external fixation in PP group	PKW: 51 (NR), range 19-77 VLP: 52 (NR), range 24-79
Marcheix et al, 2010 (29)	Parallel-group, randomized controlled trial	France	NR	103 total; 53 PKW, 50 VLP No external fixation	PKW: 73 (11) VLP: 75 (11)
Hollevoet et al, 2011 (18)	Parallel-group, randomized controlled trial	Belgium	University hospital surgeon (either consultants or trainees supervised by consultants)	40 total; 20 PKW, 20 VLP No external fixation	PKW: 66 (NR) VLP: 67 (NR)
McFadyen et al, 2011 (30)	Parallel-group, randomized controlled trial	United Kingdom	Senior orthopaedic consultant for most cases (74% VLP; 62% PKW). Trainees or associate specialists for other cases	56 total; 29 PKW, 27 VLP No external fixation	PKW: 65 (NR), range 18-80 VLP: 61 (NR), range 26-80
Karantana et al, 2013 (20)	Parallel-group, randomized controlled trial	United Kingdom	Senior orthopaedic consultant surgeons from tertiary center	130 total; 64 PKW, 66 VLP 11/64 (17%) required external fixation in PP group	PKW: 51 (16) VLP: 48 (15)
Goehre et al, 2014 (12)	Parallel-group, randomized controlled trial	United Kingdom	Experienced senior orthopaedic surgeons	40 total; 19 PKW; 21 VLP No external fixation	PKW: 73.8 (8.9) VLP: 71.3 (5.7)
Costa et al, 2014 (5)	Parallel-group, randomized controlled trial	United Kingdom	Multiple surgeons from multiple different centers (68% had performed >20 prior VLPs; 74% had performed >20 prior PKWs)	461 total; 230 PKW; 231 VLP No external fixation	PKW: 59.7 (16.4) VLP: 58.3 (14.9)

PKW = percutaneous Kirschner wires; VLP = volar locking plate; PP = percutaneous pins; DASH = Disabilities of the Arm, Shoulder and Hand; NR = not reported; PEM = Patient Evaluation Measure; PRWE = Patient Rated Wrist Evaluation.

Fracture types in PKW group (AO classification)	Fracture types in VLP group (AO Classification	Outcomes
A2: 4 A3: 2 C1: 6 C2: 9	A2: 2 A3: 8 C1: 2 C2: 11	DASH (3 weeks, 6 weeks, 12 weeks, 1year) Wrist ROM Digit motion Grip strength Pinch strength Patient satisfaction Return to work Radiographic outcomes Complications
A2: 1 A3: 22 C2: 23 C3: 6	A2: 0 A3: 17 C2: 25 C3: 8	DASH (12 weeks, 26 weeks) Herzberg score Wrist ROM Radiographic outcomes Complications
Frequencies by AO type NR 7 extra-articular, 13 intra-articular	Frequencies by AO type NR 11 extra-articular, 9 intra-articular	DASH (3 months, 1 year) Wrist ROM Grip strength Radiographic outcomes Complications
Only Type A fractures, frequencies NR	Only Type A fractures, frequencies NR	QuickDASH (3 months, 6 months) Gartland and Werley score Wrist ROM Grip strength Radiographic outcomes Complications
A3: 28 C2: 30 C3: 6	A3: 27 C2: 37 C3: 2	QuickDASH (6 weeks, 12 weeks, 1 year) PEM EQ-5D Wrist ROM Grip strength Radiographic outcomes Complications
A2: 6 A3: 9 C1: 4	A2: 4 A3: 14 C1: 3	DASH (3 months, 6 months, 12 months) PRWE Castaing score Wrist ROM Grip strength Radiographic outcomes Complications
A2: 73 A3: 84 B1: 1 B2: 1 B3: 1 C1: 33 C2: 26 C3: 7	A2: 73 A3: 78 B1: 4 B2: 1 B3: 0 C1: 30 C2: 34 C3: 11	DASH (12 months) PRWE EQ-5D Complications

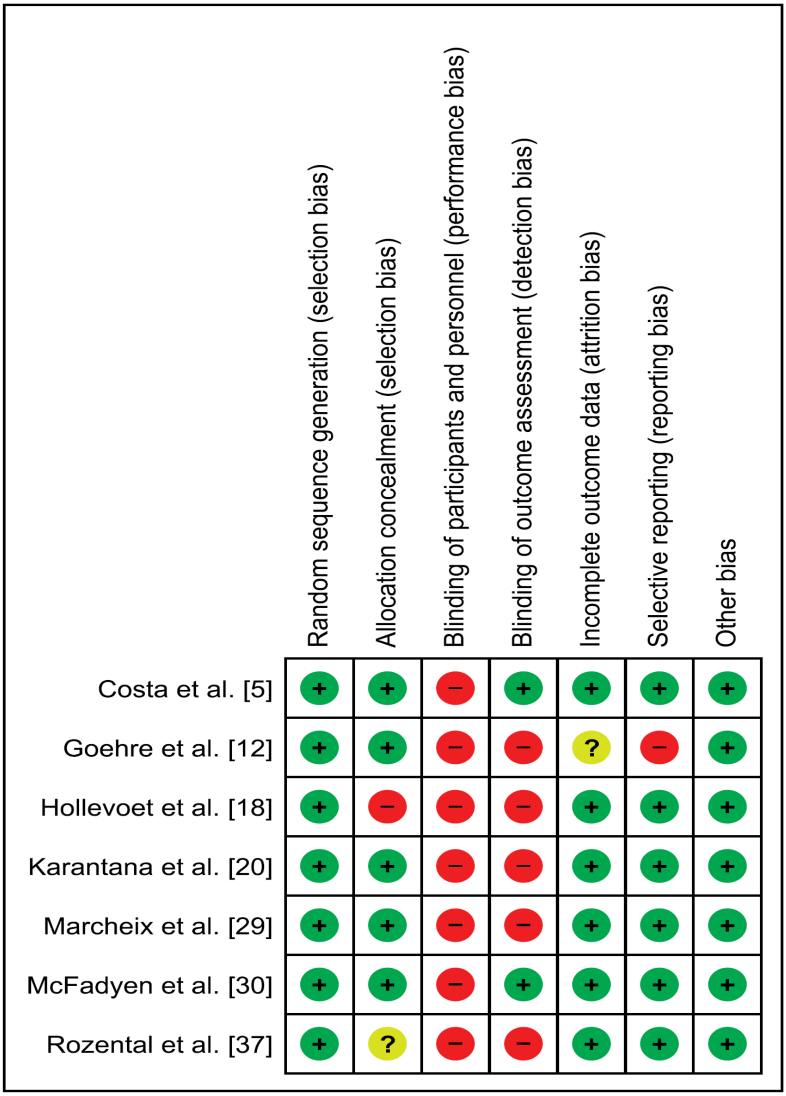

Figure 3: The risk of bias for each trial included in the meta-analysis using the Cochrane Risk of Bias tool is shown.

Risk of Bias

None of the trials reported any attempt to blind surgeons or patients. Outcome assessors were blinded in two trials. One study had a risk of selective reporting bias. The majority of trials were at low risk of bias in terms of random sequence generation, allocation concealment, completeness of followup, selective reporting, or other biases (Fig. 3).

Study or Subgroup	Percutaneous K-Wires Mean	SD	Total	Volar Plating Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Goehre et al. [12]	15.8	10.6	19	15.8	16.5	21	13.0%	0.00 [-8.52, 8.52]
Hollevoet et al. [18]	27	24	20	21	21	20	4.8%	6.00 [-7.98, 19.98]
Karantana et al. [20]	27	20	64	21	17	66	23.2%	6.00 [-0.39, 12.39]
Marcheix et al. [29]	33	22	53	25	21	50	13.7%	8.00 [-0.30, 16.30]
McFadyen et al. [30]	27.2	9.6	29	18.3	9.6	27	37.4%	8.90 [3.87, 13.93]
Rozental et al. [37]	26	23	22	11	13	23	7.8%	15.00 [4.02, 25.98]
Total (95% CI)			**207**			**207**	**100.0%**	**7.28 [4.21, 10.36]**

Heterogeneity: Chi² = 5.32, df = 5 (P = 0.38); I² = 6%
Test for overall effect: Z = 4.64 (P < 0.00001)

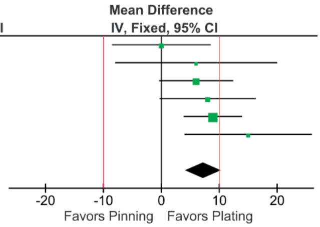

Figure 4: The individual and pooled 3-month mean differences in DASH scores and 95% CIs are shown in the forest plot. The minimum clinically important difference is indicated by the red lines. IV = inverse variance

RESULTS

Functional Outcomes

Patients treated with volar locking plates had slightly better DASH scores than did patients treated with K-wires at 3 months and at final followup. At 3-month followup, the mean DASH score was 7.5 points lower (i.e., better) in the volar locking plate group (six trials, 414 participants; 95% confidence interval [CI], 4.4-10.6; p < 0.001) (Fig. 4). The upper threshold of the 95% CI crossed the MCID of 10, and therefore we were unable to rule out a clinically important difference at 3 months. At final followup (6 or 12 months), the mean DASH score was only 3.8 points lower in the volar locking plate group (seven trials, 875 participants; 95% CI, 1.2-6.3; p = 0.004) (Fig. 5). The upper threshold of the 95% CI was below 10, suggesting that this was unlikely to be a clinically important difference at final followup. These conclusions were consistent when SMDs were used for analysis (3 month SMD, 0.42; 95% CI, 0.22-0.61, p < 0.001); 6- or 12-month SMD, 0.29, 95% CI, 0.11-0.46, p = 0.001). Removal of the two trials from the analysis that did not follow patients to 1 year (i.e., 6-month followup only) did not substantially change the results (mean difference [MD], 2.3; 95% CI, 0.3-4.4, p = 0.03). Based on GRADE, there was low confidence in the 3-month estimate of effect and moderate confidence in the 6- to 12-month estimate of effect (Table 2). There were a total of 875 participants for which these data were available. There were no differences found between K-wires and volar locking plates in terms of the PRWE score at either 3 months or final followup (6 months or 1 year) in either of the two trials that reported on this endpoint (5, 12).

6

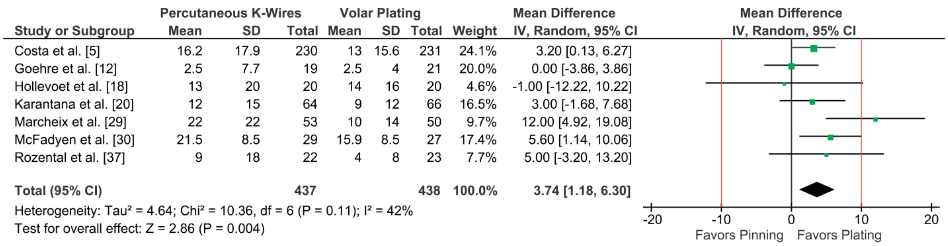

Figure 5: The individual and pooled 6- and 12-month mean differences in DASH scores, along with 95% CIs, are shown in the forest plot. The minimum clinically important difference is indicated by the red lines. IV = inverse variance.

Table 2. GRADE summary of findings
Volar locking plate compared with percutaneous Kirschner wires for displaced distal radius fracture Bibliography (systematic reviews)

Outcomes	Number of participants (studies) followup	Quality of the evidence (GRADE)	Anticipated absolute effects	
			Risk with percutaneous K-wires	Risk difference with volar locking plate
Function at 3 months (function at 3 months) assessed with: DASH followup: 3 months	414 (6 RCTs) 3 months	⊕⊕○○ Low[1,2]	The mean function at 3 months in the control group was 27.4	MD 7.5 lower (4.4 lower to 10.6 lower)
Function at 6-12 months (final function) assessed with: DASH followup: range 6-12 months	875 (7 RCTs) 6-12 months	⊕⊕⊕○ Moderate[1]	The mean function at 6-12 months in the control group was 15.5	MD 3.8 lower (1.2 lower to 6.3 lower)

[1] Lack of blinding of outcome assessors in most trials; [2] High imprecision in pooled estimate; the risk in the intervention group (and its 95% CI) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI); RCTs = randomized controlled trials; MD = mean difference. GRADE Working Group grades of evidence: High quality = We are very confident that the true effect lies close to that of the estimate of the effect; Moderate quality = We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different; Low quality = Confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect; Very low quality = We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Wrist ROM

Flexion and supination were both slightly greater in the volar locking plate group at 3 months (four trials, 3.7° [95% CI 0.3°-7.1°, p = 0.04] and 4.1° [95% CI 0.6°-7.6°, p = 0.02] greater, respectively), but not at final followup. There were no differences in wrist extension, pronation, radial deviation, or ulnar deviation at 3 months or final followup. Wrist ROM was reported as an outcome in five trials (12, 18, 21, 29, 37), four of which reported sufficient information to enable pooling. Radial deviation and ulnar deviation were reported in only two of these trials (12, 37).

Radiographic Outcomes

There were no differences in radiographic outcomes at the latest reported followup between the two interventions. Among trials that reported sufficient information for meta-analysis, there were no differences in volar tilt (four trials, 0.1° greater in K-wire group; 95% CI, -4.6° to 4.9°; p = 0.96), radial inclination (four trials, 0.4° greater in K-wire group; 95% CI, -0.9° to 1.7°; p = 0.58), or radial height (three trials, 0.4 mm greater in K-wire group; 95% CI, -0.3 mm to 1.0 mm; p = 0.31) at final follow-up. Of the two trials that could not be included in meta-analysis, one reported better volar tilt, radial height, and radial inclination with the volar locking plate, but did not report absolute values (30); the other trial reported a greater median volar tilt in the K-wire group (4° vs 0° in volar locking plate group) but did not provide any data to estimate variance (e.g. SD, interquartile range) or statistically analyze the data. Only one study reported articular step-off postoperatively (21), and it detected no difference between the two interventions. Radiographic outcomes were reported in six of the seven trials.

Table 3. Common complications

Complication	Percutaneous Kirschner wires	Volar locking plate	p value
Superficial infection	36	14	0.001
Deep infection	2	2	1.00
Neurological injury* (carpal tunnel)	33 (22)	32 (28)	0.89 (0.39)
Tendon rupture†	6	6	1.00
Reoperations	14	17	0.59

*Includes transient nerve palsies; †does not include tendinitis

Complications

There were more total complications in the K-wire group than in the volar locking plate group. This difference was driven predominantly by a difference in superficial wound infections (8.2% versus 3.2%; RR = 2.6; p = 0.001), all of which were successfully treated with oral antibiotics. There were no differences in the risks of any of the other reported complications (deep infection, neurologic injury, tendon rupture, or re-operations) between the two groups (Table 3). In total, there were 102 complications reported in the K-wire group and 66 reported in the volar locking plate group. All trials reported complications.

DISCUSSION

Distal radius fractures are common and costly injuries (4, 32, 39). In the setting of unstable dorsally displaced fractures requiring surgical intervention, the optimal surgical treatment option remains equivocal. Clinical practice guidelines have bemoaned the lack of high-quality evidence to inform orthopaedic practice in this area (22, 25). Despite this lack of evidence, there has been a large shift in the treatment of dorsally displaced distal radius fractures toward the use of the volar locking plates, especially among younger orthopaedic surgeons (20, 23). To our knowledge, there have been no published meta-analyses to date comparing volar locking plates with K-wires for dorsally displaced distal radius fractures. In our meta-analysis of 875 patients, we found lower (i.e., better) DASH scores with use of volar locking plates at both 3 months and 12 months. Although we cannot exclude the possibility of a small clinically important difference at 3 months, the magnitude of improvement by 12 months is most likely imperceptible to patients.

An important limitation of our review is that followup of all included trials was limited to a maximum of 1 year and in some trials just 6 months. This followup interval is not long enough for development of posttraumatic arthritis, one of the long-term complications of a malreduced articular surface. One of the potential advantages of volar plating is that the fracture can be reduced under direct observation leading to more accurate articular reduction in AO Type B and C fractures. Studies with longer-term followup will be necessary to determine whether there is a difference in clinical symptoms of posttraumatic arthritis between these two treatment modalities. Included trials, in general, had low risk of bias—with the exception of blinding, which is difficult given the nature of the interventions. However, given that both interventions were surgical, the presence of a “placebo bias” is less likely. The inclusion of patients with adjunctive external fixation in the K-wire group is a potential limitation. However, only two trials included patients with external fixation and a small proportion required this adjunct (< 3% of all patients treated with K-wires). If K-wires alone would have led to a poorer outcome in these patients treated with adjunctive external fixation, then this meta-analysis may potentially be underestimating the benefit of the volar locking plate in some situations. However, given the small number of patients, the degree of this underestimate is expected to be minimal. Differences resulting from adjunctive treatments that were not reported consistently across trials (e.g., use of bone graft) is another limitation to this meta-analysis.

Another important limitation relates to the external validity—or generalizability—of the findings of this review. The results of this meta-analysis are most applicable to the low-to-moderate energy dorsally displaced distal radius fracture (with or without an intraarticular component), which is reducible under fluoroscopy and allows for good purchase of bone with K-wires. Extreme cases of either high-energy trauma (e.g., motor vehicle accidents) or very low-energy trauma in patients with osteoporosis are either underrepresented or excluded entirely in the trials constituting our meta-analysis. Therefore we cannot make any definitive conclusions regarding these subgroups.

We found that use of volar locking plates for displaced distal radius fractures showed a small improvement in DASH scores at 3 months (MD, 7.5; 95% CI, 4.4-10.6; $p < 0.001$) and 12 months (MD, 3.8; 95% CI, 1.2-6.3; $p = 0.004$) compared with K-wires. Uncertainty in the estimate precludes the conclusion that there is no clinical advantage at 3 months postoperatively; however, by 1 year the magnitude of this difference was less than our *a priori*-established MCID on the DASH scale of 10 points (10, 38). This represents the best (i.e. lowest risk of bias) estimate of functional differences in the literature to date, as we were able to pool the results of seven recent and good quality RCTs to achieve a large sample size. Our endeavor was facilitated by the use of a common and recommended functional outcome instrument across all RCTs, the DASH questionnaire. Inconsistent and varying use of outcome instruments has presented limitations to previous meta-analyses in the orthopaedic and distal radius literature (2, 17). Future trials must continue to use common outcome instruments to allow for meaningful meta-analysis.

Our analysis also found small early advantages in flexion and supination in the volar locking plate group (3.7° and 4.1° greater, respectively) at 3-month followup, but these differences disappeared at final followup. Not all trials standardized postoperative protocols for both groups (e.g., patients treated with volar locking plates were allowed to mobilize at 1 week in three trials versus 6 weeks for patients with K-wires), which may have contributed to the finding that volar locking plate leads to improved DASH scores, flexion, and supination at 3 months. However, patients treated with volar locking plates typically are permitted earlier mobilization (42), and it would be reasonable to expect this to contribute to some the early advantage in ROM. Furthermore, the three trials that standardized postoperative protocols showed possible early improvements as well; therefore, the early improvements seen with volar locking plates may not be entirely attributable to reduced immobilization times as has been suggested (27).

There were no differences in radiographic alignment (volar tilt, radial inclination, and radial height, or articular incongruity) between the two interventions. The relationship between radiographic outcomes—including articular incongruity—and clinical outcomes is controversial (11). In terms of short-term outcomes, small amounts of radial shortening (as little as 3 mm) have been shown to negatively affect function (1, 41). However, the distal radius appears to be relatively tolerant to changes in volar tilt, with no apparent functional deficits with even a small amount of dorsal angulation (44). In terms of long-terms outcomes, an articular step-off of 2 mm has been shown to result in radiographic signs of arthritis. However, this has not consistently translated into poorer clinical outcomes (6, 11). Therefore, small differences in radiographic outcomes are likely not clinically important.

Superficial infections were more frequent in patients treated with K-wires, but otherwise no differences in complication rates were found between the two treatments. It has been argued that in the absence of convincing evidence of superiority of volar locking plates, economic considerations should drive clinical decision-making and policy in the treatment of dorsally displaced distal radius fractures (8, 9, 40). However, a robust economic analysis will need to consider differences in costs associated with complications (e.g., antibiotic treatment for superficial infections) in addition to differences in costs of the implants, length of surgery, requirement for adjunctive treatments (e.g., external fixation, casting), and postoperative protocols (e.g. clinic visits, radiographs) (8).

CONCLUSION

We found that volar locking plates result in lower (i.e., better) DASH scores compared with K-wires for dorsally displaced distal radius fractures in adults. However, these differences were small, and likely to have been imperceptible to the patient, since they were smaller than the pre-defined MCID. Further research is required to better delineate if there are specific radiographic, injury, or patient characteristics that may benefit from volar locking plates in the short term. Further, the incidence of post-traumatic arthritis would not have been detected at the short-term followups in the studies included in this meta-analysis. Therefore, future research must evaluate if there are any differences in outcomes and complications between these two interventions in the long term.

REFERENCES

(1) Aro HT, Koivunen T. Minor axial shortening of the radius affects outcome of Colles' fracture treatment. J Hand Surg Am.1991; 16:392-398.

(2) Chaudhry H, Simunovic N, Petrisor B. Cochrane in CORR: surgical versus conservative treatment for acute injuries of the lateral ligament complex of the ankle in adults (review). Clin Orthop Relat Res. 2014 Oct 25 [Epub ahead of print].

(3) Chung KC, Shauver MJ, Yin Y, Kim M, Baser O, Birkmeyer JD. Variations in the use of internal fixation for distal radius fracture in the United States Medicare population. J Bone Joint Surg Am. 2011;93:2154-2162.

(4) Chung KC, Spilson SV. The frequency and epidemiology of hand and forearm fractures in the United States. J Hand Surg Am. 2001;26:908-915.

(5) Costa ML, Achten J, Parsons NR, Rangan A, Griffin D, Tubeuf S, Lamb SE. Percutaneous fixation with Kirschner wires versus volar locking plate fixation in adults with dorsally displaced fracture of distal radius: randomized controlled trial. BMJ. 2014;349:g4807.

(6) Diaz-Garcia RJ, Chung KC. Common myths and evidence in the management of distal radius fractures. Hand Clin. 2012;28:127-133.

(7) Dincer F, Samut G. Chapter 2: Physical examination of the hand. In: Hand Function. Duruoz MT (Ed.). Springer, New York, NY; 2014.

(8) Dzaja I, MacDermid JC, Roth J, Grewal R. Functional outcomes and cost estimation for extra-articular and simple intra-articular distal radius fractures treated with open reduction and internal fixation versus closed reduction and percutaneous Kirschner wire fixation. Can J Surg. 2013;56:378-384.

(9) Farner S, Malkani A, Lau E, Day J, Ochoa J, Ong K. Outcomes and cost of care for patients with distal radius fractures. Orthopedics. 2014;37:e866-e878.

(10) Franchignoni F, Vercelli S, Giordano A, Sartorio F, Bravini E, Ferriero G. Minimal clinically important difference of the Disabilities of the Arm, Shoulder and Hand outcome measure (DASH) and its shortened version (QuickDASH). J Orthop Sports Phys Ther. 2014;44:30-39.

(11) Giannoudis PV, Tzioupis C, Papathanassopoulos A, Obakponovwe O, Roberts C. Articular step-off and risk of post-traumatic osteoarthritis. Evidence today. Injury. 2010;41:986-995.

(12) Goehre G, Otto W, Schwan S, Mendel T, Vergroesen PP, Lindemann-Sperfeld L. Comparison of palmar fixed-angle plate fixation with K-wire fixation of distal radius fractures (AO A2, A3, C1) in elderly patients. J Hand Surg. 2014;39:249-257.

(13) Goldhahn J, Beaton D, Ladd A, Macdermid J, Hoang-Kim A; Distal Radius Working Group of the International Society for Fracture Repair (ISFR); International Osteoporosis Foundation (IOF). Recommendation for measuring clinical outcome in distal radius fractures: a core set of domains for standardized reporting in clinical practice and research. Arch Orthop Trauma Surg. 2014;134:197-205.

(14) GraphClick [computer program]. Arizona software. Available at: www.arizona-software.ch/graphclick; 2010. Accessed November 3, 2014.

(15) Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64:383-394.

(16) Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savovich J, Schulz KF, Weeks L, Sterne JA; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.

(17) Hoang-Kim A, Scott J, Micera G, Orsini R, Moroni A. Functional assessment in patients with osteoporotic wrist fractures treated with external fixation: a review of randomized trials. Arch Orthop Trauma Surg. 2009;129:105-111.

(18) Hollevoet N, Vanhoutte T, Vanhove W, Verdonk R. Percutaneous K-wire fixation versus palmar plating with locking screws for Colles' fracture. Acta Orthop Belg. 2011;77:180-187.

(19) Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMCMed Res Methodol. 2005;5:13.

(20) Hull P, Baraza N, Whalley H, Brewster M, Costa M. Dorsally displaced fractures of the distal radius - a study of preferred treatment options among UK trauma and orthopaedic surgeons. Hand Surg. 2010;15:185-191.

(21) Karantana A, Downing ND, Forward DP, Hatton M, Taylor AM, Scammell BE, Moran CG, Davis TR. Surgical treatment of distal radial fractures with a volar locking plate versus conventional percutaneous methods: a randomized controlled trial. J Bone Joint Surg Am. 2013;95:1737-1744.

(22) Koval K, Haidukewych GJ, Service B, Zirgibel BJ. Controversies in the management of distal radius fractures. J Am Acad Orthop Surg. 2014;22:566-575.

(23) Koval KJ, Harrast JJ, Anglen JO, Weinstein JN. Fractures of the distal part of the radius. The evolution of practice over time. Where's the evidence? J Bone Joint Surg Am. 2008 Sep;90:1855-1861.

(24) Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ. 2003;326:1167-1170.

(25) Lichtman DM, Bindra RR, Boyer MI, Putnam MD, Ring D, Slutsky DJ, Taras JS, Watters WC III, Goldberg MJ, Keith M, Turkelson CM, Wies JL, Haralson RH III, Boyer KM, Hitchcock K, Raymond L. AAOS clinical practice guideline summary: treatment of distal radius fractures. J Am Acad Orthop Surg. 2010;18:180-189.

(26) Li-hai Z, Ya-nan W, Zhi M, Li-cheng Z, Hong-da L, Huan Y, Xiao-xie L, Pei-fu T. Volar locking plate versus external fixation for the treatment of unstable distal radial fractures: a meta-analysis of randomized controlled trials. J Hand Surg Eur Vol. 2014;39:249-257.

(27) Lozano-Calderon SA, Souer S, Mudgal C, Jupiter JB, Ring D. Wrist mobilization following volar plate fixation of fractures of the distal part of the radius. J Bone Joint Surg Am. 2008;90:1297-1304.

(28) Magee DJ. Chapter 7: Forearm, wrist and hand. In: Orthopaedic physical assessment. Magee DJ (Ed.). Springer, New York, NY; 2008.

(29) Marcheix P-S, Dotzis A, Benko P-E, Siegler J, Arnaud J-P, Charissoux J-L. Extension fractures of the distal radius in patients older than 50: a prospective randomized study comparing fixation using mixed pins or a palmar fixed-angle plate. J Hand Surg. 2010;35:646-651.

(30) McFadyen I, Field J, McCann P, Ward J, Nicol S, Curwen C. Should unstable extra-articular distal radial fractures be treated with fixed-angle volar-locked plates or percutaneous Kirschner wires? A prospective randomised controlled trial. Injury. 2011;42:162-166.

(31) Moher D, Liberatri A, Tetzlaff J, Altman DG; the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;338:b2535.

(32) Nana AD, Lichtman DM. Plating of distal radius. J Am Acad Orthop Surg. 2005;13:159-171.

(33) Nellans KW, Kowalski E, Chung KC. The epidemiology of distal radius fractures. Hand Clin. 2012;28:113-125.

(34) Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care. 2003;41:582-592.

(35) Pinto R, Maher CG, Ferreira ML, Ferreria PH, Hancock M, Oliveira VC, et al. Drugs for relief of pain in patients with sciatica: systematic review and meta-analysis. BMJ. 2012;344.

(36) Review Manager (RevMan) Version 5.3. Copenhagen, The Netherlands: The Nordic Cochrane Centre, The Cochrane Collaboration; 2014. Available at http://tech.cochrane.org/revman. Accessed April 30, 2015

(37) Rozental TD, Blazar PE, Franko OI, Chacko AT, Earp BE, Day CS. Functional outcomes for unstable distal radial fractures treated with open reduction and internal fixation or closed reduction and percutaneous fixation: a prospective randomized trial. J Bone Joint Surg Am. 2009;91:1837-1846.

(38) 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.

(39) Shauver MJ, Yin H, Banerjee M, Chung KC. Current and future national costs to medicare for the treatment of distal radius fracture in the elderly. J Hand Surg. 2011;36:1282–1287.

(40) Shyamalan G, Theokli C, Pearse Y, Tennent D. Volar locking plates versus Kirschner wires for distal radial fractures--a cost analysis study. Injury. 2009;40:1279-1281.

(41) Slutsky DJ. Predicting the Outcome of Distal Radius Fractures. Hand Clin. 2005;21:289-294.

(42) Smith DW, Brou KE, Henry MH. Early active rehabilitation for operatively stabilized distal radius fractures. J Hand Ther. 2004;17:43-49.

(43) US Department and Health & Human Services, Agency for Healthcare Research and Quality. Welcome to H-CUPnet. Available at: http://hcupnet.ahrq.gov/. Accessed November 3, 2014.

(44) Warwick D, Field J, Prothero D, Gibson A, Bannister GC. Function ten years after Colles' fracture. Clin Orthop Relat Res. 1993;(295):270-274.

(45) Wei DH, Poolman RW, Bhandari M, Wolfe VM, Rosenwasser MP. External fixation versus internal fixation for unstable distal radius fractures: a systematic review and meta-analysis of comparative clinical trials. J Orthop Trauma. 2012;26:386-394.

SUMMARY AND FUTURE PERSPECTIVES

Distal radius fractures (DRFs) are the most common fracture seen in the emergency department. Unfortunately, optimal management for these injuries is far from certain. The clinical practice guideline for distal radius fractures published by the American Academy of Orthopaedic Surgeons (AAOS) has 29 recommendations; however, none of these recommendations were given a "strong" rating owing to limited strength of the evidence. The Dutch guidelines identified the same issues with the quality of evidence and partly relied on expert opinion to achieve consensus. Therefore, determining effective evidence-based treatment of patients with DRFs is crucial. To conduct best evidence clinical trials in DRF treatment, and to properly compare trial results however, there must be a consensus on the use of inclusion criteria and of outcome measures.

This thesis aimed to investigate a critical component of the methodology used in DRF research, which might help us to improve the quality of future research. In this thesis we therefore focused on an underappreciated aspect of methodology: assessment tools used for diagnosis and trial eligibility in patients with DRFs, and the quality of patient-reported outcome measures (PROMs).

To this end, we conducted reliability studies, two different types of systematic reviews and performed a clinimetric study focussing on inclusion criteria and outcome measures used in DRF research.

In **Chapter 1** we answered the following question:

- What is the influence of computed tomography on the reliability and accuracy of classification systems for distal radius fractures?

We performed a reliability study to determine the intra- and inter-observer reliability of four of the most commonly used DRF classification systems. The reliability of these classification systems using conventional radiography versus conventional radiography with an additional CT scan was compared. Our study results suggest that the additional value of CT scanning over conventional radiographs is limited in regard to reliability. However, it has significant implications for accurate scoring of the fracture types. Obtaining an additional CT scan changes how patients are classified into fracture types. Therefore, trials using conventional radiography alone to evaluate eligibility will have different patients included compared to trials using additional CT scans. This has implications for external validity (generalizability) and for comparing trials to each other.

Our findings indicate the need for a more reliable classification system, ideally based on CT scans rather than conventional radiography alone. This new classification system should also focus on giving direction about the type of treatment.

In **Chapter 2** we answered the following question:

- Can experienced surgeons predict the additional value of a CT scan in patients with displaced intra articular distal radius fractures?

In a prospective study we evaluated whether surgeons can predict the additional value of a CT scan by asking them to decide whether they would obtain a CT scan to choose their treatment plan at multiple time points. In Addition, we asked surgeons if the CT scan was useful for treatment choice and/or pre-operative planning. We determined what the implications were of always ordering an additional CT scan for displaced intra-articular DRFs or leaving the decision to the surgeon's discretion. For choice of treatment, leaving the choice up to the surgeon led to fewer scans needed and fewer unnecessary risks to the patient. Therefore, we do not recommend always ordering an additional CT scan in these patients as a standard practice. However, once a surgeon decides to treat a patient surgically, it is much harder to predict whether an additional CT scan is useful for pre-operative planning. Therefore, we cannot recommend implementing rules for or against CT scanning in these patients.

That surgeons can predict the usefulness of an additional CT scan is not necessarily related to better patient outcomes. Future randomized trials should be conducted, comparing two groups:

1. Ordering CT scans for all patients with intra-articular displaced DRF.
2. CT scanning on request of the surgeon.

Firstly, to confirm the results of this study, secondly to follow patients to determine which decision making model improves patient outcomes. Future studies should also include cost-effectiveness analyses.

In **Chapters 3 and 4** we answered the following question:

- Are validated outcome measures used in distal radius fracture studies truly valid?

We first conducted a systematic review of methodological quality of clinimetric studies that evaluated measurement properties of patient-reported outcome measures (PROMs) used in DRF patients. We evaluated both the methodological quality and the results of the different clinimetric studies to create a summary of the level of evidence for each measurement property (e.g. reliability, responsiveness, internal consistency) for each PROM. Our results showed that the most studied measurement properties were those of the Patient-Rated-Wrist-Evaluation (PRWE), followed by the Disability of the Arm, Shoulder and Hand questionnaire (DASH). However, the body of literature lacked strong evidence supporting good quality of the available PROMs in patients with DRFs. A good PROM should have a high level of evidence (i.e. be evaluated in high quality studies) supporting good quality on all measurement properties.

Our systematic review found limited to no evidence for any of the measurement properties of the PRWE and the DASH. This led us to perform a multi-centre prospective clinimetric study evaluating internal consistency, test-retest reliability, measurement error and content validity, following the COSMIN standards to ensure high methodological quality. Based on our results we now have strong evidence that both the PRWE and the DASH questionnaires have good content validity and are reliable and internally consistent instruments for the assessment of patients with distal radius fractures. Although the conclusions about these measurement properties did not change, we now have stronger evidence supporting them. However, in our systematic review, previous studies found limited evidence that the measurement error is acceptable, with a smallest detectable change (SDC) of 4.4 to 11 points for the PRWE. In our study we found a much higher SDC of 20 points, indicating that the quality of the clinimetric study can have major implications for the conclusions. We also have high quality evidence that the SDC of the DASH is 14 points. The PRWE's minimal clinically important difference (MCID) from previous (low quality) literature is smaller than the SDC, so it is difficult to distinguish important changes from measurement error when applied to individual patients. However, it is still useful for research purposes because measurement error is reduced when large groups of patients are studied.

Future studies using high quality methodology should focus on evaluating the measurement property responsiveness, defined as the ability to detect clinically important changes over time. MCID is an important aspect of responsiveness that deserves special attention. The available information we currently have for MCID of the DASH and the PRWE is either based on low quality evidence or only evaluated in pathologies other than DRFs.

In **Chapter 5** we answered the following question:

- What is the influence of spectrum bias on the intra-observer agreement of distal radius fracture treatment plans?

We performed a reliability study to determine the influence of spectrum bias on the intra-observer agreement of DRF treatment plans. Spectrum bias is the bias that occurs when studies include a population lacking therapeutic uncertainty. Therefore, we determined the influence of the surgeons' level of therapeutic uncertainty on the level of intra-observer agreement in treatment plans in patients with displaced DRFs using conventional radiographs alone or with an additional CT scan. Without taking spectrum bias into account, the intra-observer reliability decreased when using an additional CT scan. When we corrected for spectrum bias in our study, our conclusions changed. In cases with therapeutic uncertainty, the intra-observer agreement on treatment plan improved when an additional CT scan was used. In cases without therapeutic uncertainty we showed that an additional CT scan even diminished the agreement. This study clearly showed that the CT scan can indeed improve the intra-observer agreement on treatment plan, but only when there is therapeutic uncertainty. This is another clear example where conflicting results in the literature can be explained by methodological limitations.

To address the current controversies in the additional value of CT scans for agreement in treatment plans in other types of fractures we suggest that future studies take therapeutic uncertainty into account to minimize spectrum bias. Future prospective randomized trials should be conducted to determine whether the use of an additional CT scan and their resulting management, in cases of therapeutic uncertainty, influences outcomes in patients with displaced DRFs.

In **Chapter 6** we answered the following question:

- Are volar locking plates superior to percutaneous K-wires for dorsally displaced distal radius fractures?

We conducted a systematic review and meta-analysis of percutaneous Kirschner wire fixation and plaster cast (PKW) compared to volar locking plate (VLP) for displaced DRFs. In all studies included in this review, conventional radiographs were used to classify the fractures according to the AO classification and the DASH scores were used as the outcome score. This showed, with moderate confidence, that VLP demonstrates better DASH scores at both 3-month and final follow-up (6-12 months) compared to PKW for displaced distal radius fractures in adults; however, these differences did not cross the a priori determined threshold for clinical importance.

Based on these results, further research is required to better delineate if there are specific radiographic (eg intra articular fractures) or patient characteristics (eg elderly) that may benefit from volar locking plates in the short-term. The incidence of post-traumatic arthritis would not have been detected at the short-term follow-up intervals used in the studies included in this meta-analysis. Using volar plating is expected to result in a lower rate of post-traumatic arthritis. Long term studies are needed to effectively investigate the influence of post-traumatic arthritis on patient outcomes.

On the surface (e.g. looking at the Cochrane risk of bias summary) it appears that the results of these studies are of high quality, and surgeons may use these findings in practice. However, we argue that it is prudent to look more closely at methodological limitations of published research, especially the underappreciated and under-recognized methodological limitations discussed in this thesis. As shown in this thesis, using CT scans instead of conventional radiographs alone changes how patients are classified into fracture types. Looking critically at the included studies, we can see that it is likely that the results are imprecise or biased because conventional radiographs alone were used to classify fracture types to determine trial eligibility. Therefore, trials using conventional radiography alone to evaluate eligibility will have different patients included compared to trials using additional CT scans. This has implications for generalizability and applicability in clinical practice.

The included studies were conducted before the DASH was validated for use in DRF patients so they were at risk of imprecision or having biased results and therefore potentially misleading conclusions. However, in this thesis we demonstrated that the DASH is valid and reliable in this population, which improves our confidence in the conclusions of this review.

Future implications

Many clinical questions in DRF treatment remain unresolved and unanswered by this thesis. However, to answer these questions we need studies based on high quality methodology. In this thesis we identified and discussed several underappreciated and under-recognised methodological issues in DRF research. We recommend that investigators be aware of these limitations of past DRF research in order to improve the quality of the DRF literature. We aim to take our own advice in the next phase of our research program.

For our next study, we will conduct a randomized controlled trial where we aim to avoid some of the limitations of the literature identified in this thesis. Based on the evidence from this thesis, our protocol for the upcoming Effectiveness of Surgery Versus Casting for Elderly Patients with Displaced Intra-articular Radius Fractures Trial (DART Trial) will have the following features: 1) inclusion criteria using fracture characteristics instead of a fracture classification system; 2) the use of an additional CT scan will be up to the treating surgeon; and 3) the validated DASH and PRWE will be used as outcome measures. In our quest for high quality evidence for treating DRFs, we aim to take two steps forward without taking one step back.

NEDERLANDSE SAMENVATTING EN TOEKOMSTPERSPECTIEF

Distale radius fracturen (DRFs) zijn de meest voorkomende fracturen die worden gezien op de Spoed Eisende Hulp. Tot op heden is er nog geen consensus bereikt over de beste behandeling van dit type fractuur. De richtlijnen van de American Academy of Orthopaedic Surgeons (AAOS) voor het beleid rondom distale radiusfracturen bevat 29 aanbevelingen; echter al deze aanbevelingen konden slechts gedaan worden op basis van beperkt wetenschappelijk bewijs. Bij het opstellen van de Nederlandse richtlijnen werd vanwege gebrek aan wetenschappelijk bewijs mede gebruik gemaakt van “expert opinion” om consensus te bereiken. Het verkrijgen van hoogstaand wetenschappelijk bewijs voor de behandeling van DRFs is dan ook van cruciaal belang. Om onderzoeken van hoge kwaliteit naar de behandeling van DRFs uit te voeren, en ze onderling te kunnen vergelijken, moet er consensus bestaan over de te hanteren inclusie criteria en de uitkomstmaten.

Het doel van dit proefschrift is om methodologieën die gebruikt worden in wetenschappelijk onderzoek naar DRF kritisch te evalueren, wat mogelijk leidt tot verbetering van de kwaliteit van toekomstig onderzoek. In dit proefschrift hebben we ons gefocust op ondergewaardeerde aspecten van de methodologie in DRF onderzoek: instrumenten die gebruikt worden om vast te stellen of een DRF patiënt in aanmerking komt voor een studie (*eligibility*), en de kwaliteit van de patiënt gerapporteerde uitkomstmaten (PROMs).

In dit proefschrift hebben we *reliability* studies, twee verschillende type systematische reviews en een klinimetrische studie uitgevoerd waarbij we ons focusten op de inclusie criteria en de uitkomstmaten die gebruikt worden in wetenschappelijk onderzoek naar DRFs

In **Hoofdstuk 1** hebben we de volgende vraag beantwoord:

- Wat is de invloed van computertomografie (CT scan) op de betrouwbaarheid en accuraatheid van de classificatie systemen voor distale radius fracturen?

We hebben een *reliability* studie uitgevoerd om de intra- en interbeoordelaars betrouwbaarheid van de vier meest gebruikte DRF classificatie systemen vast te stellen. We vergeleken de betrouwbaarheid van deze classificatiesystemen gebruikmakend van conventionele röntgenfoto’s versus conventionele röntgenfoto’s met een toegevoegde CT scan. De resultaten van de studie laten zien dat de toegevoegde waarde van een CT scan ten opzichte van conventionele röntgenfoto’s beperkt is wat betreft de betrouwbaarheid van de classificatiesystemen. Echter, het heeft significante consequenties voor de accuraatheid van het scoren van de verschillende type fracturen. Een toegevoegde CT scan verandert de classificering van het type fractuur. Studies die alleen conventionele röntgenfoto's gebruiken om te bepalen of een DRF patiënt in aanmerking komt voor een studie, includeren dus een andere groep DRF patiënten dan studies die een aanvullende CT scan gebruiken. Dit heeft consequenties voor zowel de externe validiteit (generaliseerbaarheid) als de mogelijkheid om studies met elkaar te kunnen vergelijken.

Onze bevindingen tonen aan dat er behoefte bestaat aan een betrouwbaarder classificatiesysteem, welke idealiter gebaseerd is op CT scans in plaats van conventionele röntgenfoto's alleen. Dit nieuwe classificatiesysteem zou ook richting moeten geven aan het behandelplan.

In **Hoofdstuk 2** hebben we de volgende vraag beantwoord:

- Kunnen ervaren chirurgen de toegevoegde waarde voorspellen van een CT scan bij patiënten met gedisloceerde intra-articulaire distale radius fracturen?

In een prospectieve studie hebben we geëvalueerd of chirurgen de toegevoegde waarde kunnen voorspellen van een CT scan. Ten eerste hebben we op meerdere momenten gevraagd of ze een CT scan zouden aanvragen om hun behandelplan op te stellen. Daarna hebben we ze gevraagd of de verkregen CT scan van toegevoegde waarde was voor het op te stellen behandelplan en/of de pre-operatieve planning. Op basis van deze data hebben we vastgesteld wat de implicaties zijn van het standaard aanvragen van een aanvullende CT scan bij gedisloceerde intra-articulaire distale radius fracturen danwel het besluit om een CT scan aan te vragen over te laten aan de chirurg. Dit liet zien dat wanneer de chirurg zelf de keuze mag maken al dan niet een CT scan aan te vragen om het behandelplan op te stellen worden er relatief minder onnodige CT scans aangevraagd en neemt het onnodig risico voor de patiënt af. Om die reden adviseren we om niet standaard een aanvullende CT scan aan te vragen bij patiënten met een gedisloceerde distale radius fractuur. Echter indien de chirurg besluit om een patiënt operatief te behandelen op basis van de conventionele röntgenfoto's, is het veel moeilijker de toegevoegde waarde van een CT scan voor preoperatieve planning te voorspellen. We kunnen dus geen aanbeveling doen wat betreft het wel of niet standaard aanvragen van een aanvullende CT scan bij deze groep patiënten.

Dat een chirurg de toegevoegde waarde van een aanvullende CT scan (deels) kan voorspellen, betekent niet automatisch dat dit ook leidt tot een betere uitkomst voor de patiënt. Toekomstige gerandomiseerde studies zouden de volgende twee groepen moeten vergelijken:

1. Een CT scan aanvragen bij alle patiënten met een gedisloceerde intra-articulaire DRF
2. CT scan alleen op verzoek van de chirurg

Ten eerste, om de resultaten van deze studie te bevestigen. En ten tweede, om de patiënten te vervolgen, zodat vastgesteld kan worden welk beslissingsmodel de beste patiëntenuitkomst geeft. Toekomstige studies zouden ook een kosten-effectiviteits-analyse moeten uitvoeren.

In **Hoofdstuk 3 en 4** hebben we de volgende vraag beantwoord:

- Zijn de gevalideerde uitkomstmaten die gebruikt worden in studies naar distale radius fracturen daadwerkelijk valide?

Allereerst hebben we een systematische review gedaan naar de methodologische kwaliteit van klinimetrische studies waarin de meeteigenschappen van *patient-reported outcome measures* (PROMs) die gebruikt worden bij DRF patiënten geëvalueerd worden. We hebben zowel de methodologische kwaliteit als de resultaten van de verschillende klinimetrische studies geanalyseerd om de bewijsgraad van iedere meeteigenschap (o.a. *reliability*, responsiviteit, interne consistentie) van de PROMs vast te stellen. Onze resultaten laten zien dat de meest onderzochte meeteigenschappen die van de *Patient-Rated-Wrist-Evaluation* (PRWE) zijn, gevolgd door de *Disability of the Arm, Shoulder and Hand questionnaire* (DASH). Echter, sterk bewijs dat de goede kwaliteit van de beschikbare PROMs voor patiënten met DRFs ondersteunt, mist in de beschikbare literatuur. Een goede PROM zou gebaseerd moeten zijn op een hoge bewijsgraad (d.w.z. geëvalueerd in kwalitatief goede studies), die de goede kwaliteit van de meeteigenschappen ondersteunen.

In onze systematische review is weinig tot geen bewijs gevonden voor een van de meeteigenschappen van de PRWE of de DASH. Om deze reden hebben we een multi-center prospectieve klinimetrische studie uitgevoerd, waarin de interne consistentie, *test-retest reliability*, meetfouten en content validiteit, volgens de COSMIN standaard geëvalueerd zijn. Op basis van onze resultaten bestaat nu sterk bewijs dat zowel de PRWE als de DASH vragenlijsten inhoudelijke valide, betrouwbare en intern consistente instrumenten zijn voor de beoordeling van patiënten met distale radius fracturen.

Hoewel de conclusies over deze meeteigenschappen niet zijn veranderd, hebben we nu wel sterk bewijs beschikbaar om deze conclusies te ondersteunen. Echter, in eerdere studies werd beperkt bewijs gevonden over de accepteerbaarheid van de meetfout, met een kleinst detecteerbare verandering (*Smallest Detectable Change*, SDC) van 4.4 tot 11 punten voor de PRWE. In ons onderzoek vonden we een veel hogere SDC van 20 punten, waaruit blijkt dat de kwaliteit van de klinimetrische studie van grote invloed kan zijn op de conclusies. Daarnaast hebben we een SDC van 14 punten gevonden voor de DASH, op basis van kwalitatief sterk bewijs. De *Minimal Clinically Important Difference* (MCID) voor de PRWE, gevonden in eerdere onderzoeken, is kleiner dan de SDC, waardoor het moeilijk is relevante verschillen te onderscheiden van meetfouten, wanneer het meetinstrument gebruikt wordt voor individuele patiënten. Daarentegen is het meetinstrument nog steeds bruikbaar voor onderzoeksdoeleinden, omdat de meetfout afneemt wanneer dit betrekking heeft op grote groepen patiënten.

Toekomstige studies van hoge methodologische kwaliteit zouden gericht moeten zijn op de meeteigenschap responsiviteit, gedefinieerd als het vermogen om klinisch relevante veranderingen in de tijd te detecteren. De MCID is een belangrijk onderdeel van de responsiviteit, waarvoor speciale aandacht op zijn plaats is. De informatie over de MCID van de DASH en de PRWE die we momenteel beschikbaar hebben, is ofwel gebaseerd op bewijs van beperkte kwaliteit, ofwel alleen geëvalueerd in aandoeningen anders dan DRFs.

In **Hoofdstuk 5** hebben we de volgende vraag beantwoord:

- Wat is de invloed van spectrum bias op de intra-beoordelaars overeenkomst wat betreft het behandelplan van distale radius fracturen?

We hebben een *reliability* studie verricht om de invloed van spectrum bias op de intra-beoordelaars overeenkomst wat betreft behandelplannen voor DRFs te bepalen. Spectrum bias is de bias die optreedt wanneer een populatie, waarbij therapeutische onzekerheid ontbreekt, wordt geïncludeerd in een studie. In deze studie hebben we de invloed van de mate van therapeutische onzekerheid bij de chirurg op de mate van overeenkomst van de behandelplannen binnen de beoordelaars bij patiënten met gedisloceerde DRFs bepaald op basis van alleen conventionele radiografie of in combinatie met een CT scan. Wanneer spectrum bias buiten beschouwing wordt gelaten, verminderde de intra-beoordelaars betrouwbaarheid wanneer een aanvullende CT scan werd gebruikt. Wanneer in onze studie gecorrigeerd wordt voor spectrum bias, veranderden onze conclusies. In casus waarbij therapeutische onzekerheid bestaat, verbeterde de intra-beoordelaars betrouwbaarheid door het gebruik van een aanvullende CT scan. In de gevallen waarbij geen therapeutische onzekerheid bestond, hebben we aangetoond dat een toegevoegde CT scan de betrouwbaarheid juist afneemt. Dit onderzoek heeft aangetoond dat een CT scan intra-beoordeelaars betrouwbaarheid daadwerkelijk kan verbeteren, maar alleen wanneer therapeutische onzekerheid bestaat. Dit is een ander duidelijk voorbeeld waarbij conflicterende resultaten in de literatuur verklaard kunnen worden door methodologische beperkingen.

Om de huidige controverses over de toegevoegde waarde van CT scans voor overeenstemming in behandelplannen voor andere typen fracturen vast te stellen, stellen wij voor dat bij toekomstige studies de therapeutische onzekerheid in beschouwing wordt genomen om spectrum bias te verminderen. Toekomstige prospectieve gerandomiseerde trials zouden uitgevoerd moeten worden om te bepalen of het gebruik van een aanvullende CT scan en de hieruit volgende behandeling, in gevallen waar therapeutische onzekerheid bestaat, de uitkomsten van patiënten met gedisloceerde DRFs beïnvloedt.

In **Hoofdstuk 6** hebben we de volgende vraag beantwoord:

- Zijn *hoekstabiele volair platen* superieur aan percutane *K-draad fixatie* voor dorsaal gedisloceerde distale radius fracturen?

We hebben een systematische review en meta-analyse uitgevoerd naar percutane Kirschner draad fixatie met gipsimmobilisatie (PKW) in vergelijking tot *hoekstabiele volair platen* (VLP) bij gedisloceerde DRFs. In alle studies die zijn meegenomen in dit review, is conventionele radiografie gebruikt om de fracturen te classificeren op basis van de AO classificatie en de DASH scores zijn gebruikt als uitkomstmaat. Hieruit blijkt, met matige betrouwbaarheid, dat de VLP betere DASH scores oplevert zowel na 3 maanden als aan het einde van de follow-up periode (6-12 maanden) in vergelijking tot PKW bij volwassen patiënten met gedisloceerde distale radius fracturen. Echter, deze verschillen hebben de a priori bepaalde grens voor klinische relevantie niet bereikt.

Op basis van deze resultaten, is verder onderzoek nodig om vast te stellen of VLP bij specifieke röntgenologische (bv. Intra-articulaire fracturen) of patiëntenkarakteristieken (bv. ouderen) op korte termijn een betere uitkomst kunnen geven. De invloed van post-traumatische arthritis is moeilijk vast te stellen door de korte follow-up perioden in de geïncludeerde studies in deze meta-analyse. Verwacht wordt dat minder post-traumatische arthritis wordt gezien wanneer VLPs worden gebruikt. Lange termijn studies zijn nodig om de invloed van post-traumatische arthritis op de patiëntenuitkomsten effectief te onderzoeken.

Op het eerste gezicht (op basis van de Cochrane risk of bias samenvatting), lijken de resultaten van deze studies van hoge kwaliteit en chirurgen gebruiken de bevindingen waarschijnlijk in de klinische praktijk. Echter, wij stellen dat voorzichtigheid geboden is wanneer nader gekeken wordt naar de methodologische beperkingen van gepubliceerde studies, met name gelet op de ondergewaardeerde en niet erkende methodologische beperkingen die naar voren komen in dit proefschrift. Zoals aangetoond in dit proefschrift, wordt de indeling van patiënten naar type fractuur beïnvloedt door het gebruik van aanvullende CT scans in vergelijking tot conventionele radiografie alleen. Kritisch kijkend naar de geïncludeerde studies kunnen we zien dat het waarschijnlijk is dat de resultaten onnauwkeurig of beïnvloed worden door bias, omdat alleen conventionele rontgenfoto's gebruikt werd om fracturen te classificeren om te bepalen of een patiënt in aanmerking kwam voor deelname aan de studie. Daardoor verschillen de patiëntenpopulaties in de studies waarin patiënten die geselecteerd zijn op basis van alleen conventionele rontgenfoto's van de populaties waarbij aanvullende CT scans zijn gebruikt om patiënten te selecteren. Dit heeft gevolgen voor de generaliseerbaarheid en toepasbaarheid van de resultaten in de klinische praktijk.

De geïncludeerde studies zijn uitgevoerd voordat de DASH gevalideerd is voor gebruik in DRF patiënten, waardoor het risico bestond op onnauwkeurigheid of bias en daardoor potentieel misleidende conclusies. In dit proefschrift hebben we daarentegen aangetoond dat de DASH valide en betrouwbaar is in deze patiëntenpopulatie, waardoor ons vertrouwen in de conclusies van dit review versterkt wordt.

TOEKOMSTPERSPECTIEF

Vele klinische vragen over de behandeling van DRFs blijven onopgelost en onbeantwoord in dit proefschrift. Om deze vragen te beantwoorden zijn echter studies gebaseerd op methodologie van hoge kwaliteit nodig. In dit proefschrift hebben we verschillende ondergewaardeerde en niet erkende methodologische kwesties in wetenschappelijk onderzoek naar DRFs geïdentificeerd en bediscussieerd. We adviseren onderzoekers alert te zijn op deze beperkingen van eerder onderzoek naar DRF ten einde de kwaliteit van DRF literatuur te verbeteren. In ons onderzoeksprogramma nemen we onze eigen adviezen mee in het opzetten van vervolgstudies

In ons eerstvolgende vervolgonderzoek hebben we een gerandomiseerde gecontroleerde trial opgezet waarin we de beperkingen van de literatuur, die in dit proefschrift aan het licht worden gebracht, beogen te vermijden. Op basis van het in dit proefschrift gepresenteerde bewijs, zal het protocol voor de aankomende *Effectiveness of Surgery Versus Casting for Elderly Patients with Displaced Intra-articular Radius Fractures Trial* (DART Trial) de volgende kenmerken omvatten: 1) inclusiecriteria op basis van fractuurtype in plaats van een classificatiesysteem; 2) de keuze voor het gebruik van een aanvullende CT scan wordt overgelaten aan de chirurg; en 3) de gevalideerde DASH en PRWE worden gebruikt als uitkomstmaten. In onze zoektocht naar bewijs van hoge kwaliteit voor de behandeling van DRFs, streven wij er naar twee stappen voorwaarts, zonder een stap terug, te nemen.

CONFLICTS OF INTEREST STATEMENT

The author of this thesis declares that no conflicts of interest are applicable.

PHD PORTFOLIO

Name PhD student: Ydo V. Kleinlugtenbelt
PhD period: 2015-2017
Name PhD supervisor: Prof. Dr. Goslings/Prof Dr. Bhandari

1. PhD training

	Year	Workload (Hours/ECTS)
General courses		
Principles and Practice of Clinical Research; London course directors Dr. Mohit Bhandari and Dr. Emil Schemitsch.	2010	1
SRM (surgical research methodology)McMaster University	2014-2015	2
GCP (Good Clinical Practice)	2014	0.5
Specific courses		
Experimenteel Onderzoek Heelkundige Specialismen, SEOHS	2009	0.5
Seminars, workshops and master classes		
- Journalclubs orthopaedic surgery weekly OLVG/UMCU	2010-2013	4
- Grand rounds orthopaedic surgery weekly McMaster	2014-2015	2.5
- Journal club orthopaedic surgery McMaster	2014-2015	1
- Workshop scientific presentation, M Bhandari	2014	0.2
- Workshop “getting things done”, M Bhandari	2014	0.2
- Building bridges, Clinical research focus showcase (part 1 and 2)	2014	0.4

Presentations

ERASS (European Rheumatoid Arthritis Surgical Society)	2008	0.5
Kleinlugtenbelt YV, Bakx PA, Huij J. Instrumented Bone Preserving	2008	0.5
elbow prosthesis in rheumatoid arthritis: 2-8 year follow-up	2008	0.5
NOV (Dutch Orthopedic Society) annual meeting	2013	0.5
Kleinlugtenbelt YV, Bakx PA, Huij J. Instrumented Bone Preserving elbow prosthesis in rheumatoid arthritis: 2-8 year follow-up		
ESSSE (European Society for Surgery of the Shoulder and the Elbow)	2013	0.5
Kleinlugtenbelt YV, Bakx PA, Huij J. Instrumented Bone Preserving elbow prosthesis in rheumatoid arthritis: 2-8 year follow-up		
NOV (Dutch Orthopedic Society) annual meeting	2014	0.5
Kleinlugtenbelt YV, Groen SR, Ham SJ, Kloen P, Haverlag R, Simons MP, Poolman RW, Scholtes VAB Classification systems on distal radius fractures.Does the reliability improve using an additional CT scan?		
EFORT (European Federation of National Associations of Orthopaedics and Traumatology)	2014	0.5
Kleinlugtenbelt YV, Nienhuis R, Poolman RW, Scholtes VAB Patient reported outcome in distale radius fracture research: Wich instrument should we use?A systematic review on measurement properties using COSMIN		
Invited speaker NOV (Dutch Orthopedic Society) annual meeting	2014	0.5
Patient reported outcome in hand and wrist disorders		

(Inter)national conferences

- ERASS, Barcelona	2008	0.5
- ESSSE, Brugge	2008	0.7
- NOV (najaars/voorjaars/jaarvergadering)	2008-2013	4
- COA/AOA, Montreal	2014	0.7

Other

-

2. Teaching

	Year	Workload (Hours/ECTS)
Lecturing		
- Orthopaedic Resident Review in Trauma course	2014	0.5
- Grand rounds McMaster University	2014	0.5
Tutoring, Mentoring		
-		
Supervising		
- Muzammil Menon, Medical student McMaster University; Project: Use of Google in fracture care	2014-2015	1
- Rosalie Ackerman: Tibial fixation in ACL reconstruction	2017	1
- Niek van der Hoek: Use of Google in fracture care	2017	1
Other		
- Resident teaching, McMaster University	2014-2015	2

LIST OF PUBLICATIONS

Peer-reviewed publications

1) **Kleinlugtenbelt YV**, Madden K, Groen SR, Ham SJ, Kloen P, Haverlag R, Simons MP, Bhandari M, Goslings JC, Scholtes VAB, Poolman RW. Can experienced surgeons predict the additional value of a CT scan in patients with displaced intra-articular distal radius fractures? Strategies Trauma Limb Reconstr. 2017 Apr 24. [Epub ahead of print]

2) Madden K, **Kleinlugtenbelt YV**. Cochrane in CORR ®: Decision Aids for People Facing Health Treatment or Screening Decisions. Clin Orthop Relat Res. 2017 May;475(5):1298-1304.

3) **Kleinlugtenbelt YV**, Scholtes VA, Toor J, Amaechi C, Maas M, Bhandari M, Poolman RW, Kloen P. Does Computed Tomography Change our Observation and Management of Fracture Non-Unions? Arch Bone Jt Surg. 2016 Oct;4(4):337-342.

4) Memon M, Ginsberg L, Simunovic N, Ristevski B, Bhandari M, **Kleinlugtenbelt YV**. Quality of Web-based Information for the 10 Most Common Fractures. Interact J Med Res. 2016 Jun 17;5(2):e19..

5) **Kleinlugtenbelt YV**, Nienhuis RW, Bhandari M, Goslings JC, Poolman RW, Scholtes VA. Are validated outcome measures used in distal radial fractures truly valid? A critical assessment using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. Bone Joint Res. 2016 Apr;5(4):153-61.

6) Burgers PT, Hoogendoorn M, Van Woensel EA, Poolman RW, Bhandari M, Patka P, Van Lieshout EM; **HEALTH Trial Investigators**. Total medical costs of treating femoral neck fracture patients with hemi- or total hip arthroplasty: a cost analysis of a multicenter prospective study. Osteoporos Int. 2016 Jun;27(6):1999-2008.

7) Devji T, **Kleinlugtenbelt Y**, Evaniew N, Ristevski B, Khoudigian S, Bhandari M. Operative versus nonoperative interventions for common fractures of the clavicle: a meta-analysis of randomized controlled trials. CMAJ Open. 2015 Nov 10;3(4):E396-405.

8) **Kleinlugtenbelt YV**, Hoekstra M, Ham SJ, Kloen P, Haverlag R, Simons MP, Bhandari M, Goslings JC, Poolman RW, Scholtes VA. Spectrum bias, a common unrecognised issue in orthopaedic agreement studies: do CT scans really influence the agreement on treatment plans in fractures of the distal radius? Bone Joint Res. 2015 Dec;4(12):190-4.

9) Evaniew N, McCarthy C, **Kleinlugtenbelt YV**, Ghert M, Bhandari M. Vitamin C to Prevent Complex Regional Pain Syndrome in Patients With Distal Radius Fractures: A Meta-Analysis of Randomized Controlled Trials. J Orthop Trauma. 2015 Aug;29(8):e235-41.

10) **Kleinlugtenbelt YV**, Bhandari M. Cochrane in CORR (®): Interventions for Treating Proximal Humeral Fractures in Adults (Review). Clin Orthop Relat Res. 2015 Sep;473(9):2750-6.

11) Chaudhry H, **Kleinlugtenbelt YV**, Mundi R, Ristevski B, Goslings JC, Bhandari M. Are Volar Locking Plates Superior to Percutaneous K-wires for Distal Radius Fractures? A Meta-analysis. Clin Orthop Relat Res. 2015 Sep;473(9):3017-27.

12) Bruinsma WE, Guitton T, Ring D; **Science of Variation Group**. Radiographic loss of contact between radial head fracture fragments is moderately reliable. Clin Orthop Relat Res. 2014 Jul;472(7):2113-9.

13) Hageman MG, Guitton TG, Ring D; **Science of Variation Group**. How surgeons make decisions when the evidence is inconclusive. J Hand Surg Am. 2013 Jun;38(6):1202-8.

14) van Lieshout AP, van Manen CJ, du Pré KJ, **Kleinlugtenbelt YV**, Poolman RW, Goslings JC, Kloen P. Peak incidence of distal radius fractures due to ice skating on natural ice in The Netherlands. Strategies Trauma Limb Reconstr. 2010 Aug;5(2):65-9.

15) **Kleinlugtenbelt IV**, Bakx PA, Huij J. Instrumented Bone Preserving elbow prosthesis in rheumatoid arthritis: 2-8 year follow-up. J Shoulder Elbow Surg. 2010 Sep;19(6):923-8.

Book (chapter)

1) **Kleinlugtenbelt YV**, Bhandari M; Bone stimulation for fracture healing. Extracorporeal Shock Waves, Electric Stimulation, Ultrasound. Bookchapter *Turek's Orthopaedics: Principles and Their Application 7th Edition*.

2) **Kleinlugtenbelt YV**; The Reliability Study: Measuring Surgeon Agreement.

3) Bhandari M, Sancheti P. Clinical Research made easy. A guide to publishing in medical literature. 2nd edition, 2017. Co-editor: **Y.V. Kleinlugtenbelt**

DANKWOORD

In de winter van 2011 en 2012 daalde het kwik langdurig onder nul. Erik Hulzebosch maakte overuren in de Nederlandse talkshows en de dikte van het natuurijs was onderwerp van gesprek bij iedere koffieautomaat. De Elfstedentocht kon uiteindelijk in beide winters geen doorgang vinden, maar heel Nederland kon de schaatsen eindelijk weer eens uit het vet halen. 10 dagen lang was het een groot feest op de Nederlandse wateren.

Echter, deze mooie Nederlandse traditie had ook zijn keerzijde. Er stonden rijen met patiënten voor de ingang van de Spoedeisende Hulp. Vaak hun arm ondersteunend in zelfgemaakte mitella´s en soms nog met hun schaatsen om hun nek. De SEH draaide overuren en in de daaropvolgende weken was de traumapoli volledig overboekt. Voor al die patiënten uiteraard een drama, maar voor mij als assistent in opleiding tot orthopedisch chirurg met voorliefde voor de traumatologie kwamen er onverwachts vele mogelijkheden op mijn weg. Dit moment was de officieuze start van mijn promotie.

De **maatschap orthopedie** in het **OLVG** liet mij samen met de gipsverbandmeesters mijn eigen polsfracturen poli opzetten. Ik kreeg de volledige vrijheid en vertrouwen voor dit project. Dit is mijn ogen typerend voor deze maatschap in het OLVG en dit is slechts een klein voorbeeld van alle mogelijkheden die ze tijdens mijn opleiding en zelfs daarna geboden hebben. Ik wil jullie hiervoor dan ook ontzettend bedanken en zal hier mijn gehele carrière de vruchten van plukken.

In het bijzonder wil ik mijn beide opleiders in het OLVG bedanken. **Jaap Willems**, jij hebt mij de kans gegeven om vanuit Rotterdam naar Amsterdam te komen om mijn opleiding tot orthopedisch chirurg in het OLVG te volgen. Ik zal de woensdagmiddagen waar je de voortgang van wetenschappelijke projecten met jou moest bespreken nooit vergeten. Jij hebt mij geholpen om als volledig onervaren wetenschapper mijn eerste artikel in de Journal of Shoulder and Elbow te publiceren.

Daarna begon het tijdperk van **Rudolf Poolman** als mijn opleider. Voor ieder klinisch probleem was er wel een artikel waarin je volgens Rudolf het antwoord kon vinden. Evidence Based Medicine voerde de boventoon. Ik werd iedere dag uitgedaagd en met halve waarheden kwam je niet weg. Rudolf, jij haalde het beste in mij naar boven!

Samen met jou en research coördinator **Vanessa Scholtes** werden mijn ideeën omgezet in een promotie traject. Rudolf de man van de grote ideeën, Vanessa en ik zorgden er dan samen voor dat het uitvoerbaar werd. Vanessa ik wil je bedanken voor alle energie die je vanaf het begin tot te eind in mijn promotie hebt gestopt. Met name in het begin was je kennis van de epidemiologie onmisbaar. Ik heb op dit gebied zoveel van je geleerd en zonder jouw kennis en energie was dit proefschrift nooit tot stand gekomen.

Sylvester Groen, **Rik Nienhuis** en **Reinoud Krol**, jullie heb ik tijdens jullie wetenschappelijke stage mogen begeleiden. Ik heb genoten van jullie enthousiasme en doorzettingsvermogen. Jullie dataverzameling is goud waard geweest voor dit promotietraject. Ik wil jullie hier uiteraard ontzettend voor bedanken.

De basis van mijn traumatologische vaardigheden is gelegd door **Maarten Simons** en **Robert Haverlag**. Jullie gaven mij veel vertrouwen in mijn vooropleiding en via deze weg wil ik jullie ook nog namens Djamilla Boerma bedanken, die tijdens de tweede kerstdag lekker thuis kon blijven, en mij als tweedejaars assistent een volwaardig traumaprogramma liet afwerken. Maar naast deze klinische ervaringen moet ik jullie natuurlijk ook bedanken voor jullie bijdrage, wat uiteindelijk tot drie publicaties heeft geleid, aan dit proefschrift.

Mijn speciale dank gaat uit naar **John Ham**. Vele polsfracturen hebben we samen geopereerd en je was altijd nauw betrokken bij mijn wetenschappelijke vorderingen. Als personen zijn we totaal verschillend, maar juist dat maakte onze samenwerking zo goed en bijzonder. Je kritische beoordelingen van de vele polsfracturen en de revisies van de vele artikelen waar je aan meegeholpen hebt, heeft ervoor gezorgd dat de kwaliteit van dit proefschrift sterk verbeterd is.

Peter Kloen, het heeft even geduurd, maar uiteindelijk heeft onze samenwerking geleid tot drie mooie publicaties. Je was altijd betrokken en kwam iedere afspraak na. Ik ben je hier dankbaar voor. Sylvester zal de beoordelingssessies op zaterdagochtend niet snel vergeten.

My application for a clinical and scientific fellowship at McMaster University consisted of one text message from Rudolf to **Mohit Bhandari**. Within a minute we received a positive answer: “Of course he is welcome”. I would like to thank Mo for all the opportunities he offered me. At what other fellowship do you have the opportunity to have private dinners with Wayne Paprosky, Peter Giannoudis and Thomas Einhorn? You gave me complete freedom to plan my scientific and clinical activities. And I will never forget the following words during our first meeting: “If you are not going to travel in Canada with your family for a couple of weeks, you will be fired.” Mo, thanks for every single minute of our time in Canada and being one of my promotors.

It is not only Mo Bhandari which made this fellowship so special; his team at the Centre for Evidence Based Orthopaedics is one of the best in the world. **Sheila Sprague**, assistant professor, how would our start in Hamilton have been without you? Everything was arranged and we could always call you for help. And even after my fellowship you kept supporting me and also visited me at the Deventer Hospital, so we could continue our collaboration in science. I am extremely grateful to you. **Nicole Simunovic**, research coordinator, it never took you more than 24 hours to correct any of my articles. The door of your office was always open for me and I enjoyed the conversations about complex methodological issues. Thank you for all your help.

Kim Madden, research coordinator, without you I would have never finished this thesis. Shortly after my fellowship you came to the Netherlands to work at the Deventer Hospital for three months. We had multiple sessions behind our laptops, live or through skype. Most of the time I was tired after a day of hard working, but with your energy you kept me awake. For you the sky is the limit, although you made it through the first rounds, unfortunately you didn't make it to become an astronaut.

Ted Berenschot, medewerker bibliotheek OLVG, jij valt in de groep stille krachten. Je hebt vele complexe zoekopdrachten in Pubmed voor me uitgevoerd. Het maakte je niet uit of ik in het OLVG, het UMCU of aan McMaster University werkzaam was. Dank voor al je werkzaamheden die onmisbaar waren voor dit proefschrift.

Als orthopedisch chirurg promoveren bij een traumachirurg zal bij sommigen de wenkbrauwen doen fronsen. Echter voor mij was het een logische keuze. De samenwerking tussen de traumachirurgen en orthopedisch traumachirurgen is in mijn ogen belangrijk om traumachirurgie in Nederland te optimaliseren. Mijn bezoeken aan het AMC zorgden er altijd weer voor dat ik opnieuw energie kreeg om de volgende stap te nemen. **Carel Goslings**, ik was altijd onder de indruk van je commentaar op mijn manuscripten. Zonder jouw steun en betrokkenheid was dit proefschrift niet tot stand gekomen. **Jacqueline Brockhoff**, secretaresse professor Goslings, jij zorgde er voor dat er geen deadline gemist werd. Stress komt in jouw woordenboek niet voor. Iedere promovendus verdient zo'n secretaresse!

Een goed voorbeeld van een uitstekende samenwerking in de traumachirurgie is onze trauma-unit in het Deventer Ziekenhuis. **Herbert Roerdink en Elvira Flikweert**, traumachirurgen DZ, jullie enthousiasme is aanstekelijk. Jullie steunen me in alle onderzoeken die we in een korte tijd hebben opgezet in het Deventer Ziekenhuis. De kans dat er een patiënt in jullie dienst niet geïncludeerd wordt, is praktisch nul. De titel “best including center of the month” in de HEALTH trial kwam vaak grotendeels op jullie conto. Tijdens onze nieuwe distale radius trial (DART) wilden jullie eigenlijk al includeren voordat we eigenlijk officieel begonnen waren. Herbert, ik kijk uit naar onze verdere samenwerking in de toekomst en hoop oprecht voor je dat je met hoog kwalitatief onderzoek kan aantonen dat de Gannet het eerste keus implantaat wordt bij mediale collum fracturen. Elvira, zoals ik in een van mijn laatste stellingen aangeef, kost het laatste deel van je proefschrift heel veel energie. Maar ik ben er van overtuigd dat op korte termijn ook jouw proefschrift af is en we een trauma unit hebben met enkel gepromoveerde (orthopedisch) traumachirurgen. En ik zeg het nog een keer: “als je hulp nodig hebt dan ben je altijd welkom!”

Sinds 2015 maak ik deel uit van een fantastische vakgroep orthopedie in het Deventer Ziekenhuis. Zonder steun van je maten is goed wetenschappelijk onderzoek doen haast niet mogelijk. **Danielle Langeloo**, hoewel je je misschien soms afvroeg waar ik allemaal mee bezig was, heb je me vanaf dag 1 gesteund in al mijn wetenschappelijke plannen. Samen met **Hans-Peter van Jonbergen** hebben we al snel de stichting ”Deventer Evidence Based Orthopaedics” opgericht. Een stichting waar we grote plannen mee hebben! Hans-Peter, ik geniet van de dialogen die we voor en na ons werk hebben, dit houdt me scherp. Wetenschappelijk onderzoek maakt echt deel uit van jouw praktijk. Ik kijk ontzettend uit naar ons gezamenlijke project dat we hopelijk binnenkort samen met Zimmer kunnen beginnen. Grote data met mogelijk baanbrekende uitkomsten die de orthopedie echt kunnen veranderen. De aankomende jaren zijn we wetenschappelijk in ieder geval nog aan elkaar verbonden. **Rinco Koorevaar**, terwijl ik gestaag mijn promotie doorzette, heb jij een enorme eindsprint ingezet. Indrukwekkend om te zien hoe snel je het allemaal afgerond hebt. Hopelijk kunnen we ons nu richten op het verder uitbreiden van onze schouderpoli en naast de regio functie die we hier in hebben er ook een officieel STZ speerpunt van maken. Ik weet zeker dat hier ook mooie wetenschappelijke projecten uit voort gaan komen. **Lex Barnaart**, twee promoties in je afscheidsjaar, dat is maar weinig orthopeden gegeven. Ik voelde me vereerd dat ik de laatste jaren van je carrière nog mee heb kunnen maken. Er zijn denk ik maar weinig orthopeden die nog zoveel passie en energie in hun vak stoppen een paar maanden voor hun afscheid. Lex, daar mag je trots op zijn! **Joost Reuver**, jij hebt nooit onder stoelen of banken geschoven dat je geen wetenschappelijke ambities hebt. Desondanks ben je onmisbaar in de onze wetenschappelijke projecten. Ook in jouw dienst kwam je in huis om HEALTH trial patiënten te opereren en voor PRONO-

MOS zal jij één van de twee includerende orthopeden worden. Zonder inclusies geen data, zonder data geen wetenschappelijk bewijs.

Esther van 't Riet, research coördinator wetenschapsbureau DZ, jij hebt er voor gezorgd dat ik één van mijn dromen heb kunnen laten uitkomen. Een eigen onderzoeksstichting voor de afdeling Orthopedie in het Deventer ziekenhuis met een gepromoveerde researchmedewerker gesteund door het ziekenhuis. Jij hebt de Raad van Bestuur kunnen overtuigen van onze ideeën, wat er toe geleid heeft dat de onderzoekstrein echt is gaan rijden.

Wat was ik trots dat ik een gepromoveerde research medewerker had gevonden met zelfs internationale werkervaring die samen met mij al mijn wetenschappelijke plannen wilde uitwerken. **Ellie Landman**, ondanks dat jij het soms nog te langzaam vindt gaan, mogen we best wel trots zijn op wat we in het afgelopen jaar hebben neergezet. Maar ook voor dit proefschrift heb je je steentje bijgedragen. Met name tijdens die laatste zware loodjes was je van onschatbare waarde. Ik kijk uit naar onze samenwerking de komende jaren!

Peter de Klein en **Jules de Beer** jullie hebben altijd met veel interesse mijn vorderingen gevolgd. Ik ben er trots op dat deze twee economen mijn paranimfen willen zijn.

Bij mijn schoonouders **Mechteld** en **Michiel Feilzer** heb ik heel wat uren achter mijn laptop gezeten, zowel op kantoor als in Portugal. Jullie vroegen je iedere keer af wanneer het nu eens af is en wat ik nog moest doen. Nu is het dan eindelijk klaar. Dank voor jullie geduld.

Papa en Mama, "ikke zelf doen" is iets wat ik van kleins af aan al zei. Jullie hebben me altijd gesteund om mezelf te ontplooien en mijn eigen weg te belopen. Mama, we zijn totaal anders maar toch lijken we ook veel op elkaar. En papa ik ben opgegroeid met weinig complimenten, maar dat heeft er bij mij altijd toe geleid dat ik weer dat extra stapje deed. Papa en mama ik ben jullie heel dankbaar.

Rosalie, lieverd, wat hebben we de laatste jaren samen veel meegemaakt. We hebben de twee zwaarste winters in jaren doorgemaakt in Canada. Ik aan het werk en jij met de kinderen thuis bij -30 graden Celsius terwijl je geen kant op kon. Zonder jouw steun en geduld had ik het nooit gered. We zijn meerdere malen voor mijn werk verhuisd. De dag na mijn verdediging voorlopig voor het laatst naar onze mooie huis in Gorssel. We gaan samen met de jongens daar een hele mooie tijd tegemoet!

Bram en **Gijs** jullie hebben er hopelijk niet al te veel van meegekregen. Maar op de laatste dag tijdens het schrijven van dit dankwoord kwamen jullie toch echt vragen waarom ik de hele dag achter de computer moest zitten. Gelukkig net op tijd klaar. De kans dat jullie in mijn voetsporen treden is op dit moment heel erg klein (en ook helemaal niet erg). Jullie worden geen dokter want dat vinden jullie veel te gemakkelijk: “Als dokter hoef je maar één ding te kunnen: Mensen beter maken. Als politieagent moet je boeven vangen, bekeuringen uitschrijven, het verkeer regelen… dat is pas moeilijk!!”

CURRICULUM VITAE

Ydo (Vincent) Kleinlugtenbelt is geboren op 31 maart 1978 in Aalburg. Zijn jeugd heeft hij doorgebracht in Drunen, waar hij op het d'Oultremontcollege zijn VWO afrondde in 1996. De keuze om geneeskunde te gaan studeren stond al op jonge leeftijd vast. Echter door het numerus fixus systeem heeft hij de eerste 2 jaar na de middelbare school in het buitenland doorgebracht. In Brighton, Engeland heeft hij zijn Cambridge Proficiency examen gehaald. Vervolgens heeft hij zijn eerste kandidatuur behaald aan het Rijks Universitair Centrum Antwerpen. In 1998 werd hij toegelaten aan de faculteit geneeskunde van de Erasmus Universiteit te Rotterdam. In 2002 behaalde hij zijn doctorandus titel, waarna hij zich 2 jaar fulltime op zijn roei carrière richtte. In 2006 behaalde hij, cum laude, zijn artsexamen. Na een jaar als Agnio chirurgie gewerkt te hebben in het toenmalige MCRZ, werd hij in januari 2008 toegelaten tot de opleiding tot orthopedisch chirurg aan de Universiteit van Utrecht (opleiders Prof. Dr. R.M. Castelein en Prof. Dr. D.B.F. Saris). Zijn chirurgische vooropleiding (opleider Dr. M.F. Gerhards) en het grootste deel van zijn orthopedische opleiding (opleiders Dr. W.J. Willems en Dr. R.W. Poolman) heeft hij doorlopen in het Onze Lieve Vrouwe Gasthuis in Amsterdam. In 2014-2015 heeft hij gewerkt als "Clinical and Research fellow in upper extremity and trauma" aan Mcmaster University, Hamilton,Canada onder professor M. Bhandari. Per 1 juni 2015 is hij begonnen als orthopedisch-traumachirurg in het Deventer Ziekenhuis en heeft aldaar de stichting Deventer Evidence-Based Orthopaedics opgericht. Hij heeft samen met Rosalie Feilzer twee zonen, Bram en Gijs.