



■ TRAUMA

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?

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

From Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands

■ Y. V. Kleinlugtenbelt, MD, Orthopaedic Surgeon, Department of Orthopaedic and Trauma Surgery Deventer Ziekenhuis, Nico Bolkesteinaan 75, 7416 SE Deventer, The Netherlands.
 ■ M. Hoekstra, MD, House Officer, Department of Orthopaedic and Trauma Surgery
 ■ S. J. Ham, MD, PhD, Orthopaedic Surgeon, Department of Orthopaedic and Trauma Surgery
 ■ R. Haverlag, MD, Trauma Surgeon, Department of General and Trauma Surgery
 ■ M. P. Simons, MD, PhD, Trauma Surgeon, Department of General and Trauma Surgery
 ■ R. W. Poolman, MD, PhD, Orthopaedic Surgeon, Department of Orthopaedic and Trauma Surgery
 ■ V. A. B. Scholtes, PhD, Program Manager
 JointResearch Onze Lieve Vrouwe Gasthuis, Oosterpark 9, P.O. Box 95500, 1090 HM, Amsterdam, The Netherlands.
 ■ P. Kloen, MD, PhD, Orthopaedic Surgeon, Department of Orthopaedic and Trauma Surgery Academic Medical Centre, P.O. Box 22660, 1100 DD, Amsterdam, The Netherlands.
 ■ M. Bhandari, MD, PhD, Orthopaedic Surgeon, Division of Orthopaedic Surgery McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Ontario, Canada.
 ■ J. C. Goslings, MD, PhD, Trauma Surgeon, Trauma Unit, Department of Surgery Academic Medical Center, P.O. Box 22660, 1100 DD, Amsterdam, The Netherlands.
 Correspondence should be sent to Dr Y. V. Kleinlugtenbelt; e-mail: ijdokib@hotmail.com

doi:10.1302/2046-3758.412.2000433
 \$2.00

Bone Joint Res
 2015;4:190–194.
 Received 8 May 2015; Accepted after revision 3 November 2015

Objectives

Current studies on the additional benefit of using computed tomography (CT) in order to evaluate the surgeons' agreement on treatment plans for fracture 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 intra-observer agreement of treatment plans for fractures of the distal radius.

Methods

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

Results

In cases with high therapeutic certainty, intra-observer agreement based on radiograph was almost perfect (0.86 to 0.90), but decreased to moderate based on a radiograph and CT (0.47 to 0.60). In cases with high therapeutic uncertainty, intra-observer agreement was slight at best (-0.12 to 0.19), but increased to moderate based on the radiograph and CT (0.56 to 0.57).

Conclusion

Spectrum bias influenced the outcome of this agreement study on treatment plans. An additional CT scan improves the intra-observer agreement on treatment plans for a fracture of the distal radius only when there is therapeutic uncertainty. Reporting and analysing intra-observer agreement based on the surgeon's level of certainty is an appropriate method to minimise spectrum bias.

Cite this article: *Bone Joint Res* 2015;4:190–194.

Keywords: Agreement study, spectrum bias, intraobserver, distal radial fracture, computed tomography

Article focus

- To determine the influence of spectrum bias on the intra-observer agreement of treatment plans for a fracture of the distal radius.

- Reporting and analysing intra-observer agreement based on the surgeon's level of therapeutic certainty is an appropriate method to minimise spectrum bias.

Key messages

- Spectrum bias influences the intra-observer agreement of treatment plans for a fracture of the distal radius
 - An additional CT scan improves the intra-observer agreement on treatment plans for a fracture of the distal radius only when there is therapeutic uncertainty.

Strengths and limitations

- Strength: the current study is the first agreement study to implement the surgeon's level of therapeutic certainty in their analysis to minimise the effect of spectrum bias.
 - Strength: The COAST criteria were used to ensure we addressed all components of an agreement study.

- Limitation: The distribution over the different groups of therapeutic certainty was skewed in this study.

Introduction

Various treatment methods are available for fractures of the distal radius, mostly guided by fracture characteristics and surgeons' expertise.¹ Historically, plain radiographs have played a large role in characterising these different type of fractures. However, it is known that plain radiographs are not the most reliable modality for accurate assessment of the distal part of the radius. The use of computed tomography (CT) is becoming a popular additional imaging modality to assess the exact morphology of fractures of the distal radius, especially when a surgeon cannot evaluate this from radiographs alone.²⁻⁴

The increased popularity of using both radiograph and CT may be supported by previous study results which show that, compared with plain radiographs alone, the addition of CT is a more accurate method of assessing certain fracture characteristics (e.g. the amount of comminution, involvement of the distal radioulnar joint and the extent of articular surface depression).^{2,3}

Therefore, the addition of CT improves the accuracy of assessing fracture characteristics of the distal radius. What is less clear is whether CT improves the agreement on treatment planning. Studies have found that the treatment plan (conservative or surgical) may shift after the addition of CT. More specifically, when a treatment plan is based on both radiograph and CT, a surgeon is more likely to treat the patient with a fracture of the distal radius surgically than when the treatment plan is based on radiograph 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 does not consistently improve with the addition of CT, when compared with radiographs alone.⁵⁻⁷

One explanation for these apparently inconsistent results may be the differences in the chosen study population in these agreement studies. Ideally, the test results should be evaluated in a study population that is a perfect reflection 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 significantly mislead clinicians.⁸

For example, when only cases with grossly dislocated extra-articular fractures, with inadequate positions after closed reduction, are selected for these studies, the intra-observer agreement will probably be very high, either based on radiographs or on radiographs and CT. This is because the therapeutic uncertainty will be low: surgeons will most likely plan to operate, based on a radiograph alone, and one would not expect them to change their treatment plan when they reassess a case with the addition of CT. Therefore, this group of patients would not be an appropriate study population. If chosen, it would give 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 compared with using radiographs alone.

On the other hand, when only cases are selected in which the radiograph 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 radiographs, 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 on treatment planning with an 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 cases in which they lack therapeutic certainty.

Therefore, the aim of this study was 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. To address the potential influence of spectrum bias, we will determine the influence of the surgeon's level of therapeutic certainty on the intra-observer agreement in treatment plan in patients with displaced fractures of the distal radius using radiograph alone or radiograph and CT. We hypothesised that 1) the intra-observer agreement is positively related to the surgeon's therapeutic certainty, both on radiograph and radiograph plus CT, 2) the level of certainty is most strongly related to the intra-observer agreement based on radiograph, and 3) the intra-observer agreement only improves by the addition of CT in therapeutically uncertain cases.

Materials 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 1, 2007 and March 2, 2011, a database was established of patients with a displaced fracture of the distal radius 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 presented with a displaced fracture of the distal radius in the Emergency Department, were 18 years of age or older, had no prior fracture or pathology of the distal radius, had both pre- and post-reduction plain posterior-anterior and lateral radiographs of the wrist, and had an additional post-reduction CT, made within five days after the reduction in case of any doubt of the characteristics of the fracture, or when there was a possible indication for surgery.

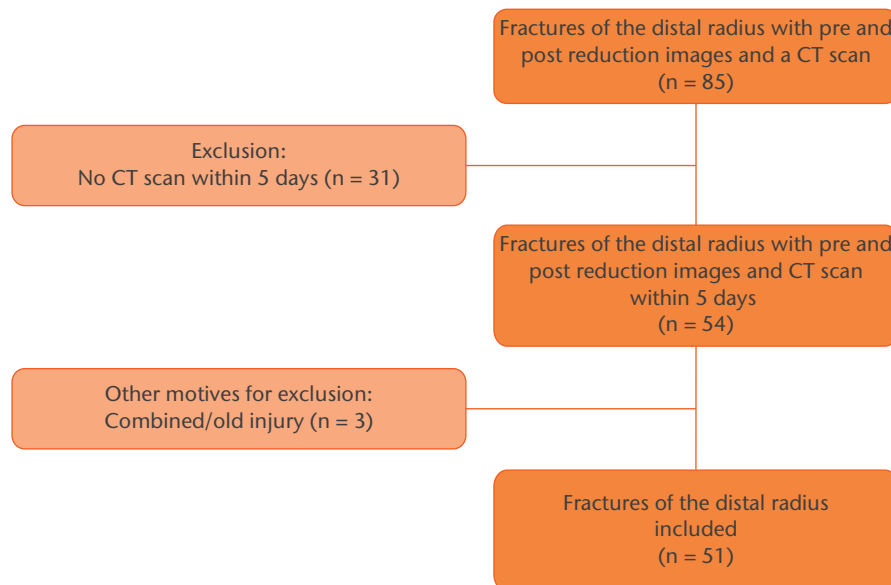


Fig. 1

Flowchart showing the exclusion criteria.

Observers. The panel consisted of four experienced Dutch surgeons, two of whom are trauma surgeons (MPS, RH) and two of whom are orthopaedic surgeons (SJH, PK). All four surgeons have over ten years of experience in the treatment of fractures and all are responsible for the care of patients with a fracture of the distal radius within their department.

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

- T1: pre- and post-reduction plain radiographs (T1 radiograph).

- T2: pre- and post-reduction plain radiographs (T2 radiograph).

- T3: pre- and post-reduction plain radiographs and axial, sagittal and coronal planes CT (T3 radiograph and CT).

- T4: pre- and post-reduction plain radiographs and axial, sagittal and coronal planes CT (T4 radiograph and CT).

All images were converted to digital format and anonymised. The cases were also presented with the relevant clinical data (e.g., age of the patient, gender, dominant hand, profession and specific hobbies).

Scoring form. Scoring included choice of treatment plan (non-operative treatment with plaster after closed reduction, or operative treatment) and therapeutic certainty on the treatment plan: very uncertain; uncertain; somewhat uncertain; certain; and very certain).

Therapeutic certainty was defined as how confident the surgeon was about his treatment plan. For example, if the surgeon was completely sure that he would treat a

patient operatively, he scored a five on the level of certainty. If he was unsure about the type of treatment, he scored a one or two on the level of certainty.

Statistical analysis. We determined the intra-observer agreement in two different ways. Firstly, we determined the intra-observer agreement on treatment plans for each surgeon separately and calculated the mean agreement for the four surgeons. Secondly, we analysed the intra-observer agreement by the surgeon's therapeutic certainty, scored at T1. Due to 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 intra-observer reliability for a radiograph.

T3 and T4 were used to determine the intra-observer reliability for a radiograph and 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 to 0.20 slight, 0.21 to 0.40 fair, 0.41 to 0.60 moderate, 0.61 to 0.80 substantial, and > 0.81 indicates almost perfect agreement.

Results

Study participants. During the study period, a post-reduction CT scan was undertaken for 85 patients who entered the Emergency Room with a displaced fracture of the distal radius. A total of 51 patients met the complete inclusion criteria (Fig. 1). Their mean age was 50 years (standard deviation (SD), 14)). 75% of the patients were female. The CT scan was performed at a mean of 2.53 days post-reduction (SD 2.21).

Table I. Kappa statistics of the four observers and the mean with 95% confidence interval (CI) in parentheses for treatment plan on plain radiographs and plain radiographs with an additional CT scan (radiograph and CT scan)

Intra-observer agreement for all cases		
Observer	Radiograph (T1 to T2)	Radiograph and CT scan (T3 to T4)
Observer 1	0.48 (0.23 to 0.72)	0.60 (0.39 to 0.81)
Observer 2	0.50 (0.14 to 0.86)	0.40 (0.14 to 0.66)
Observer 3	0.61 (0.34 to 0.87)	0.79 (0.39 to 1.00)
Observer 4	0.83 (0.67 to 0.99)	0.44 (0.23 to 0.65)
Mean	0.69 (0.58 to 0.79)	0.57 (0.45 to 0.69)

Table II. Kappa statistics based on the surgeon's level of certainty with 95% confidence interval (CI) in parentheses for treatment plan on plain radiographs and plain radiographs with an additional CT scan (radiograph and CT scan)

Intra-observer agreement, based on surgeon's level of certainty			
	Radiograph (T1 to T2)		Radiograph and CT scan (T3 to T4)
(Very) uncertain	-0.12 (-0.62 to 0.38)	No	0.57 (0.18 to 0.95) Moderate
Somewhat uncertain	0.19 (-0.11 to 0.48)	Slight	0.56 (0.26 to 0.87) Moderate
Certain	0.90 (0.76 to 1.00)	Almost perfect	0.47 (0.19 to 0.75) Moderate
Very certain	0.86 (0.76 to 0.96)	Almost perfect	0.60 (0.44 to 0.76) Moderate

Agreement treatment plans. The mean intra-observer agreement, regardless of the level of therapeutic certainty, on treatment plan based on radiograph is substantial (0.69). Adding a CT scan resulted in moderate agreement (0.57) (Table I).

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

For those cases where there was therapeutic uncertainty on the treatment plan based on radiograph, adding a CT improved the intra-observer agreement from 'none' to 'slight agreement based on radiograph, to 'moderate agreement' based on radiograph and CT.

For those cases where there was therapeutic certainty on the treatment plan based on radiograph, adding a CT worsened the intra-observer agreement, which decreased from almost perfect agreement (range 0.86 to 0.90) based on radiograph, to moderate agreement (range 0.47 to 0.60) based on radiograph and CT.

Discussion

Using radiographs alone, the level of therapeutic certainty is positively related to the intra-observer agreement, and even leading to no agreement when the surgeon is uncertain on the treatment plan. This influence is not seen on the intra-observer agreement based on radiograph and CT scan.

In therapeutically uncertain cases, the intra-observer agreement on treatment plan improves when an additional CT scan is used. In therapeutically certain cases, the agreement is already perfect, thus, there is little to 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 intra-observer 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 have led us to conclude differently and, therefore, may have been misleading for clinicians. These potentially misleading results showed us that the intra-observer agreement on treatment plan did not increase when using additional CT scanning for decision making in treatment plans for fractures of the distal radius. Moreover, it was even less reliable (radiograph alone: Kappa 0.69; radiograph 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 minimised spectrum bias, and so were able to determine the intra-observer agreement in a population with and without therapeutic uncertainty.

The controversy seen in literature regarding fractures of the distal radius on agreement in a treatment plan is also seen in other types of fracture, including fractures of the proximal humerus¹¹⁻¹³ and fractures of the tibial plateau.¹⁴⁻¹⁷ Although the CT scan has been shown to be

more accurate in assessing characteristics of fractures, 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.

The strength of our study is that all observers were experienced in judging imaging and treatment of fractures of the distal radius. As seen in many agreement studies, the average intra-observer agreement will probably be slighter lower when you have fewer experienced surgeons.⁹ However, we would still expect a similar pattern, that the agreement based on radiograph 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. In addition, the order of images was randomised, 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. Surgeons are generally confident in their decisions, which probably explains why these groups were relatively small.

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 fractures of the distal radius. However, this does not mean that the additional study has influence on the outcome of the patient. Prospective randomised studies should indicate whether the use of an additional CT scan and the resulting management in cases of therapeutic uncertainty would influence outcomes in patients with displaced fractures of the distal radius.

To the best of our knowledge, no previous agreement studies implemented the surgeon's level of certainty in their analysis to minimise 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 using this method to minimise spectrum bias.

In conclusion, our study shows that spectrum bias may influence the outcome of agreement studies on treatment plans. An additional CT scan improves the intra-observer agreement on plans for the treatment of fractures of the distal radius only when there is therapeutic uncertainty. Reporting and analysing intra-observer agreement based on the surgeon's level of certainty is an appropriate method of minimising spectrum bias.

References

1. Koval K, Haidukewych GJ, Service B, Zircgibel BJ. Controversies in the management of distal radius fractures. *J Am Acad Orthop Surg* 2014;22:566–575.
2. Cole RJ, Bindra RR, Evanoff BA, et al. 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;22:792–800.
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;19:720–727.
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;107:491–498.(Article in German).
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;92-A:2523–2532.
6. Katz MA, Beredjikian PK, Bozentka DJ, Steinberg DR. Computed tomography scanning of intra-articular distal radius fractures: does it influence treatment? *J Hand Surg Am* 2001;26:415–421.
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;83:676–678.
8. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;299:926–930.
9. Karanicolas PJ, Bhandari M, Kreder H, et al. Evaluating agreement: Conducting a reliability study. *J Bone Joint Surg [Am]* 2009;91-A(Suppl 3):99–106.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
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;6:38.
12. Ramappa AJ, Patel V, Goswami K, et al. Using computed tomography to assess proximal humerus fractures. *Am J Orthop (Belle Mead NJ)* 2014;43:E43–E47.
13. Brouwer KM, Lindenhovius AL, Dyer GS, et al. Diagnostic accuracy of 2- and 3-dimensional imaging and modeling of distal humerus fractures. *J Shoulder Elbow Surg* 2012;21:772–776.
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;41:173–178.
15. Chan PS, Klimkiewicz JJ, Luchetti WT, et al. Impact of CT scan on treatment plan and fracture classification of tibial plateau fractures. *J Orthop Trauma* 1997;11:484–489.
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;18:279–283.
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;16:632–637.

Funding statement:

- M. Bhandari reports funding received from Smith & Nephew, Stryker, Amgen, Zimmer, Moximed, Bioventus, Merck, Eli Lilly, Sanofi, Ferring, Commed, DePuy and Bioventus, none of which is related to this article. R. Poolman and V. Scholtes report funding for Joint Research Onze Lieve Vrouwe Gasthuis from ZonMw, Achmea, Link/Lima, Tornier, Stryker and NuVasive, none of which is related to this article.
- R. Poolman and V. Scholtes report funding for Joint Research Onze Lieve Vrouwe Gasthuis from ZonMw, Achmea, Link/Lima, Tornier, Stryker and NuVasive, none of which is related to this article.

Author contributions:

- Y. V. Kleinlugtenbelt: Design of study; Acquisition, analysis and interpretation of data; Writing and revision of manuscript
- M. Hoekstra: Acquisition, analysis and interpretation of data; Writing and revision of manuscript
- S. J. Ham: Acquisition of data; Revision of manuscript
- P. Kloen: Acquisition of data; Revision of manuscript
- R. Haverlag: Acquisition of data; Revision of manuscript
- M. P. Simons: Analysis and interpretation of data; Revision of manuscript
- M. Bhandari: Analysis and interpretation of data; Revision of manuscript
- J. C. Goslings: Analysis and interpretation of data; Revision of manuscript
- R. W. Poolman: Design of study; Analysis and interpretation of data, Revision of manuscript
- V. A. B. Scholtes: Design of study; Analysis and interpretation of data, Revision of manuscript

ICMJE Conflict of Interest:

- None declared
- ©2015 Kleinlugtenbelt et al. This is an open-access article distributed under the terms of the Creative Commons Attribution licence (CC-BY-NC), which permits unrestricted use, distribution, and reproduction in any medium, but not for commercial gain, provided the original author and source are credited.