CLINICAL RESEARCH METHODOLOGY IN SPINAL TRAUMA PATIENTS:

NATURAL EXPERIMENTS AND EQUIPOISE

AGNITA STADHOUDER

Clinical research methodology in spinal trauma patients

natural experiments and clinical equipoise

Agnita Stadhouder

2025

Colofon

CLINICAL RESEARCH METHODOLOGY IN SPINAL TRAUMA PATIENTS:
NATURAL EXPERIMENTS AND EQUIPOISE

Author: Agnita Stadhouder

Layout en design; Ester Picavet (esterpicavet.nl)

Printing: boekengilde.nl

ISBN: 9789464919271

Financial support for printing of this thesis was kindly provided by The Dutch Orthopedic Association (NOV), Euros, D.H.Heijne Stichting / Basko Healthcare, Inspine and LEUK Orthopedie.

NEDERLANDSE ORTHOPAEDISCHE VERENIGING | NOV

FSC *This thesis is printed on 100% FSC-certified paper (sourced entirely from responsibly managed forests or recycled paper)*

Clinical research methodology in spinal trauma patients

natural experiments and clinical equipoise

Klinische onderzoeks methoden in spinale trauma patiënten

Natuurlijke experimenten en equipoise

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. ir. W. Hazeleger,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

donderdag 15 mei 2025 des middags te 4.15 uur

door

Agnita Stadhouder

geboren op 14 oktober 1972
te Enkhuizen

Promotor:

Prof. dr. F.C. Öner

Copromotoren:

Dr. R.M. Houwert

Dr. S.P.J. Muijs

Beoordelingscommissie:

Prof. dr. F.W. Bloemers

Prof. dr. L.P.H. Leenen

Prof. dr. L.W. van Rhijn

Prof. dr. P.A.J.T. Robe (voorzitter)

Prof. dr. J.J. Verlaan

Contents

1

Introduction

CHAPTER 1

The history

The evolution of clinical research has a long history. The first description was recorded in the bible 562 BC in the book of Daniel. King Nebuchadnezzar of Babylon ordered his men to eat only meat and drink wine to improve their health. Some noblemen objected and the king allowed them to continue their diet of legumes and water for 10 days. At the end of the trial the vegetarian noblemen appeared healthier than the meat-eaters so the king allowed them to continue their diet.[1] Ibn-Sina (Avicenna) in 1025 also mentioned in his 'Canon of Medicine' certain rules of clinical research, especially in drug testing experiments. He mentions that in a clinical trial a remedy should be used in its natural state in disease. However, there is no record of the application of these principles in practice.[2] In 1747 James Lind performed the actual first controlled clinical trial on scurvy. He gathered 12 patients with scurvy and divided them in 6 groups each with a different diet. The 2 patients that took oranges and lemons had the most visible and good effect of the treatment.[3] Then, 150 years later in 1800 the word placebo was mentioned in medical research, but it took until 1863 when in the United States for the first-time a placebo-controlled trial was published by Austin Flint.[2,4] He treated 13 patients suffering from articular rheumatism with an herbal placebo and compared these patients with the standard therapy in those days. In 1943-44 the first double-blind controlled comparative trial was performed with the herb patulum (golden cup) against common cold. The trial was ordered by the Medical Research Council in the UK. Over a 1000 factory and office workers were enrolled, doctors and patients being blinded for the actual treatment. Results showed no difference between patulum and the control solution (placebo). Then, in 1946 the era of randomization started by the trial of streptomycin for tuberculosis, also performed in the UK, under strict supervision of a statistician, Dr. Hill. The study was published in 1948 in the British Medical Journal and is considered a landmark paper in clinical research methodology. The trial influenced virtually every area of clinical medicine since then.[2,5]

Epidemiology

From an epidemiological standpoint, clinical trials can be of two types: observational or experimental. Observational studies usually are used to generate a hypothesis. They are descriptive or analytic. Descriptive studies provide a description of the observed phenomena while analytical studies try to find an association between the measurements and the outcomes. Experimental studies are hypothesis-testing studies, where there is an intervention that tests the association between exposure and outcome.[6] Observational studies can be subdivided in case reports/ case series, ecologic studies, cross sectional studies, case control studies with a defined cohort, cohort studies and case cohort studies. Experimental study designs can be classified in two groups: controlled or uncontrolled. These can further be subdivided in three broad groups: clinical trials, field trials and com-

munity trials. Clinical trials are considered a 'gold standard' approach in epidemiological medical research. Clinical trials can be further subdivided in randomized trials, non-randomized trials, cross over and factorial trials.[6] Randomization is a methodology that is developed to prevent bias. The concept of bias is the lack of internal validity or incorrect assessment of the association between an exposure and an effect in the target population.[7] Kleinbaum et al divided bias in medical research in three main groups: selection bias, information bias and confounding.[8] Selection bias is when there is a systematic error in including subjects for a study, thus affecting the external validity of the study. Subjects are not a good representative of the population being studied and therefore the results cannot be generalized. It can also influence the internal validity of the study if the selection of subjects of treatment groups is influenced by certain factors. To decrease selection bias, one should be aware and select subjects that are representative of the population studied or use randomization techniques. Information bias is present when there is a systematic error in obtaining data from the study subjects. This can also be on the part of the investigator, where the investigator is influenced by certain characteristics of the group, making it observer bias. Blinding of the patients/ study results can prevent this form of bias.[6] The simple definition of confounding is "the confusion of effects". A confounding variable is a variable that correlates (positively or negatively) with both the exposure and the outcome. Especially in observational studies this can be a problem since they are not randomized. If a possible confounding is not recognized this results often in a distorted or incorrect association of treatment effect.[9] Selection bias can be a common cause of confounding. This happens when one or more of the predictor variables that determine the assignment to the intervention also directly affects the outcome. This leads to a type I error in which the outcome of the intervention is falsely attributed to the intervention while it is actually the confounding variable causing the effect. It can also lead to a Type II error on the other side, when the study incorrectly concludes that there is no treatment effect from the intervention.[9] Preventing confounding is very important in clinical research and is performed by implementing different study designs as randomization and participant matching. Most confounding though is removed by statistical procedures when analyzing data (multivariable regression analysis, propensity score methods) of clinical trials, especially in observational studies.[9]

Clinical Research in Surgical Fields

The aforementioned summary shows clearly that performing good clinical research is always a challenge. Performing good quality clinical research in surgical patients is even more a challenge and may require alternative methodologies. Already in 1996 the editor of the Lancet wrote an editorial with the title: "Surgical research or comic opera: questions, but few answers."[10] After anesthesia and antiseptic techniques were developed surgical techniques were rapidly developed for many different conditions that could not be treated before. Many operative procedures were therefore introduced before the concept of randomized controlled trials was

developed. Once a treatment has been accepted as a standard treatment, testing this in a RCT (Randomized Controlled Trial) setting becomes problematic.[11] In the surgical profession other factors leading to conducting less RCT`s are the learning curve of operative procedures, influence of commercial parties, difficulties in blinding, lack of funding, lack of education in clinical epidemiology, recruitment of patients with rare conditions, lack of agreement on standard outcome measurements and difficulties of performing randomization in emergency settings as compared to conducting non-surgical RCT's.[11,12] But there is progress, Hanzlik et al looked at the quality of publications in the Journal of Bone and Joint Surgery in 1975, 1985, 1995 and 2005. They concluded that the percentage of Level-I studies increased from 4% in 1975 to 21% in 2005. The average level of evidence rating also improved from 3.72 to 2.90.[13] In 2013 another article analyzed the published surgical RCT's between 1999 and 2009. In this extensive review the number of surgical papers mentioning RCT methodology increased from 300 in 1999 to 450 in 2009 (50%), although this was especially in other continents than North America where there was a decrease of 23%. The quality of these RCT's was also assessed according to the Cochrane guidelines by means of a 9-item list. They defined 'low risk of bias trials' when there was adequate generation of allocation, adequate concealment of allocation, intention-to-treat analysis, and adequate handling of dropouts. The proportion of low risk of bias trials increased from 6% in 1999 to 14% in 2009. There were important geographical differences in the quality of RCT's in those 10 years. The quality of European trials improved substantially from 7.5% in 1999 of low risk of bias to 23% in 2009 (PR 3.03; 95% CI 1.65–5.52; P < 0.001). North American studies also improved from 2.8% low risk of bias trials in 2009 to 16% in 2009 (PR 5.79; 95% CI 1.30–25.7; P = 0.01). Trials from Asia/Oceania did not show any improvement in reported methodological characteristics over these 10 years, with 5% low risk of bias trials in both years. The authors discussed that this was a worrisome development as there was no sign of improvement over the past 10 years. Countrywise, The Netherlands had the highest rate of low risk of bias trials (50%).[14] In 2023 an update of this review was presented with the same methodology on surgical research papers until the year 2019. In this study, 438 papers were included, which is a stable number compared to 2009. Gastrointestinal/ oncologic surgery was the most common subspecialty (50.1%) using this design while trauma surgery had only 3.6% of the total published surgical RCT's. The quality of studies improved especially in Asia where the percentage of low risk of bias studies increased to 18.1 percent (RR 3.50, 1.70 to 7.32; P < 0.001). RCT's from Africa/ South America still remained very low in risk of bias (<10%) and Europe and North America did not significantly improve. The authors discuss that the stable number of RCT's shows that a steady state might have been reached, which is not necessarily a negative development. The quality did improve the last decade and that should be the aim of surgical research in the coming decade.[15] Robinson et al also performed a systematic review about the characteristics of surgical randomized clinical trials between 2008 and 2020 and identified 388 papers in two high-impact surgical

journals. 29.9% of these trials were in the field of orthopedic surgery; trauma surgery was not specified as a subcategory. The authors concluded that registration of the trials was suboptimal and in 31.5% of the registered trials there was a discrepancy between the registered and published trial. Also interesting, 78.1% of the trials did not control for surgeon expertise. The authors concluded that 'these data suggest that improvements in the design, implementation and reporting of randomized clinical trials in surgery are warranted'.[16]

Research in Trauma and spinal trauma

There are several possible explanations why clinical research in trauma patients seems to be more challenging than surgical research in general: Trauma surgeons often have a preference for certain operative or conservative treatments because of their training, past experiences and technical skills. The local hospital culture plays an important role in how patients are treated and the infrastructure and daily routines in the hospital taking care of trauma patients is difficult to change when alternative treatment modalities are introduced. Then treatments differ substantially in trauma care which makes surgeons reluctant to include certain trauma patients in a RCT. Further, some of these conditions are infrequent pathologies preventing enrollment of patients delaying inclusion in the fast-developing surgical field. In trauma patients, urgent clinical decision making may be necessary which makes recruitment extra problematic when informed consent is needed before inclusion and treatment.[17-19] In spinal trauma research, all of the above apply but there are extra contributing factors in the scarcity of RCT`s in this field. One of the attributing difficulties is the variation in classification system which have been used in spinal injuries. In the past 50 years several classification systems have been used.[20-23] With the introduction in 2013 and 2016 of the AOSpine thoracolumbar and subaxial injury classification this hurdle might have been overcome.[20-24] In 2021 the classification was extended with a surgical decision making tool, the AOSpine Injury Score (AOSIS). Another complicating factor in spinal trauma research was the absence of a universal disease-specific outcome instrument for spinal trauma patients. For decades spinal trauma patient outcome was scored with pain-, general outcome- and chronic low back pain outcome measurements. Issues specific to spinal trauma patients are not measured adequately in this way.[25-27] Work has been done on specific spinal trauma patient populations. Most studies are from a critical-care prospective or have their focus on Spinal Cord Injury (SCI) patients[28, 29] But these patients represent a minority of all spinal trauma patients.[27] Also here the AO Spine Knowledge Forum Trauma took initiative and developed universal disease-specific outcome instruments for spine trauma patients.[26] This development was done in a thorough way by identifying components of the International Classification of Functioning, Disability, and Health (ICF) of the World Health Organization (WHO). Then 150 spine experts from all world regions identified 13 ICF categories as most relevant through a web-based survey.[30] The patients` perspective and measurement format was also investigated.[31] A first draft of the AOSpine Patient Reported

Outcome Spine Trauma (AOSpine PROST) was developed containing 19 items.[26] The AOSpine PROST was further validated and implemented and will be another valuable tool in improving the quality of spinal trauma research.[32-34] Beside these issues that have almost been resolved, there remains a serious variation in the treatment of most of these injuries, thus one major research question still needs to be answered: what is the optimal treatment for patients with spinal injuries?[35] In other words how can we create clarity in the existing equipoise about the best treatment for these patients?

Clinical equipoise and natural experiment studies

Clinical equipoise, as described by Freedman, exists when there is a genuine uncertainty within the expert medical community about the optimal treatment of a certain disease.[36] In 2008 this concept was introduced in a retrospective comparative study on spinal fractures.[37,38] Since then there is an increasing number of clinical studies that use equipoise as a starting point for clinical research. In 2023 there were 212 publications in PubMed with equipoise as a headword comparing to 53 in 2008 when it was originally introduced for comparative clinical research. In the presence of equipoise in a clinical setting it is common to form `schools` based on convictions of clinical superiority among treating doctors or surgeons.[37,39] In spinal trauma care especially, non-surgical and surgical schools have become well recognized and established in different regions and hospitals. All with adequate resources and clinical experience to perform optimal care in either a non-operative or operative way.[39] Instead of forcing surgeons in a prospective randomized clinical trial setup the historical `school` formation can be used in an observational study design using the expertise in non-operative and operative treatment as an advantage in optimal treatment management.[37]

Although the concept of 'equipoise' has been regularly used in other contexts, the threshold of disagreement has not been clearly defined. Medical ethics researchers in 2 papers suggested a trial to be unethical when agreement among experts is above 70% or 80%.[40-42] In subcategories of spinal trauma, this 70-80% threshold of clinical equipoise is not met and invites the spinal community to perform good quality research.

Methodologically, RCT's as discussed above, are not the preferred method in this kind of populations. Observational studies are increasingly believed to provide reliable evidence complementary to RCT's, provided they are of sufficient quality.[19,43] Vandenbroucke advised that to ensure similar credibility of observational studies compared to randomized studies, three essential restrictions should be taken into account. First the selection of research topics is limited to where allocation of exposure is minimally associated with the outcome of interest. The second restriction involves that a study design is required to have at least a quasi-random allocation of exposure to treatment. The third is restriction to topics where potential confounding variables can be identified, accurately measured, and appropriately adjusted for in statistical analysis.[44] Having said this, observational

studies are more representative of common daily practice where physicians can perform the treatment they prefer. These studies are much faster in generating results and are less costly.[19] However, without randomization, incomparability of treatment groups may lead to confounding bias.[19] To overcome this, natural experiments (NEs) as a type of observational study can be introduced. Natural experiments (NEs) have a long history in public health research, dating back to John Snow's classic study of London's cholera epidemics in 1854 where Dr. John Snow identified a specific street pump as the source of an intense cholera outbreak by plotting the location of cholera deaths on a dot-map.[45] Recently NEs and other alternatives to RCTs are implemented as an alternative to evaluate health interventions, not only in public health care but more and more in clinical research.[19,46,47] Natural Experiments are described by van de Wall et al in their introduction paper as `observational studies in which patients are exposed to either the experimental or the control condition, whereby treatment allocation is determined by factors outside the control of the investigators'. The process governing treatment allocation arguably resembles the random assignment in an experimental setting, hence the name natural experiment.[19]

Aim of this thesis

Design and conduction of good quality research in trauma patients is not an easy task. There is a wide variation in patient presentation; clinical equipoise is strongly felt among spine surgeons about the optimal conservative or operative treatment; outcome instruments are not yet well defined and accepted making randomization of spinal trauma patients extremely challenging in an acute trauma setting. There is a need for alternative methodologies to provide answers in this diverse patient group.

This thesis consists of four sections. The first section describes the common practice in spinal trauma care in the Netherlands and a clinical RCT performed on the conservative treatment of spinal fractures. The second part explores the different outcome scores for patients with a traumatic spinal injury. The third part introduces clinical equipoise as the starting point of a comparative study on non-operative and operative treatment in spinal trauma patients. The theoretical concept is discussed. This concept is brought into practice in a large retrospective cohort study comparing operative and non-operative schools. The fourth part builds upon this concept and introduces the natural experiment in a (spinal) trauma setting as a good alternative for RCT`s. Last, a review is performed to explore the use of natural experiments in spinal trauma and investigates if this is a viable good quality alternative for RCT`s.

1. Collier R. Legumes, lemons and streptomycin: a short history of the clinical trial. CMAJ. *2009;180(1):23-4*.

2. Bhatt A. Evolution of clinical research: a history before and beyond james lind. Perspect Clin Res. *2010;1(1):6-10*.

3. Hughes RE. James Lind and the cure of scurvy: an experimental approach. Med Hist. *1975;19(4):342-51*.

4. Flint A. ART. I.--*A contribution toward the Natural History of Articular Rhcumatism; consisting of a report of thirteen cases treated solely with palliative measures. The American Journal of the Medical Sciences (1827-1924). 1863;46(91):17*.

5. Hill CA, Riley C, Gifford JH. The Use of Intrathecal Heparin in Conjunction with Streptomycin in the Treatment of Tuberculous Meningitis: Preliminary Report. J Clin Pathol. *1948;1(3):157-61*.

6. Chidambaram AG, Josephson M. Clinical research study designs: The essentials. Pediatr Investig. *2019;3(4):245-52*.

7. Delgado-Rodriguez M, Llorca J. Bias. J Epidemiol Community Health. *2004;58(8):635-41*.

8. Kleinbaum DG, Morgenstern H, Kupper LL. selection bias in epidemiologic studies. American Journal of Epidemiology. *1981;113(4):452-63*.

9. Vetter TR, Mascha EJ. Bias, Confounding, and Interaction: Lions and Tigers, and Bears, Oh My! Anesth Analg. 2017;125(3):1042-8.

10. Horton R. Surgical research or comic opera: Questions, but few answers. Lancet. *1996;347(9007):984*.

11. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. BMJ. *2002;324(7351):1448-51*.

12. Hirst A, Agha RA, Rosin D, McCulloch P. How can we improve surgical research and innovation?: the IDEAL framework for action. Int J Surg. *2013;11(10):1038-42*.

13. Hanzlik S, Mahabir RC, Baynosa RC, Khiabani KT. Levels of evidence in research published in The Journal of Bone and Joint Surgery (American Volume) over the last thirty years. J Bone Joint Surg Am. *2009;91(2):425-8*.

14. Ahmed Ali U, van der Sluis PC, Issa Y, Habaga IA, Gooszen HG, Flum DR, et al. Trends in worldwide volume and methodological quality of surgical randomized controlled trials. Ann Surg. *2013;258(2):199-207*.

15. Pronk AJM, Roelofs A, Flum DR, Bonjer HJ, Abu Hilal M, Dijkgraaf MGW, et al. Two decades of surgical randomized controlled trials: worldwide trends in volume and methodological quality. Br J Surg. *2023;110(10):1300-8*.

16. Robinson NB, Fremes S, Hameed I, Rahouma M, Weidenmann V, Demetres M, et al. Characteristics of Randomized Clinical Trials in Surgery From 2008 to 2020: A Systematic Review. JAMA Netw Open. *2021;4(6):e2114494*.

17. Stadhouder A, Öner FC, Wilson KW, Vaccaro AR, Williamson OD, Verbout AJ, et al. Surgeon equipoise as an inclusion criterion for the evaluation of nonoperative versus operative treatment of thoracolumbar spinal injuries. Spine J. *2008;8(6):975-81*.

18. Stadhouder A, van Rossenberg LX, Kik C, Muijs SPJ, Öner FC, Houwert RM. Natural Experiments as a Study Method in Spinal Trauma Surgery: A Systematic Review. Global Spine J. *2023:21925682231220889*.

19. van de Wall BJM, Stadhouder A, Houwert RM, Öner FC, Beeres FJP, Groenwold RHH, Group NS. Natural experiments for orthopaedic trauma research: An introduction. Injury. *2023;54(2):429-34*.

20. Denis F. The three column spine and its significance in the classification of acute thoracolumbar spinal injuries. Spine (Phila Pa 1976). *1983;8(8):817-31*.

21. Magerl F, Aebi M, Gertzbein SD, Harms J, Nazarian S. A comprehensive classification of thoracic and lumbar injuries. Eur Spine J. *1994;3(4):184-201*.

22. McCormack T, Karaikovic E, Gaines RW. The load sharing classification of spine fractures. Spine (Phila Pa 1976). *1994;19(15):1741-4*.

23. Vaccaro AR, Öner C, Kepler CK, Dvorak M, Schnake K, Bellabarba C, et al. AOSpine thoracolumbar spine injury classification system: fracture description, neurological status, and key modifiers. Spine (Phila Pa 1976). *2013;38(23):2028-37*.

24. Vaccaro AR, Koerner JD, Radcliff KE, Öner FC, Reinhold M, Schnake KJ, et al. AOSpine subaxial cervical spine injury classification system. Eur Spine J. *2016;25(7):2173-84*.

25. Muller U, Roder C, Greenough CG. Back related outcome assessment instruments. Eur Spine J. *2006;15 Suppl 1(Suppl 1):S25-31*.

26. Sadiqi S, Lehr AM, Post MW, Dvorak MF, Kandziora F, Rajasekaran S, et al. Development of the AOSpine Patient Reported Outcome Spine Trauma (AOSpine PROST): a universal disease-specific outcome instrument for individuals with traumatic spinal column injury. Eur Spine J. *2017;26(5):1550-7*.

27. Stadhouder A, Buckens CF, Holtslag HR, Öner FC. Are existing outcome instruments suitable for assessment of spinal trauma patients? J Neurosurg Spine. 2010;13(5):638-47.

28. Akmal M, Trivedi R, Sutcliffe J. Functional outcome in trauma patients with spinal injury. Spine (Phila Pa 1976). *2003;28(2):180-5*.

29. Holtslag HR, Post MW, Lindeman E, Van der Werken C. Long-term functional health status of severely injured patients. Injury. *2007;38(3):280-9*.

30. Öner FC, Sadiqi S, Lehr AM, Dvorak MF, Aarabi B, Chapman JR, et al. Towards the development of an outcome instrument for spinal trauma: an international survey of spinal surgeons. Spine (Phila Pa 1976). *2015;40(2):E91-6*.

31. Öner FC, Sadiqi S, Lehr AM, Aarabi B, Dunn RN, Dvorak MF, et al. Toward developing a specific outcome instrument for spine trauma: an empirical cross-sectional multicenter ICF-based study by AOSpine Knowledge Forum Trauma. Spine (Phila Pa 1976). *2015;40(17):1371-9*.

32. Buijs GS, de Gendt EEA, Sadiqi S, Post MW, Muijs SPJ, Öner FC. Long-Term Reliability and Validity of the AO Spine PROST (Patient-Reported Outcome Spine Trauma). Spine (Phila Pa 1976). *2022;47(17):E562-E9*.

33. Sadiqi S, de Gendt EEA, Muijs SPJ, Post MWM, Benneker LM, Holas M, et al. Validation of the AO Spine CROST (Clinician Reported Outcome Spine Trauma) in the clinical setting. Eur Spine J. *2024;33(4):1607-16*.

34. Sadiqi S, Dvorak MF, Vaccaro AR, Schroeder GD, Post MW, Benneker LM, et al. Reliability and Validity of the English Version of the AOSpine PROST (Patient Reported Outcome Spine Trauma). Spine (Phila Pa 1976). *2020;45(17):E1111-E8*.

35. Kweh BTS, Tee JW, Dandurand C, Vaccaro AR, Lorin BM, Schnake K, et al. The AO Spine Thoracolumbar Injury Classification System and Treatment Algorithm in Decision Making for Thoracolumbar Burst Fractures Without Neurologic Deficit. Global Spine J. *2024;14(1_suppl):32S-40S*.

36. Freedman B. Equipoise and the ethics of clinical research. N Engl J Med. *1987;317(3):141-5*.

37. Stadhouder A, Buskens E, de Klerk LW, Verhaar JA, Dhert WA, Verbout AJ, et al. Traumatic thoracic and lumbar spinal fractures: operative or nonoperative treatment: comparison of two treatment strategies by means of surgeon equipoise. Spine (Phila Pa 1976). *2008;33(9):1006-17*.

38. Stadhouder A, Buskens E, Vergroesen DA, Fidler MW, de Nies F, Öner FC. Nonoperative treatment of thoracic and lumbar spine fractures: a prospective randomized study of different treatment options. J Orthop Trauma. *2009;23(8):588-94*.

39. Dandurand C, Öner CF, Hazenbiller O, Bransford RJ, Schnake K, Vaccaro AR, et al. Understanding Decision Making as It Influences Treatment in Thoracolumbar Burst Fractures Without Neurological Deficit: Conceptual Framework and Methodology. Global Spine J. *2024;14(1_suppl):8S-16S*.

40. Dandurand C, Dvorak MF, Hazenbiller O, Bransford RJ, Schnake KJ, Vaccaro AR, et al. Using Equipoise to Determine the Radiographic Characteristics Leading to Agreement on Best Treatment for Thoracolumbar Burst Fractures Without Neurologic Deficits. Global Spine J. *2024;14(1_suppl):25S-31S*.

41. Johnson N, Lilford RJ, Brazier W. At what level of collective equipoise does a clinical trial become ethical? J Med Ethics. 1991;17(1):30-4.

42. Mhaskar R, B BB, Djulbegovic B. At what level of collective equipoise does a randomized clinical trial become ethical for the members of institutional review board/ethical committees? Acta Inform Med. 2013;21(3):156-9.

43. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. *2000;342(25):1887-92*.

44. Vandenbroucke JP. When are observational studies as credible as randomised trials? Lancet. 2004;363(9422):1728-31.

45. McLeod KS. Our sense of Snow: the myth of John Snow in medical geography. Soc Sci Med. *2000;50(7-8):923-35*.

46. Hoepelman RJ, Beeres FJP, Beks RB, Sweet AAR, Ijpma FF, Lansink KWW, et al Non-operative vs. operative treatment for multiple rib fractures after blunt thoracic trauma: a multicenter prospective cohort study. Eur J Trauma Emerg Surg. *2023;49(1):461-71*.

47. Hoepelman RJ, Minervini F, Beeres FJP, van Wageningen B, FF IJ, van Veelen NM, et al. Quality of life and clinical outcomes of operatively treated patients with flail chest injuries: A multicentre prospective cohort study. Front Surg. *2023;10:1156489*.

2

Management of spinal trauma patients: a national survey in The Netherlands

Agnita Stadhouder
Sayf Faraj
F. Cumhur Öner

Acta Orthop. Belg., 2019, 85, 40-46

ABSTRACT

Lack of consensus in spinal trauma management and differences in the practical organization between trauma regions can have significant consequences on the fate of patients with spine trauma. For this reason a national survey was conducted among the 11 trauma regions in the Netherlands. Representative surgeons were sent a survery on seven areas of spinal trauma management: treatment protocol, referral, advisory committee, classification used, responsible medical specialist, timing of surgical intervention, and the current view on spinal trauma care.
All 11 centers completed the survey yielding a response rate of 100%.
The results of this study shows that in a relative small country, all seven areas in the management of spine trauma differs substantially and can be of use to show the possible areas of discrepancies between trauma centers in comparable European countries.

INTRODUCTION

Spinal trauma with or without spinal cord injury (SCI) may lead to significant disability with poor functional outcomes.[7] Motor vehicle accidents, falls, violence, and sports are the leading causes of spinal injuries.[3,20] Associated neurologic damage is a cause of lasting and serious disability. A worldwide SCI incidence ranging from 10.4 to 83 per million inhabitants per year has been reported in which males are disproportionately affected with a male-female ratio of 4:1 and a mean age of 33 years. [3,23] The Netherlands is one of the enlisted countries with the lowest reported incidence of a mere 10.4 per million inhabitants per year.[2,23] Despite these relatively low rates, SCI has not only been associated with a negative impact on the lives of sufferers, but also with extremely high economic costs.[21]

Most of the thoracolumbar spinal fractures without neurologic involvement are treated nonoperatively with favorable long-term outcomes.[15] In recent literature, reviewing operative versus nonoperative treatment in thoracic and lumbar fractures, no definite conclusions could be drawn with regards to complication rates and long-term outcome between the two methods.[8,10] The common treatment of SCI is surgical stabilization followed by rehabilitation and complication prevention.[19] However, SCI remains a heterogeneous group of injuries and therefore various treatments can be associated with good clinical outcomes. On top of that it has to be noted that not every hospital has the proper facilities to give optimal care to trauma patients with SCI and therefore specific criteria are mentioned in SCI guidelines to determine whether patients should be transported to a specialized trauma center.[17] However, the choice of optimal treatment remains difficult to determine due to the limited number of high-quality studies and the multiple clinical variables that accompany spinal trauma (e.g. the degree of ligamentous and bone injury, the presence of neurologic deficits, associated other traumatic lesions and overall health status).

Throughout the literature several conservative and surgical procedures have been mentioned and proposed, and numerous studies on the management of traumatic SCI have been conducted. However, to date, there is lack of consensus in treatment with regard to fracture and neurologic deficit, classification, scoring system, the decision to operate, ideal timing for surgery, and surgical approaches. [1,9,13,16] Organization of trauma care in a country or region can have significant consequences on the fate of patients with spinal column injuries. In the Netherlands, the Ministry of Health appointed in 1997, 11 trauma centers, each responsible for emergency health care in their region. The goal is to create intensive collaboration between different hospitals in a trauma region, as with Medical Mobile Teams and Ambulances. Trauma protocols are synchronized between these regional hospitals and there is a regional registration of trauma patients. Another initiative in optimizing Spinal Care comes from the Spinal Cord Injury Organization Netherlands (DON). This patients' organization with 1300 members was founded in 1976. They presented a health care report in 2013 on Spinal Cord Injury with the intention to investigate the complete pathway of healthcare from patients' perspective. The

report was supported by the Dutch Flemish Spinal Cord Injury Society, the Dutch Spine Society, The National Society Acute Health Care and the Dutch Society of Neurology.[24]

Despite these guidelines from professionals' and patients' perspectives the practical organization and management of spinal trauma patients and differences between trauma regions are largely unknown. We conducted a survey among the trauma regions for the purpose of clarification of these differences.

MATERIALS AND METHOD

We approached all 11 trauma centers and asked them to appoint a representative surgeon involved in the acute care of spinal trauma patients in their regions. All centers received an invitation to participate in the study. A repeat email was sent to non-responders after 4, 6 and 8 weeks. After 12 weeks, physicians were contacted by phone. No financial compensation was granted to participants.

The survery consisted of 9 multiple choice questions and 7 open questions on seven areas of spinal trauma management: (1) treatment protocol, (2) referral, (3) advisory committee, (4) classification used, (5) responsible medical specialist in spinal trauma care, (6) timing of surgical intervention, and (7) the current view of health care professionals involved in the management of spinal trauma patients.

Data was collected from September 2013 to December 2014. All responses were manually recorded and analysed with Microsoft Excel 2011.

RESULTS

All the 11 centers completed the survey yielding a response rate of 100%. Trauma centers were represented by a neurosurgeon, orthopaedic surgeon or general trauma surgeon. Eight of the 11 trauma centers have a protocol on the care, transfer, and treatment of patients with spine trauma in the region. Table I provides inisght in the treatment protocols of the 11 trauma centers regarding spinal injury.

All 11 trauma centres have an advisory board regarding spinal trauma patients (Table II). This advisory board sets the policy for patients with traumatic spinal injury and consists of a board of medical doctors with various background specialities. Neurosurgeons were present in all trauma centres' advisory boards. Orthopaedic surgeons in 10 out of 11. Varying between trauma centres, trauma surgeons, general surgeons and rehabilitation physicians supported neurosurgeons and orthopaedic surgeons. In some cases a neurologist, intensive-care physician, or a radiologist constituted support (Figure 1).

In nine of the 11 trauma centers both neurosurgeons and orthopaedic surgeon are together responsible for surgical treatment, spinal surgery, cervical or thoracolumbar. In the other 2 centers both trauma general surgeons and orthopaedic surgeons operate on thoracolumbar spinal fractures. In these 2 centers neuro-

surgeons are responsible for operative procedures of the cervical spine, with or without neurological deficit (Figure 2).

Patients suffering from neurological deficit due to spinal trauma are transported to a trauma center or a hospital specialized in this type of injury. A ratio of 1:1.6 was found when comparing available hospitals in the regions of the trauma centeres for spinal trauma with neurological deficit versus without neurological deficit, respectively (Table III).

In cervical spine fractures the SLIC, AO and AO revised classifications are used. Five trauma centers did not use a classification system in the assessment of patients with cervical spinal fractures. The assessment of patients with thoracolumbar fractures varies from AO, AO revised, and TLICS classification. Five trauma centers use a combination of the classification systems, 3 use the AO-Magerl classification, 2 use the AOSpine revised (this was just published during data gathering), and 1 uses the TLICS classification.

The classification system for neurological deficit due to spinal trauma is more straightforward than the classification for spinal fractures. Of the 11 trauma centres, 6 use the American Spinal Injury Association classification (ASIA), 3 use Frankel and 2 reported to use both classification systems (Table IV).

In patients with incomplete neurological deficit: 6 trauma centers operate within 6 hours, 4 trauma centers within 24 hours, and 1 trauma center within 48 hours. Patients suffering from complete neurological deficit after spinal trauma are less likely to be operated within 6 hours (only 3 trauma centers). Trauma centers prefer to wait longer before performing an operation in patients with complete neurological deficit (Figure 3).

Questions	Yes	No	Partly
Is there a protocol on the care, transfer, and treatment of patients with spine trauma in the region?	8	2	1
Is the trauma center in the region the same as the neurosurgery center?	11	0	0
Are all patients with suspected neurological deficit transferred to a trauma center in the region?	11	0	0

Table I. - Questions on spinal trauma treatment protocols

Questions	Yes	No
Is there a specific advisory board of spinal trauma patients in the trauma center?	11	0

Table II. - Presence of an advisory board regarding spine trauma

Questions	Total N of hospitals
How many hospitals are eligble to treat patients with spinal fracture without neurological impairment?	28
How many hospitals are eligble to treat patients with spinal fracture with neurological impairment?	17

Table III. - Questions on the management of spinal trauma patients with and without neurological deficit

Spinal injury	Classification			
Cervical fractures	SLIC	AO	AO Revised	None
	2	2	2	5
Thoracolumbar fracture	TLICS	AO	AO Revised	Combination
	1	3	2	5
Neurological impairment	ASIA	Frankel	Combination	
	6	3	2	

Table IV. – Classification of spinal injuries

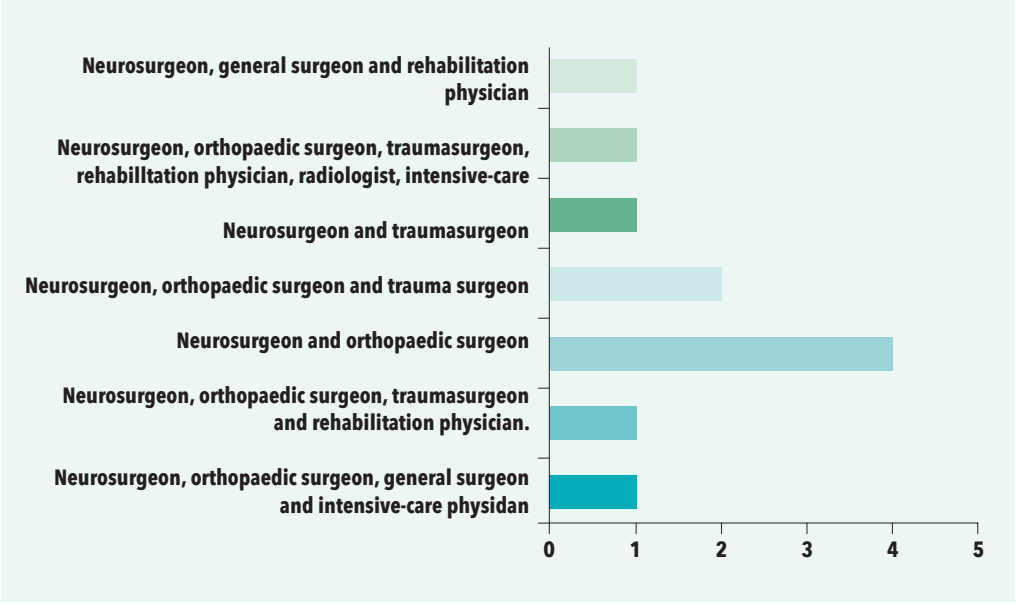

Fig. 1. – Advisory board on traumatic spinal injury

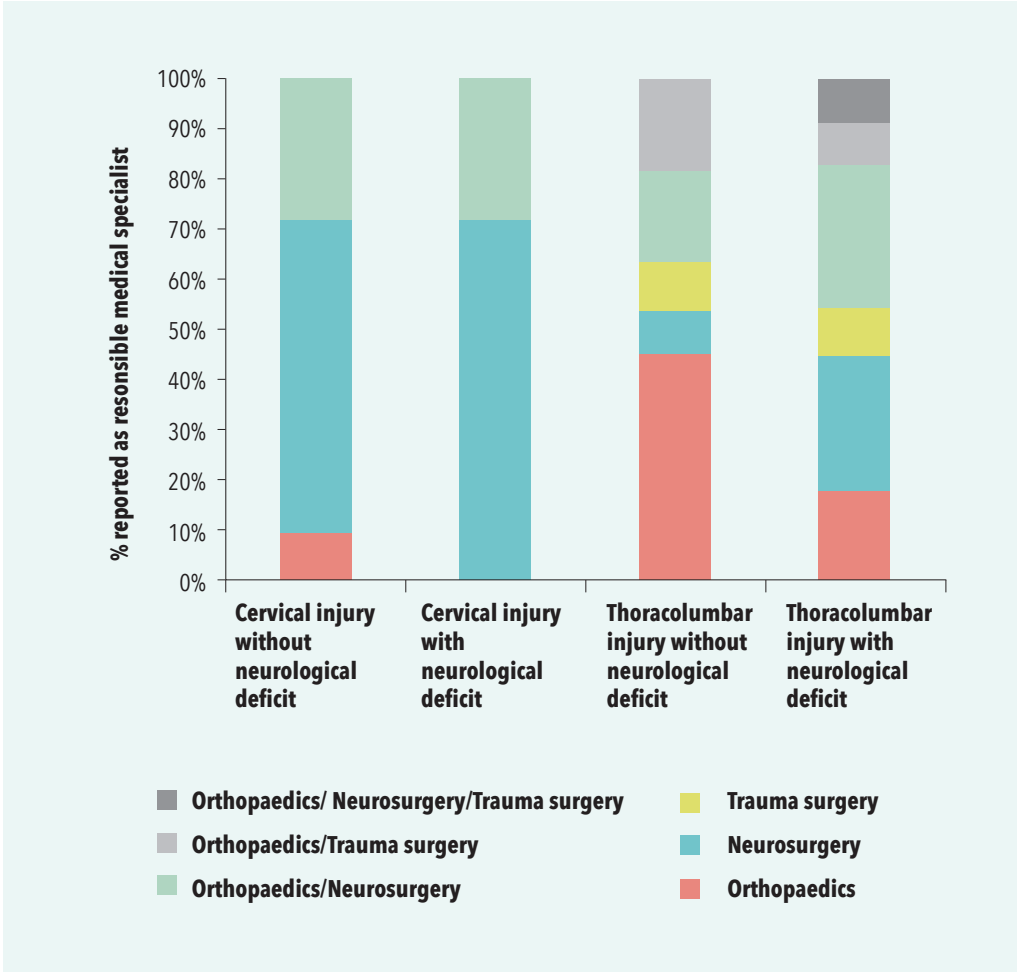

Fig. 2. – Responsible medical specialist in spinal trauma care

All participants were asked to grade spinal trauma management in their region with a score between 0 and 10. This resulted in 5 being the lowest grade awarded and 10 the highest, and an average of 7.7 (range 5-10) points given. Sixty-four percentage of the health care professionals involved in spinal trauma care answered that there is a need for a more concentrated care for patients with spinal trauma (Table V).

Eight participants gave suggestions to improve the management around spinal trauma patients. In short, seven participants of the nationwide health care professionals involved in the management of spinal trauma patients agreed there is need for a more centralized management of spinal trauma. In addition the following suggestions were made:

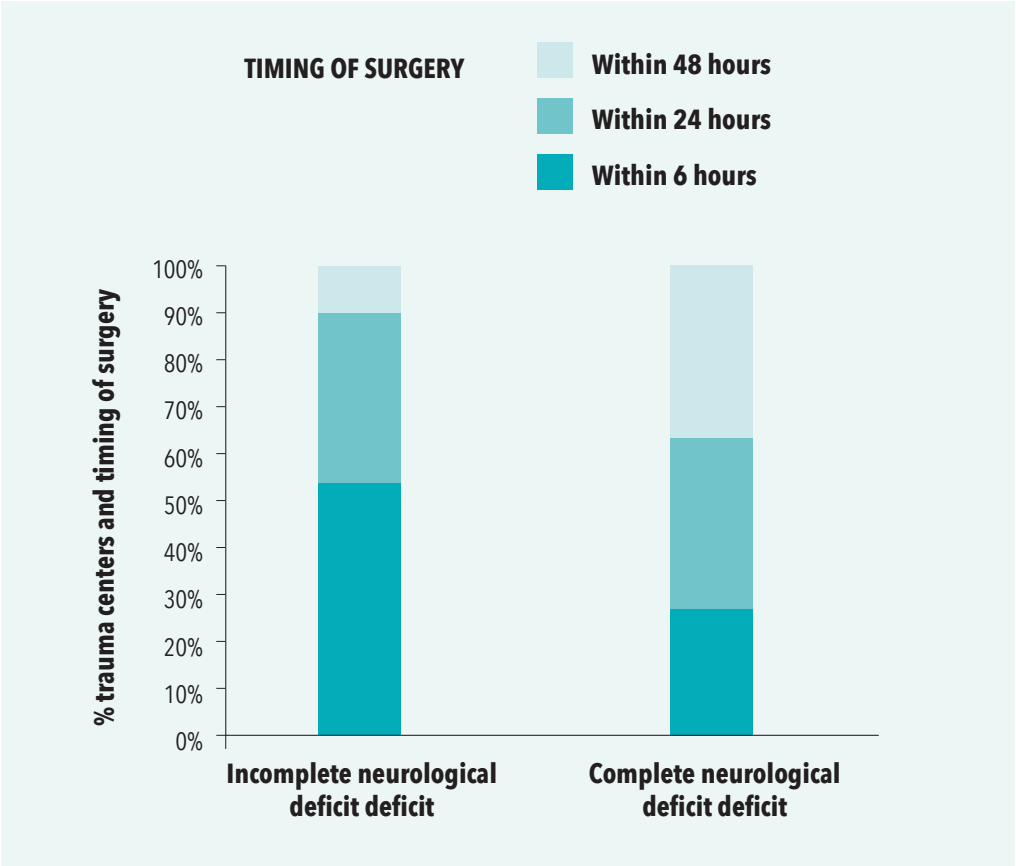

Fig. 3. – Comparison timing of surgery after traumatic spinal injury: incomplete versus complete neurological deficit

Spinal injury	Classification			
Cervical fractures	SLIC	AO	AO Revised	None
	2	2	2	5
Thoracolumbar fracture	TLICS	AO	AO Revised	Combination
	1	3	2	5
Neurological impairment	ASIA	Frankel	Combination	
	6	3	2	

Table IV. – Classification of spinal injuries

Questions	Yes	No
Is there need for a more concentrated care of patients with spinal injury?	8	2
Is there a co-operation with the rehabilitation of patients with spinal injury in the trauma region?	11	0

Table V. - Questions on the need for concentrated care for patients with spinal trauma

- There is need for clearer classification and referral guidelines for clinics not specialized in spinal injuries;
- There is need for an improved standardized evaluation when patients with spinal trauma arrive at the emergency room;
- There is need for more collaboration with trauma general surgeons;
- There should be a better transfer of imaging data;
- Investments in a digital communication network are desired;
- Establishing a team unit with surgeons, rehabilitation and intensive-care physicians is suggested;
- All spinal trauma patients should be directed immediately to the level-1 trauma center in the region;
- There should be specific demands for surgical health care professionals involved with this type of injury, and;
- Surgeons should be up to date with the recent developments and scientific research and perform a minimum, sufficient number of spinal operations.

Three participants had no comments or suggestions. However, in one region there is a regional think tank with all spine surgeons that meets biannually. Additionally there is frequent consultation on clinical cases (two to three times a week). This group graded their spinal trauma management with an 8 out of 10.

DISCUSSION

This survey reveals some variations in the initial assessment and treatment among the 11 trauma centers in the Netherlands. Besides large variations in the composition of advisory committees on management of traumatic spinal injury, there are variations in policy concerning classification systems, leading practitioners, and timing of surgical intervention. Ultimately, the survey demonstrates that more concentrated care and better communication is required for the optimal management in patients with traumatic spinal injury.

In the guideline on Acute Traumatic Spinal Injuries they notice the differences between the regional trauma centers but cannot conclude if there is a difference in quality in treatment of patients with spinal cord injury. They advise to make clear arrangements in stabilizing patients, transferring patients, diagnostics and treatment of patients between the hospitals in the specific trauma region. They also advise a more concentrated care of patients with spinal injuries, something the patient federation also agrees on. The United States started with centralizing acute health care for spinal cord injury patients, creating Acute Spinal Cord Injury Units. This Unit is closely attached to the Intensive Care Unit and provides multidisciplinary Health Care and has a minimum of 50 admitted patients a year with spinal cord injury.[4] Since there are around 200 patients a year with spinal cord injuries in the Netherlands, the patient federation advises 3-4 hospitals.[2,24]

Variation between trauma centers was found with regards to the composition of their advisory committees on traumatic spinal injury (Figure 1). The latest national guidelines indicate that patients with (poly-)trauma, arriving at a trauma center, should receive treatment by a team of medical doctors (with various backgrounds) under supervision of a trauma general surgeon where there should be a trauma protocol for patients with spinal injury.[17] The results of our survey demonstrate that this is currently not the case in some trauma centers. One can imagine that due to the low incidence of traumatic spinal injury, and its widespread complex clinical presentation of symptoms, a guideline for the composition of an advisory committee could be beneficial in each trauma center dealing with traumatic spinal injury to obtain a more thoroughly and multidisciplinary approach which also could improve registration of these patients in order to create prospective databases and perform high quality outcome analysis of treatment.

Our data suggests that there is a need for a new classification system regarding cervical spinal trauma with 5 trauma centers in our study not using a specific classification for these fractures. During the writing of this paper the AOspine subaxial cervical spine injury classification system was published following the revised thoracolumbar one. We expect that this newly designed AO Spine subaxial cervical classification system will be a valuable tool for communication, patient care, and research purposes. [22] In addition, we believe the new classification system to improve the communication and multidisciplinary approach of cervical traumatic spinal injury. Concerning thoracolumbar classification schemes there is more consensus, although still 4 different systems are used. There are pro's and con's for each system but patient care could benefit of 1 universally accepted classification system. For this the AO revised classification of traumatic thoracolumbar injuries could be used published in 2013, although it should still be evaluated after 1-2 year usage, as planned.[18]

Another variation is seen in the background of the surgeons involved in the treatment of traumatic spinal injury (Figure 2). In the Netherlands there is an ongoing discussion about the acknowledgement of spinal surgery operations and surgeons that perform these operations. The start of implementing the Dutch

Spine Surgery Registry one year ago gives more insight in performed spinal surgery and outcome in the Netherlands. Eventually, accreditation should be given to a spinal surgeon when performing an adequate number of surgeries a year. This could be of influence when care is concentrated to a few hospitals since expertise in spinal surgery is lost in this way.

Figure 3 shows notable diversity in timing with regards to surgical intervention of traumatic spinal injury with and without neurological deficit. To date, there is still no (inter)national consensus on when to operate traumatic spinal injury. Although various studies have been conducted on the topic of timing, a lack of sufficient evidence is reflected in the debate on timing in the recently updated AANS/CNS guidelines.[11] A recent review on the effects of timing in spinal surgery after traumatic SCI shows that "early" surgical intervention is associated with improved neurological and length of stay outcomes.[14] However, this study has a low level of evidence due to heterogeneity within and between studies. An observational multicenter cohort study compared "early" surgical intervention (< 24 hours) with "late" surgical intervention (> 24 hours) in acute spinal cord injury. This study found significant motor recovery improvement in incomplete acute spinal cord injury in the cervical, thoracic, or thoracolumbar spine, and shorter length of hospital stay.[5] In addition, another recent study suggested superior neurological recovery after traumatic cervical spinal cord injury if surgical intervention was performed within 8 hours after injury.[12] On the other hand, The STASCIS study revealed that patients with cervical SCI operated within 24h had a 2.83 times higher chance of improving 2 grades on the ASIA scale than patients operated later than 24h.[6] With these results no recommendations can be made with certainty in the case of timing of surgical intervention in traumatic spinal injury. This uncertainty is also reflected in our survey and suggests more clinical research on timing is necessary. An AO Spine sponsored study (SCI-POEM) is conducted on this issue at the moment, the final report to be delivered in 2017.

In the Netherlands, a majority of health care professionals (73%) involved in spinal trauma is in favor of more concentrated care for patients with spinal trauma (Table V). At the moment there are 11 national trauma centers where spinal trauma patients are eligible for treatment. As we have mentioned before, lack of agreement on when to operate imposes a prominent barrier for the implementation of a more concentrated level of spine care. Reimbursement is without doubt also a barrier on the path to implementing more concentrated care although this was not mentioned in this survey. However, more concentrated spinal care could lead to faster implementation of recent developments, guidelines and classifications, more possibilities for scientific research and higher quality of surgical experience. This may eventually result in better patient care and outcomes.

In conclusion, (inter)national collaboration in treating traumatic spinal injury is indispensable in order to achieve better communication, more spinal expertise, more research, and eventually good practical results. This survey has provided insight into the opinions of medical professionals involved in traumatic spinal inju-

ries in the Netherlands. It is of interest that in this relatively small country opinions on the treatment of spinal injury differ substantially. However, being relatively small and with very good logistic possibilities small countries, such as the Netherlands could be one of the countries to lead the way on research in timing of surgery in SCI patients. This survey can be of use to show the discrepancies between trauma regions and further motivate conducting good clinical research in this important field.

REFERENCES

1. **Abudou M, Chen X, Kong X, Wu T**. Surgical versus non-surgical treatment for thoracolumbar burst fractures without neurological deficit. *Cochrane database Syst Rev.* 2013 ; 6 : CD005079.

2. **van Asbeck FW, Post MW, Pangalila RF**. An epidemiological description of spinal cord injuries in The Netherlands in 1994. *Spinal Cord* 2000 ; 38 : 420-424.

3. **Auses O**. Spinal cord injury facts and figures at a glance. *J Spinal Cord Med.* 2014 ; 37 : 479-480.

4. **Consortium for Spinal Cord Medicine**. Early Acute Management in Adults with Spinal Cord Injury : *A Clinical Practice Guideline for Health-Care Professionals* 2008; 31 : 403-479.

5. **Dvorak MF, Noonan VK, Fallah N** *et al.* The Influence of Time from Injury to Surgery on Motor Recovery and Length of Hospital Stay in Acute Traumatic Spinal Cord Injury: An Observational Canadian Cohort Study. *J Neurotrauma* 2015 ; 32 : 645-654.

6. **Fehlings MG, Vaccaro A, Wilson JR** *et al.* Early versus Delayed Decompression for Traumatic Cervical Spinal Cord Injury: Results of the Surgical Timing in Acute Spinal Cord Injury Study (STASCIS). Di Giovanni S, ed. *PLoS One* 2012 ; 7 : e32037.

7. **Fisher CG, Noonan VK, Dvorak MF**. Changing face of spine trauma care in North America. *Spine* (Phila Pa 1976) 2006 ; 31 : 2-8.

8. **Ghobrial GM, Maulucci CM, Maltenfort M** *et al.* Operative and nonoperative adverse events in the management of traumatic fractures of the thoracolumbar spine: a systematic review. *Neurosurg Focus* 2014 ; 37 : E8.

9. **Glaser JA, Jaworski BA, Cuddy BG** *et al.* Variation in surgical opinion regarding management of selected cervical spine injuries. A preliminary study. *Spine* (Phila Pa 1976) 1998 ; 23 : 975-982.

10. **Gnanenthiran SR, Adie S, Harris I** *et al.* Nonoperative versus operative treatment for thoracolumbar burst fractures without neurologic deficit: a meta-analysis. *Clin Orthop Relat Res.* 2012 ; 470 : 567-577.

11. **Hadley MN, Walters BC**. Methodology of the Guidelines for the Management of Acute Cervical Spine and Spinal Cord Injuries. *Neurosurgery* 2013 ; 72 : 17-21.

12. **Jug M, Kejžar N, Vesel M** *et al.* Neurological Recovery after Traumatic Cervical Spinal Cord Injury Is Superior if Surgical Decompression and Instrumented Fusion Are Performed within 8 Hours versus 8 to 24 Hours after Injury: A Single Center Experience. *J Neurotrauma* 2015 ; 32 : 1385-1392

13. **Kepler CK, Vroome C, Goldfarb M** *et al.* Variation in the Management of Thoracolumbar Trauma and Postoperative Infection. *J Spinal Disord Tech.* 2015 ; 28 : E212-8.

14. van Middendorp JJ, Hosman AJF, Doi S a R. The effects of the timing of spinal surgery after traumatic spinal cord injury: a systematic review and meta-analysis. *J Neurotrauma* 2013 ; 30 : 1781-1794.

15. Moller A, Hasserius R, Redlund-Johnell I, Ohlin A, Karlsson MK. Nonoperatively treated burst fractures of the thoracic and lumbar spine in adults: a 23- to 41-year follow-up. *Spine J* 2015 ; 7 : 701-707.

16. Öner FC, Sadiqi S, Lehr AM *et al.* Towards the development of an outcome instrument for spinal trauma: an international survey of spinal surgeons. *Spine* (Phila Pa 1976) 2015 ; 40 : E91-6.

17. Nederlandse Orthopaedische Vereniging, Acute traumatische wervelletsels. https://www.dutchspinesociety.nl/images/stories/Publiek/Richtlijnen/richtlijn_acute_ traumatische_wervelletsels_2009.pdf Accessed January 20, 2016.

18. Reinhold M, Audigé L, Schnake KJ, Bellabarba C, Dai LY, Öner FC. AO spine injury classification system: A revision proposal for the thoracic and lumbar spine. *Eur Spine J* 2013 ; 22 : 2184-2201.

19. Schiller MD, Mobbs RJ. The historical evolution of the management of spinal cord injury. *J Clin Neurosci* 2012 ; 19 : 1348-1353.

20. Schwartz E, Flanders A. Spinal Trauma. *Lippincott Williams & Wilkins* ; 2007.

21. Tilford JM, Porter A. Hospitalizations and Medical Care Costs of Serious Traumatic Brain Injuries. *Spinal Cord Injuries and Traumatic Amputations* 2013.

22. Vaccaro AR, Koerner JD, Radcliff KE *et al.* AO Spine subaxial cervical spine injury classification system. *Eur Spine J.* 2016 ; 25 ; 2173-2184.

23. Wyndaele M, Wyndaele J-J. Incidence, prevalence and epidemiology of spinal cord injury: what learns a worldwide literature survey? *Spinal Cord* 2006 ; 44 : 523-529.

24. Zorgstandaard Dwarslaesie Organisatie Nederland. https://www.zorginzicht.nl/bibliotheek/Dwarslaesie/RegisterKwaliteitsstandaardenDocumenten/Zorgstandaard Dwarslaesie 2013 digitaal.pdf. Accessed on 21 January 2016.

3

Nonoperative Treatment of Thoracic and Lumbar Spine Fractures: A Prospective Randomized Study of Different Treatment Options

Agnita Stadhouder
Erik Buskens
Diederik A. Vergroesen
Malcolm W. Fidler
Frank de Nies
F. Cumhur Öner

J Orthop Trauma. 2009 Sep;23(8):588-94. DOI: 10.1097/BOT.0b013e3181a18728.

ABSTRACT

Objectives: To evaluate and compare nonoperative treatment methods for traumatic thoracic and lumbar compression fractures and burst fractures.
Design: Prospective randomized controlled trial with long-term follow-up.
Setting: Two general hospitals in the Netherlands.
Patients/Participants: Patients with a traumatic thoracic or lumbar spine fracture, without neurologic damage, with less than 50% loss of height of the anterior column and less than 30% reduction of the spinal canal were included.
Intervention: Patients in the compression group were randomized to physical therapy and postural instructions, a brace for 6 weeks or a Plaster of Paris cast for 6 or 12 weeks. Patients in the burst group received a brace or a Plaster of Paris cast, both for 12 weeks.
Main Outcome Measurements: Follow-up examinations included radiographs, Visual Analogue Scores for toleration of treatment and persistent pain and an Oswestry
Disability Index at long-term follow-up.
Results: There were 133 patients: 108 in the compression group and 25 in the burst group. For compression fractures, physical therapy and brace were considered the most tolerable. Brace therapy scored significantly better on the Visual Analogue Scores for residual pain and on the Oswestry Disability Index. None of the treatments had any significant effect on the residual deformity measurements. For burst fractures, no significant differences were found.
Conclusions: Brace treatment with supplementary physical therapy is the treatment of choice for patients with compression fractures of the thoracic and lumbar spine.
Furthermore, more than 20% of all patients had moderate or severe back pain at long-term follow-up.

INTRODUCTION

Nonoperative treatment for thoracic or lumbar anterior wedge compression type and stable burst spine fractures is considered to be safe with an acceptable long-term outcome concerning pain, employability, and residual deformity for the majority of patients.[1-6] Treatment options vary from bed rest, via the use of various orthoses, to functional treatment with postural instructions by physiotherapists.[4,7-10] However, there is no consensus in the literature about the optimal treatment. There is also a paucity of direct evidence of the effectiveness of any of the different treatment schemes, although Shen and Shen[5] and Mehta et al[11] referred to research done by Patwardhan et al[12] in which the stabilizing value of a Jewett hyperextension orthosis appeared to depend on the initial posttrauma segmental stiffness. They concluded from their own studies that it was not necessary to wear a brace as this provided no additional therapeutic benefit. Despite these reports and the tendency to treat these injuries usually with "benign neglect," every spine surgeon knows cases of dissatisfied patients with substantial residual pain after different kinds of nonoperative treatment schemes who occasionally require operative intervention.[13] To try and identify the optimal method of nonoperative treatment, we conducted a prospective randomized comparison of 4 treatment options for compression fractures (AO type A1 and A2) and 2 for burst fractures (AO type A3).[14] As far as we are aware, such a long-term study has not been previously reported.

PATIENTS AND METHODS

The study was carried out in 2 general hospitals in Amsterdam. Patients were enrolled from July 1991 until March 1997. Inclusion criteria were patients with a traumatic thoracic or lumbar fracture without neurologic impairment and younger than 80 years. Only fractures with less than 50% loss of anterior height, with less than 30% reduction of the spinal canal, and without signs of posterior element involvement were included.

There were 133 patients: 72 (54.1 %) women and 61 (45.9%) men. Patients were admitted to hospital after initial radiographs had been made. A computed tomography scan was performed within 48 hours of admission in all cases. The fractures were classified according to the AO classification, and the severity of trauma, high or low energy, was also assessed. Bed rest was prescribed for the first 3–5 days depending on pain and general condition. After written informed consent had been given, patients were randomized to one of the following treatments for compression fractures: (1) physical therapy alone for 6 weeks, (2) thermoplastic removable brace for 6 weeks, and (3) plaster of Paris (POP) cast for 6 or (4) 12 weeks. For burst fractures, thermoplastic removable brace was compared with POP cast, both for 12 weeks. All patients treated with orthoses also received physical therapy, and in the compression group, braces were allowed to be removed at night. Discharge followed after adequate mobilization.

Table 1 shows the demographic data of the patients after randomization into treatment groups.

Follow-up was planned at 6 and 12 weeks and 6 and 12 months with at least 1 long-term follow-up visit minimally 1 year later. Initially, the study focused also on radiological parameters: 5 measurements were made on the supine lateral radiographs, that is, the C1 (actual Cobb angle[15]) between the superior end plate of the vertebrae above and the inferior end plate of the vertebrae below the fractured level; the C2, which is the wedge angle of the affected vertebra; the C3 measuring the wedge angle of the fractured vertebra and adjacent intervertebral discs of the fractured vertebra; the C4, which is the ratio between the heights of theanterior and posterior parts of the vertebral body; and the C5 angle, which includes the fractured vertebra and the superior intervertebral disc (Fig. 1).

Radiologic deformity, residual pain, and functional outcome were set as primary outcome parameters. At follow-up, patients were also asked about any pretrauma back pain and disability. At the long-term follow-up visit, Visual Analogue Scores (VAS) were used to assess toleration of treatment (0 = easily tolerated, 100 = intolerable) and residual pain (0 = no pain, 100 = unbearable pain), and an Oswestry Disability Index (ODI) was calculated.[16]

	Compression Fractures					Burst Fractures		
	n	Physical Therapy	Brace for 6 Wk	POP for 6 Wk	POP for 12 Wk	n	Brace for 12 Wk	POP for 12 Wk
No. patients (%)	108	29 (27)	29 (27)	27 (25)	23 (21)	25	9 (36)	16 (64)
Gender–male, %	45	48	45	48	39	48	56	44
Mean age, yrs (range)	47 (18-76)	50 (21-70)	46 (19-69)	48 (18-75)	45 (18-76)	47 (21-73)	45 (21-64)	48 (26-73)
Male	43 (18-75)	47 (21-70)	39 (19-64)	46 (18-75)	38 (24-71)	44 (21-71)	36*(21-59)	49 (27-71)
Female	51 (20-76)	52 (24-70)	52 (26-69)	50 (22-68)	49 (20-76)	51 (26-73)	57 (46-64)	48 (26-73)
Mean admission time, d	8.8 (0-60)	9.7 (0-60)	8.8 (0-21)	8.3 (0-27)	8.9 (0-60)	12.5 (0-25)	12.4 (6-21)	12.6 (0-49)
High-energy trauma, %	82	86	86	85	78	80	86	75
*Statistically significant								

Table 1. Patient demographic data

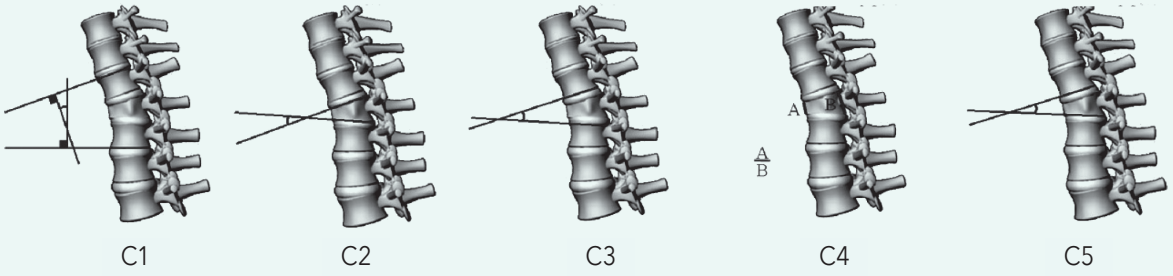

Fig. 1. –Different kyphosis measurements.

Because the majority of the fractures occurred at the thoracolumbar junction (T11-L2), we also evaluated these patients separately. In addition, analyses were repeated after exclusion of postmenopausal patients because this may be an independent parameter.

A power analysis beforehand was performed on the basis of a presumed difference in kyphosis angle of 5 degrees as significant difference (alpha 0.05, beta 0.20, SD 10), which required 22 patients per group. Statistical analyses were performed with SPSS 11.0 (SPSS Inc, Chicago, IL) to compare the different treatment schemes; compression and burst fractures were analyzed separately. Using an independent sample t test, mean differences in C measurements, VAS, and ODI between 2 treatments at a time were determined at follow-up inclusive of 95% confidence intervals (CI). Post hoc analyses were not conducted. A P value of less than 0.05 was considered significant. Power and sample size calculation was performed. In addition, possible prognostic factors for persistent back pain and disability were looked for with multivariate analysis.

RESULTS

In total, there were 133 patients: 108 compression fractures, 22 burst fractures, and 3 patients with both compression and burst fractures. Patients who had both compression and burst fractures were allocated to the burst fracture group, making a total of 25 patients in this group (Table 1).

Table 2 shows the number of fractures, subdivided according to the AO classification, in each treatment group. The ''split'' (A2.2) fractures were included with the compression fractures for treatment as, regarding their severity, they seemed more like these than like the burst fractures. One B1.2 fracture was included in the

	AO								
	A1.1	A1.2	A1.3	A2.2	A3.1	A3.2	A3.3	B1.2	Total
Physical therapy	4	30	1	1	–	–	–	–	36
Brace for 6 wk	2	28	2	–	–	–	–	–	32
POP for 6 wk	1	29	2	1	–	–	–	–	33
POP for 12 wk	1	28	2	–	–	–	–	–	31
Brace for 12 wk	–	–	–	–	4	6	–	–	10
POP for 12 wk for burst fracture	–	–	–	–	7	4	4	1	16
Total	8	115	7	2	11	10	4	1	158

Table 2. Treatment Randomization and Fracture Classification (AO), and Total Number of Fractures

burst fracture group because there was only minimal posterior disruption. Twenty patients had 2 compression fractures, 1 patient 2 burst fractures, 2 patients a compression and burst fracture and, 1 patient 2 compression fractures and 1 burst fracture. This gave a total of 158 thoracic and lumbar spine fractures: 132 were compression fractures (A1/ A2) and 26 were burst fractures (A3+B1).

The fracture level varied from Th3 to L5, 74% of the fractures were at the thoracolumbar junction (T11– L2), 15% exclusively thoracic, and 11% lumbar localized.

For compression fractures, there were no significant differences regarding sex, age, high-energy trauma, and admission time between the different treatment schemes. For patients with burst fractures treated with a brace, the mean ages of men (36 years) and women (57 years) showed a significant difference with a mean difference of 21.7 years and a CI of 1.8–41.5, as women were older than men (n = 9).

Twenty-seven women (38%) were postmenopausal with a mean of 10.4 years (range 0–30 years) between menopause and fracture. Twenty-one (78%) of these women did have a high-energy trauma, and they were all equally distributed among treatment groups ($\gamma 2$ P = 0.25).

Thirty-two patients (24%), when asked, reported pretrauma episodes of back pain, only 2 patients actually had elevated ODIs of 7 and 20 at the time of admission, the remainder did not. Although planned follow-up was at 6 and 12 weeks and 6 and 12 months and long-term follow-ups in 1998 and 2003, not all patients attended on all occasions. In 1998, a clinical and radiological long-term follow-up was carried out on 67.4% of the patients. At this time, 2 patients had died of unrelated causes and could not be included in the follow-up. In 2003, by means of telephone calls and questionnaires, follow-up was possible in 61%, corrected for the 14 patients who, by then, had died. Eleven patients who could not be traced in 1998 were contacted in 2003. Using a paired sample *t* test, we compared the VAS and ODI scores from 1998 with those from 2003 and concluded that there were no sig-

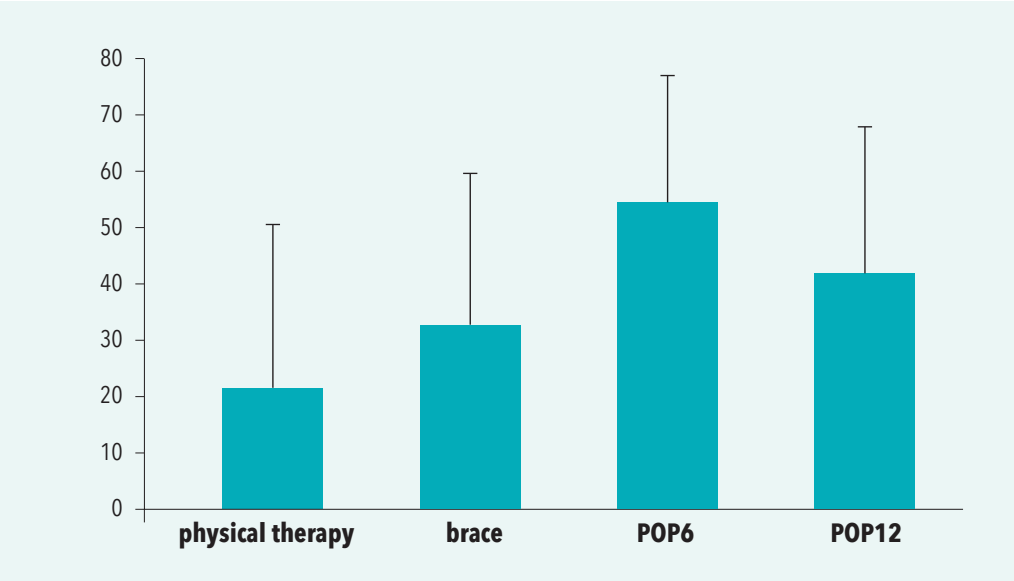

Fig. 2. Compression fractures, all patients: mean VAS toleration of treatment (0 = easily tolerated 100 = intolerable).

nificant differences. We therefore combined the scores from 1998 and 2003 for the VAS and ODI for long-term follow-up. This gave a long-term follow-up percentage for 1998/2003 of 75.4%. The mean follow-up period was 7.11 years with a range of 1–12 years (SD 3.0).

Radiologic Measurements

Table 3 shows the mean C measurements made on the lateral radiographs directly after the trauma, 1 year later, and at the first long-term follow-up in 1998. The mean measurements and individual treatment methods are presented. There were no significant differences between treatment groups at trauma, 1- year follow-up, and follow-up in 1998, also because of the large SDs. Within each treatment group, the kyphosis measurements at trauma and follow-up did not show any significant differences; in particular, there was no deterioration of the kyphosis angles.

VAS and ODI Scores

For the treatment of compression fractures, physical therapy was tolerated better than a POP for 6 and 12 weeks (mean difference 33.9, CI of 16.6–51.3, calculated power 0.97; mean difference 21.6, CI 3.4–39.8, calculated power 0.81). Brace therapy was tolerated better than a POP for 6 weeks with a mean difference of 21.6 less on the VAS scale (CI 5.8–37.4, calculated power 0.77) (Fig. 2).

For the VAS score for residual pain, a brace was significantly better than a POP for 12 weeks (mean difference 19.0, CI 1.87–36.2, calculated power 0.60) (Fig. 3).

The ODI showed a significant difference in favor of brace therapy compared with a POP for 12 weeks (mean difference 10.1, CI 0.25–20.0, calculated power 0.57)

	Com-pression Trauma	1 Yr	Last FU		Burst Trauma	1 Year	Last FU
C1 mean (SD)	7.7 (11.9)	9.3 (14.1)	7.2 (10.8)	C1 mean	11.8 (8.4)	8.3 (12.3)	11.8 (9.5)
Physiotherapy	6.8 (13.6)	7.9 (10.2)	3.8 (17.5)	Brace 12	12.6 (6.2)	9.5 (10.4)	12.3 (10.8)
Brace	6.4 (14.6)	12.0 (14.0)	7.9 (12.9)	POP 12	11.2 (10.0)	7.5 (13.8)	11.2 (9.1)
POP for 6 wk	9.4 (9.6)	10.6 (10.5)	8.7 (11.0)	–	–	–	_
POP for 12 wk	8.3 (8.9)	9.3 (21.2)	7.7 (11.4)	–	–	–	_
C2 mean	9.9 (5.2)	12.3 (7.3)	11.0 (6.0)	C2 mean	12.2 (7.1)	13.2 (6.9)	11.8 (6.1)
Physiotherapy	9.1 (6.0)	10.1 (12.4)	9.5 (8.1)	Brace 12	13.2 (4.2)	15.0 (6.6)	10.6 (7.9)
Brace	10.5 (4.5)	14.2 (5.8)	11.6 (5.1)	POP 12	11.4 (9.1)	12.0 (7.1)	13.3 (2.6)
POP for 6 wk	9.9 (5.6)	11.9 (5.2)	11.5 (5.2)	–	–	–	_
POP for 12 wk	10.1 (4.7)	12.0 (4.4)	10.4 (5.4)	–	–	–	_
C3 mean	3.1 (9.6)	4.0 (11.7)	2.0 (11.8)	C3 mean	5.9 (6.4)	4.3 (7.3)	4.0 (7.0)
Physiotherapy	2.4 (13.4)	1.5 (8.8)	-1.4 (16.1)	Brace 12	4.3 (7.3)	4.0 (9.5)	5.0 (7.2)
Brace	2.3 (7.2)	5.2 (8.1)	3.9 (8.6)	POP 12	7.4 (5.4)	4.5 (5.9)	2.8 (7.2)
POP for 6 wk	5.5 (8.0)	5.0 (10.9)	4.7 (10.0)	–	–	–	_
POP for 12 wk	2.7 (8.7)	6.8 (16.0)	0.8 (11.6)	–	–	–	_
C4 mean	0.8 (0.1)	0.7 (013)	0.8 (0.13)	C4 mean	0.73 (0.14)	0.67 (0.18)	0.69 (0.17)
Physiotherapy	0.78 (0.09)	0.72 (0.16)	0.77 (0.12)	Brace 12	0.71 (0.10)	0.62 (0.19)	0.71 (0.23)
Brace	0.77 (0.08)	0.68 (0.12)	0.71 (0.15)	POP 12	0.75 (0.17)	0.71 (0.16)	0.67 (0.09)
POP for 6 wk	0.78 (0.1)	0.74 (0.14)	0.74 (0.11)	–	–	–	_
POP for 12 wk	0.76 (0.15)	0.70 (0.17)	0.78 (0.12)	–	–	–	_
C5 mean	7.7 (7.3)	10.4 (9.6)	8.6 (9.5)	C5 mean	10.3 (7.2)	11.6 (8.3)	9.2 (6.7)
Physiotherapy	8.8 (9.7)	9.8 (7.0)	6.2 (13.3)	Brace 12	9.0 (5.5)	12.6 (7.9)	8.1 (8.6)
Brace	5.5 (5.3)	11.1 (6.9)	9.6 (7.1)	POP 12	11.5 (8.5)	10.9 (8.9)	10.3 (3.9)
POP for 6 wk	8.6 (5.9)	10.0 (9.1)	9.9 (8.0)	–	–	–	_
POP for 12 wk	7.6 (6.9)	10.9 (14.6)	8.2 (9.0)	–	–	–	_

FU, follow-up.

Table 3. Mean Measurements Compression and Burst Fractures (degrees)

and physical therapy (mean difference 14.9, CI 2.7–27.1, calculated power 0.70) (Fig. 4). These significant differences are summarized in Table 4.

When only patients with compression fractures of the thoracolumbar junction (n = 79) were analyzed, there were no significant differences regarding toleration of treatment. The VAS for residual pain was significantly lower after brace therapy compared with POP for 12 weeks (mean difference 28.1, CI 10.5–45.8) and POP for 6 weeks compared with POP for 12 weeks (mean difference 28.4, CI 9.6–47.3). The ODI was significantly lower after brace therapy than after physical therapy only

	Physical Therapy	Brace	POP for 6 Wk	POP for 12 Wk
Physical therapy	–	–	**VAS Treatment**	**VAS Treatment**
Brace	ODI	–	**VAS Treatment**	**VAS Treatment**
POP for 6 wk	–	–	–	–
POP for 12 wk	–	–	–	–

Table 4. Compression fractures, all patients; summary of significant differences between treatments: cross table

	Physical Therapy	Brace	POP for 6 Wk	POP for 12 Wk
Physical therapy	–	–	–	–
Brace	ODI	–	ODI	**VAS pain ODI**
POP for 6 wk	ODI	–	–	**VAS pain**
POP for 12 wk	–	–	–	–

Table 5. Patients with thoracolumbar junction compression fractures; summary of significant differences between treatments: cross table

	Physical Therapy	Brace	POP for 6 Wk	POP for 12 Wk
Physical therapy	–	–	**VAS treatment**	–
Brace	ODI	–	ODI	ODI
POP for 6 wk	–	–	–	–
POP for 12 wk	–	–	–	–

Table 6. All patients with compression fractures, except postmenopausal women; summary of significant differences between treatments: cross table

(mean difference 26.9, CI 11.4–42.3), POP for 6 weeks (mean difference 7.7, CI 0.35–15.0) and POP for 12 weeks (mean difference 14.4, CI 0.63–26.1). Also, there was a significant difference in favor of POP for 6 weeks compared with physical therapy (mean difference 19.2, CI 3.8–34.7) (Table 5).

After excluding postmenopausal women from the total compression fracture population (n = 81), physical therapy was tolerated better than a POP for 6 weeks (mean difference 28, CI 4.8–52.5). The VAS for residual pain did not show significant differences. The ODI was significantly lower for brace therapy compared with phys-

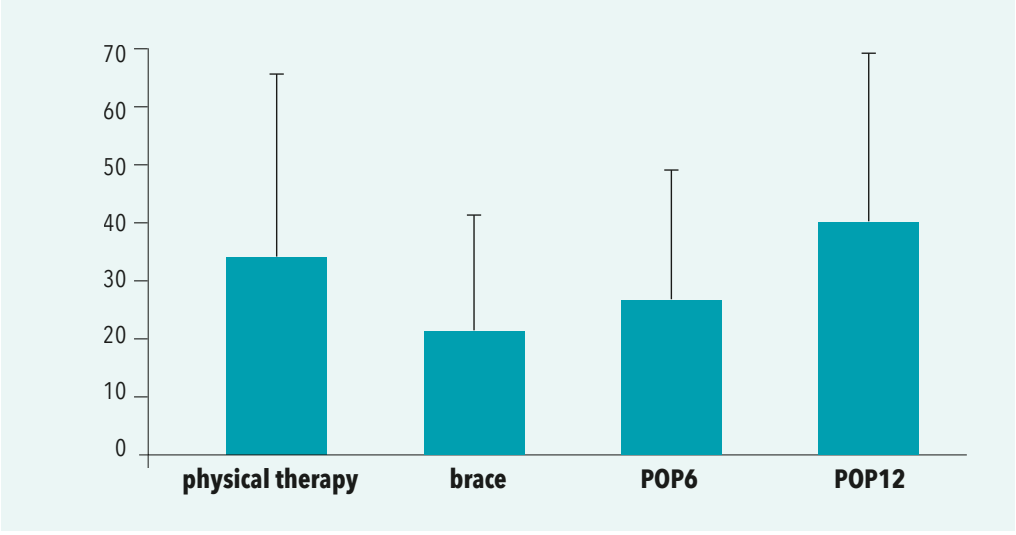

Fig. 3. Compression fractures, all patients: mean VAS residual pain.

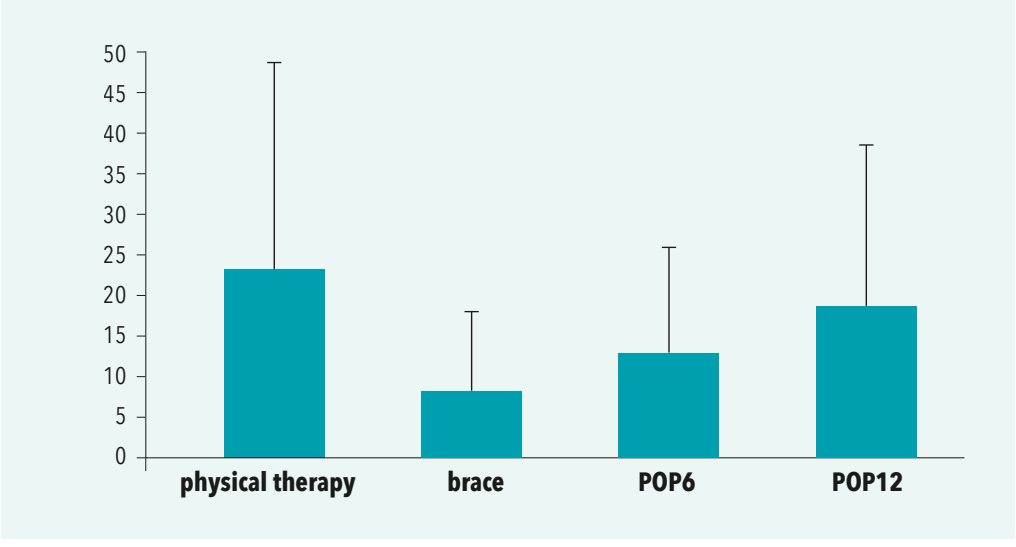

Fig. 4. Compression fractures, all patients: mean ODI.

iotherapy (mean difference 19.1, CI 3.3–35.0), a POP for 6 weeks (mean difference 8.8, CI 0.65–17) and a POP for 12 weeks (mean difference 13.9, CI 2.8–24.9) (Table 6).

For burst fractures, the VAS and ODI scores were both worse than those for the compression group and did not show any significant differences between treatments.

In the compression fracture group, 20 patients (18%) had a VAS score for persistent pain of greater than 50 (moderate pain), and 10 (9%) of these had a VAS score of >70, which implies severe pain. Twelve patients (11%) in the compression fracture group had ODI scores of >40; 8 of these with an ODI of 60–80 and 1 with an ODI of 80–100. A multivariate analysis did not show any significant relationship with the type of treatment, fracture classification, or C measurements on the lateral radiographs. Prognostic factors of poor outcome could therefore not be identified.

In the burst fractures group, no patient had a VAS pain score higher than 70; 3 patients (12%) had a VAS score >50. One patient (4%) in the burst fracture group had an ODI of 76%. Multivariate analysis did not reveal significant prognostic parameters.

One patient in the burst group treated with a brace was operated on because of progressive deformity and pain 2 years after the traumatic event.

DISCUSSION

The first part of this study considered possible alterations in the measurements of the traumatic kyphosis after various treatments. Kuklo et al[17] and Dai and Jin[18] showed that the intra- and interobserver reliability of measurements on lateral radiographs and computed tomography scans vary but the C1 measurement as used in our study and the McCormack[19] classification are the most accurate. The Spine Study Trauma Group also included the Cobb angle (our C1 measurement) and the anterior vertebral compression percentage (our C4 measurement), the vertebral body translation percentage, and the sagittal to transverse canal diameter ratio in their list of recommended measurements for assessing thoracolumbar fractures.[15] If we had restricted our measurements to these recommended ones, our results would not have been different.

Our observation that nonoperative treatment, using the methods described, does not significantly improve or, and more importantly, lead to deterioration in the final kyphosis angle is in agreement with the findings of Ohana et al[20] and Folman and Gepstein,[21] who treated patients with compression fractures functionally or with a brace. Alanay et al,[1] Agus et al,[7] and Wood et al[22] who investigated burst fractures treated nonoperatively, came to the same conclusions as did Tropiano et al[10] where burst fractures were reduced before application of the cast.

The VAS and ODI scores were more revealing. For patients with compression fractures, the scores of all the patients, of just those with thoracolumbar fractures, and of all patients after exclusion of the postmenopausal women indicated that the best of our treatment options is a brace for 6 weeks; for burst fractures, there was

no difference between a brace or a POP. We did not separately analyze the results for the 2 groups of patients with fractures above T11 or below L2, as the numbers would have been too small.

These results are in contrast with Ohana et al[20] and Folman and Gepstein[21] who concluded that they did not see any difference in outcome between patients treated with a brace or functionally with physical therapy. Braun et al[8] also did not see a difference in outcome of patients treated functionally or with a 3-point brace. The difference in results may be explained by the retrospective nonrandomized nature of their studies compared with ours.

We wondered why for patients with compression fractures, brace treatment was better than a POP? Perhaps a removable brace provided the optimal combination of support, exercise, and comfort; in other words, the brace gave the patient sufficient spinal support, reduction of discomfort, and confidence to encourage exercise during the day while removal of the brace at night facilitated sleep and a feeling of general well-being.

The scores also showed disturbing features. According to the VAS, 20 (18%) of the 108 patients with compression fractures suffered from moderate or severe back pain at long-term follow-up; 12 patients had an ODI score greater than 40 indicating moderate disability. Of the 25 patients with burst fracture, 3 (12%) had chronic moderate pain and one more was operated on because of severe persistent pain. Such chronic pain after the nonoperative treatment of thoracolumbar fractures has also been observed by other authors where treatments have varied from several weeks bed rest, different braces, and physiotherapy to supervised neglect.[2–6,9,21,23]

The multivariate analysis of our results unfortunately did not reveal any prognostic factors for persistent pain and disability. In particular, there was no association between final kyphosis measurements and residual pain, a fact that has also been noted by various other authors,[6,23–25] although Gertzbein[26] observed that a kyphotic deformity of greater than 30 degrees at 2-year follow-up was associated with an increased incidence of significant back pain. Folman and Gepstein studied 85 patients with a thoracolumbar vertebral wedge fracture treated with either physical therapy or a 3-point brace and found that 69.4% of the patients complained of chronic back pain, although there was no difference between the 2 nonoperative treatments.[21] The mean ODI for this patient group was 56.3, which is considerable higher than that of our population. He found a weak correlation between pain intensity and local kyphosis angle.

We should consider seriously the relatively high incidence of persistent pain, disability, and dissatisfaction after these relatively ''minor''spinal injuries.[27] This incidence is much higher than seen after comparable injuries to the extremities. Almost 20% of patients suffering moderate to severe pain after a minor injury of the ankle, knee, or wrist would not be accepted as ''good results.'' We should ask ourselves how we can predict these unsatisfactory results and whether we can prevent disability.

This study did have some drawbacks. First, the recruitment of patients was slow, 133 over almost 6 years, and this lead to a relatively small number of patients in each group, especially in the burst fracture group. However, these numbers were sufficient to show that there were no treatment-related statistically significant differences between the kyphosis measurements at the long-term follow-up and also to show that there were significant differences for the VAS and the ODI scores in the compression group. Second, patient compliance was not optimal, although the patients were well informed. The combined attendance at the long-term follow-up was 75%. We feel, however, that this has not resulted in a systematic bias because the random absentees applied equally to all groups. Third, patients' toleration of treatment, persistent pain, and disability were only recorded at the long-term follow-ups and not at the earlier controls as well, when they might have provided insight into how quickly patients could function independently after various treatments of a spinal fracture. Fourth, the percentage of postmenopausal women is relatively high and osteoporosis may have influenced the results. However, almost 80% of them had a high-energy trauma, and none of them had spontaneous back pain at inclusion; this probably excludes any true "spontaneous" osteoporotic fractures. Separate analyses also showed that the ODI was significantly better for brace therapy after exclusion of postmenopausal women.

We included patients with a thoracic or lumbar fracture, of whom 74% had a fracture of the thoracolumbar junction. The numbers of exclusively thoracic or lumbar fractures were too small to split our patient population in 3 groups. However, the number of patients with a fracture of the thoracolumbar junction was sufficient for separate analysis; brace therapy significantly had the best outcome on the ODI compared with the other treatment modalities.

Despite the fact that our study shows some methodologic flaws, it is one of the few studies that compares nonoperative treatment schemes based on a reasonable number of patients. A prospective, probably multicenter, study with inclusion of a sufficient number of patients would seem appropriate to search for the possible factors predicting poor outcome.

CONCLUSIONS

None of our nonoperative treatments had an effect on the post-traumatic kyphosis measurements. After a compression fracture, physical therapy alone is the most easily tolerated treatment. Brace treatment, however, results in the least residual pain and the least disability on the long term. Despite the fact that our study has some drawbacks, we tentatively recommend brace treatment as the treatment of choice for patients with moderate compression fractures of the thoracic and lumbar spine. For burst fractures, neither treatment had a clear advantage.

REFERENCES

1. **Alanay A, Yazici M, Acaroglu E,** *et al.* Course of nonsurgical management of burst fractures with intact posterior ligamentous complex: an MRIstudy. *Spine.* 2004;29:2425–2431.

2. **Cantor JB, Lebwohl NH, Garvey T,** *et al.* Nonoperative management of stable thoracolumbar burst fractures with early ambulation and bracing. *Spine.* 1993;18:971–976.

3. **Chan DP, Seng NK, Kaan KT**. Nonoperative treatment in burst fractures of the lumbar spine (L2-L5) without neurologic deficits. *Spine.* 1993;18: 320–325.

4. **Mumford J, Weinstein JN, Spratt KF,** *et al.* Thoracolumbar burst fractures. The clinical efficacy and outcome of nonoperative management. *Spine.* 1993;18:955–970.

5. **Shen WJ, Shen YS**. Nonsurgical treatment of three-column thoracolumbar junction burst fractures without neurologic deficit. *Spine.* 1999;24: 412–415.

6. **Weinstein JN, Collalto P, Lehmann TR**. Long-term follow-up of nonoperatively treated thoracolumbar spine fractures. *J Orthop Trauma.* 1987;1:152–159.

7. **Agus H, Kayali C, Arslantas M**. Nonoperative treatment of burst-type thoracolumbar vertebra fractures: clinical and radiological results of 29 patients. *Eur Spine J.* 2005;14:536–540.

8. **Braun W, Markmiller M, Ruter A**. Conservative therapy of fractures of the thoracic and lumbar spine. Indications, treatment regimen, results. *Chirurg.* 1991;62:404–408.

9. **Karjalainen M, Aho AJ, Katevuo K**. Painful spine after stable fractures of the thoracic and lumbar spine. What benefit from the use of extension brace? *Ann Chir Gynaecol.* 1991;80:45–48.

10. **Tropiano P, Huang RC, Louis CA,** *et al.* Functional and radiographic outcome of thoracolumbar and lumbar burst fractures managed by closed orthopaedic reduction and casting. *Spine.* 2003;28:2459–2465.

11. **Mehta JS, Reed MR, McVie JL,** *et al.* Weight-bearing radiographs in thoracolumbar fractures: do they influence management? *Spine.* 2004;29: 564–567.

12. **Patwardhan AG, Li SP, Gavin T,** *et al.* Orthotic stabilization of thoracolumbar injuries. A biomechanical analysis of the Jewett hyperextension orthosis. *Spine. 1990;15:654–661.*

13. **Been HD, Poolman RW, Ubags LH**. Clinical outcome and radiographic results after surgical treatment of post-traumatic thoracolumbar kyphosis following simple type A fractures. *Eur Spine J.* 2004;13:101–107.

14. **Magerl F, Aebi M, Gertzbein SD,** *et al.* A comprehensive classification of thoracic and lumbar injuries. *Eur Spine J.* 1994;3:184–201.

15. Keynan O, Fisher CG, Vaccaro A, *et al.* Radiographic measurement parameters in thoracolumbar fractures: a systematic review and consensus statement of the spine trauma study group. *Spine.* 2006;31:E156–E165.

16. Fairbank JC, Couper J, Davies JB, *et al.* The Oswestry low back pain disability questionnaire. *Physiotherapy.* 1980;66:271–273.

17. Kuklo TR, Polly DW, Owens BD, *et al.* Measurement of thoracic and lumbar fracture kyphosis: evaluation of intraobserver, interobserver, and technique variability. *Spine.* 2001;26:61–65.

18. Dai LY, Jin WJ. Interobserver and intraobserver reliability in the load sharing classification of the assessment of thoracolumbar burst fractures. *Spine.* 2005;30:354–358.

19. McCormack T, Karaikovic E, Gaines RW. The load sharing classification of spine fractures. *Spine.* 1994;19:1741–1744.

20. Ohana N, Sheinis D, Rath E, *et al.* Is there a need for lumbar orthosis in mild compression fractures of the thoracolumbar spine?: a retrospective study comparing the radiographic results between early ambulation with and without lumbar orthosis. *J Spinal Disord.* 2000;13:305–308.

21. Folman Y, Gepstein R. Late outcome of nonoperative management of thoracolumbar vertebral wedge fractures. *J Orthop Trauma.* 2003;17: 190–192.

22. Wood K, Buttermann G, Mehbod A, *et al.* Operative compared with nonoperative treatment of a thoracolumbar burst fracture without neurological deficit. A prospective, randomized study. *J Bone Joint Surg Am.* 2003;85-A:773–781.

23. Hazel WA Jr, Jones RA, Morrey BF, *et al.* Vertebral fractures without neurological deficit. A long-term follow-up study. *J Bone Joint Surg Am.* 1988;70:1319–1321.

24. Kraemer WJ, Schemitsch EH, Lever J, *et al.* Functional outcome of thoracolumbar burst fractures without neurological deficit. *J Orthop Trauma.* 1996;10:541–544.

25. Shen WJ, Liu TJ, Shen YS. Nonoperative treatment versus posterior fixation for thoracolumbar junction burst fractures without neurologic deficit. *Spine.* 2001;26:1038–1045.

26. Gertzbein SD. Scoliosis Research Society. Multicenter spine fracture study. *Spine.* 1992;17:528–540.

27. Lombao D, Pellise F, Bago J, *et al.* Quality of life in patients who did not undergo surgery for fracture of the thoracolumbar junction. *Eur Spine J.* 2006;15S:467–468.

4

Are existing outcome instruments suitable for assessment of spinal trauma patients? A review

Agnita Stadhouder
Constantinus F. M. Buckens
Herman R. Holtslag
F. Cumhur Öner

J neurosurgery Spine, Volume 13, Number 5, November 2010. 638-647.

ABSTRACT

Object. Valid outcome assessment tools specific for spinal trauma patients are necessary to establish the efficacy of different treatment options. So far, no validated specific outcome measures are available for this patient population. The purpose of this study was to assess the current state of outcome measurement in spinal trauma patients and to address the question of whether this group is adequately served by current disease-specific and generic health-related-quality-of-life instruments.

Methods. A number of widely used outcome measures deemed most appropriate were reviewed, and their applicability to spinal trauma outcome discussed. An overview of recent movements in the theoretical foundations of outcome assessment, as it pertains to spinal trauma patients has been attempted, along with a discussion of domains important for spinal trauma.

Commonly used outcome measures that are recommended for use in trauma patients were reviewed from the perspective of spinal trauma. The authors further sought to select a number of spine trauma-relevant domains from the WHO's comprehensive International Classification of Functioning, Disability and Health (ICF) as a benchmarkfor assessing the content coverage of the commonly used outcome measurements reviewed.

Results. The study showed that there are no psychometrically validated outcome measurements for the spinal trauma population and there are no commonly used outcome measures that provide adequate content coverage for spinal trauma domains.

Conclusions. Spinal trauma patients are currently followed either as a subset of the polytrauma population in theacute and early postacute setting or as a subset of neurological injury in the long-term revalidation medicine setting.

INTRODUCTION

As greater numbers of individuals survive serious trauma and their life expectancies increase substantially, the measurement of the quality of care and the health-related QOL of this group becomes more important. Additionally, the overall strain on medical services is making evidence-based justification of different treatment strategies imperative. Among survivors of major trauma, those with spinal trauma comprise a significant fraction, both in numbers and in the amount of care they require. In a recent consecutive series in the Netherlands, spinal injuries occurred in 24% of all high-energy trauma survivors and 6% had concurrent SCI.[31]

Issues specific to spinal trauma patients may not be adequately measured by generic outcome measures, or by "spine-specific" outcome measures that were designed for common chronic spinal conditions. Little work has been done on developing and validating outcome assessments in spinal trauma patients. Most studies have either assessed spine injuries from the critical care perspective or focus solely on SCI from the standpoint of rehabilitation medicine. Researchers who do investigate outcomes in (spinal) trauma patients typically have to resort to a combination of several outcome measures, or improvise their own measures.[10,60]

The SCI patient population, which usually includes only patients with serious permanent injury to the spinal cord and not those with transient or less severe neurological involvement, represents a minority of all spinal trauma patients.[1,31,32] This fact complicates the translation of outcome research performed in these populations to a general spinal trauma population. On the other hand, work in (poly)trauma populations typically contains a subset of spinal trauma patients, leading to similar obstacles when interpreting outcome data and evaluating outcome measures from the spinal trauma perspective.

Furthermore, spinal trauma patients are in a fundamental way dissimilar to patients with nontraumatic chronic back conditions, the population toward which many of the existing "spine" outcome measures have been directed.[44]

As there is no consensus on the treatment strategies for many types of spinal injuries, prospective studies are necessary to compare these treatment options. However, validated outcome assessment tools specific to the characteristics of the spinal trauma population are necessary to be able to establish the efficacy of interventions and rationalize management decisions.

For the purposes of this review, we used the WHO's International Classification of Functioning, Disability and Health (ICF) as an expansive theoretical underpinning for newly developed measures targeting, among others, trauma patients.[64] This describes, in detail, various health-related QOL domains, which can be used when comparing instruments and assessing their validity, and thus facilitates a more meaningful analysis.

The aim of this review is to evaluate the current state of outcome measurement as applied to spinal trauma patients. To be able to better assess the applicability and validity of existing outcome measures to spinal trauma patients, an overview of recent movements in the theoretical foundations of outcome assessment, as it

pertains to spinal trauma patients, has been attempted, along with a discussion of domains important for spinal trauma. This permits an evaluation of the suitability of existing outcome measures to spinal trauma patients. A selected number of widely used outcome measures deemed most appropriate are reviewed and their applicability to spinal trauma outcome is discussed.

Abbreviations used in this paper: AAOS = American Association of Orthopaedic Surgeons; ADL = activities of daily living; BDI =Beck Depression Inventory; FCI = Functional Capacity Index; FIM= Functional Independence Measure; GOS = Glasgow Outcome Scale; HADS = Hospital Anxiety and Depression Scale; HORS = Hamilton Depression Rating Scale; HSU = health service use; HUI2 = Health Utilities Index Mark 2; HUB = HUI Mark 3; ICF = International Classification of Functioning, Disability and Health; ISS = Injury Severity Score; LBOS = Low-Back Outcome Score; MFA = Musculoskeletal Function Assessment; MVAS = Million Visual Analog Scale; NASS = North American Spine Society; ODI = Oswestry Disability Index; QBPDS = Quebec Back Pain Disability Scale; QOL = quality of life; RMDQ = Roland-Morris Disabiliity Questionnaire; RTW = return to work; SCI = spinal cord injury; SCIM = Spinal Cord Independence Measure; SF-36 = 36-Itern Short Form Health Survey; SIP = Sickness Impact Profile; VAS = visual analog scale; WISCI = Walking Index for Spinal Cord Injury.

METHODS

A literature search was performed on PubMed and Embase using the Medical Subject Headings (MeSH) terms "Outcome Assessment (Health Care)" and "Spine" and "Spinal Fractures" or "Spinal Injuries" or "Spinal Cord Injuries," with limits "Humans," "Clinical Trial," "Meta-Analysis," "Randomized Controlled Trial," "Review," "English," "French," and "German;" 6090 papers were retrieved. The abstracts were reviewed by 2 authors and if deemed relevant the full-length article was sought. The full-length articles found were then thoroughly evaluated for relevant information on the outcome measurements used. The references of these papers were also manually screened for other potentially relevant articles, as were the related articles lists as generated by PubMed. Articles discussing outcomes in populations including spinal trauma components were included for analysis, with particular attention to those including psychometric data on outcome instruments in these populations. We also sought publications investigating and discussing the implementation of the ICF to spinal trauma. For those measures in which no psychometric/outcomes studies were found pertaining directly to spinal trauma patients as a unified population, the best available evidence was included.

RESULTS

Analysis of the literature on spinal injury outcome measurements has led to several important conclusions. Firstly, there is no unanimity regarding which instruments should be used when measuring nonmortality trauma outcomes in general, and spinal trauma outcomes in particular. This fact is evidenced by the plethora of instruments in use.[23] Therefore we categorized the different outcome measurements into: QOL physical measures, QOL mental/psychological measures, spinal disability measures, functional measures, and SCI measures. We also chose to draw an additional distinction between specific measurements primarily designed and used for trauma patients and those for SCI patients, due to the significantly different nature and content of these 2 groups of measures. Also, information on validation in spinal trauma was investigated.

Described in Table 1 are 4 QOL physical measures, 5 QOL mental/psychological measures, 8 disability measures, 4 functional measures, and 2 specific SCI questionnaires. These were the most commonly used measures in the literature that we thought relevant and possibly useful in trauma/spinal trauma patients.

Commonly used outcome measures that are primarily used in trauma populations in general and which we identified include the Functional Independence Measure (FIM), the GOS, the SF-36, the EQ-5D (standardized in strument of the EuroQol Group), the Musculoskeletal Function Assessment (MFA), and the Health Utilities Index Mark 3 (HUI 3). The only injury-specific outcome measure we identified is the Functional Capacity Index (FCI).

We found that the SCI patient population, whether the injuries are of traumatic origin or not, is largely treated as a separate and distinct population in the literature. Similar to the trauma outcomes field, there is no consensus about which outcome assessments are to be used in SCI patients. Outcome measures that are frequently used in SCI populations include: the Walking Index for Spinal Cord Injury (WISCI), the Spinal Cord Independence Measure (SCIM), the FIM, the SF-36, and the GOS.

Psychological outcome and well-being measurement is also an important aspect of polytrauma and spinal trauma outcome. The psychological outcome assessments commonly used in both trauma and SCI populations, such as the Hospital Anxiety and Depression Scale (HADS), the Beck Depression Inventory (BDI), and the Hamilton Depression Rating Scale (HDRS), are also discussed here. Components of generic health-related QOL tools (such as the SF-36) also measure psychological well-being. These instruments have been shown to be psychometrically sound, at least in SCI patients.

Quality-of-Life Physical Measures

The SF-36 is a widely applied generic measure that has been validated in numerous patient populations and consists of 36 items in 8 health domains.[28,45,61] It is not designed for measuring disability and has limitations such as inappropriate items for SCI patients.[46] Additionally, the psychometrics of the SF-36 have not been extensively tested in SCI populations.[25,36]

The EQ-5D is a self-administered generic health-related QOL measure with 5 dimensions (mobility, self-care, activities, pain, anxiety/depression) and a general health state VAS item. It was developed by the interdisciplinary EuroQol Group.[16] Although initially intended to complement condition-specific measures, it is being used increasingly as a stand-alone measure in spine research.[45] The EQ-5D has the added benefit of being able to generate utilities allowing economic evaluations. The HUI is a "generic multi-attribute preference-based measure of health status and health-related quality of life."[20,22] It encompasses both physical and emotional dimensions of health and emphasizes functional potential over performance in order to avoid the assumption that patients always choose to realize their full functional potential and to directly measure impairment. The HUI also allows for economic evaluation. The applicability of the HUI in trauma populations has yet to be tested.[60]

The Sickness Impact Profile (SIP) is a 136-item patient-oriented measurement. It relates to 12 areas of activity with statements that patients are asked to endorse, or patients are asked to check only those statements that they are sure describes their health on that day. This instrument was developed to detect changes or differences in heath status over time and between groups. It is useful for evaluation, program planning, and policy formulation.[6] It is also used in several spine studies on cervical disc herniation, spinal stenosis and back pain, vertebral deformities and osteoporosis, and in evaluating iliac crest donor problems.[45]

Other generic outcome instruments such as the WHO Disability Assessment Schedule and the GOS are not discussed because they are similar and used less often than the generic instruments discussed above, or are not directly applicable to spinal trauma.

Quality-of-Life Mental/Psychological Measures

The SF-36 and HUI both incorporate psychological features. Neither is validated for spinal trauma patients.

The HDRS was developed in 1960 and is one of the questionnaires most used in depression research.[26] It contains 17 variables measured on a 5- or 3-point scale. Important aspects for spinal trauma patients are depressed mood, work and interests, somatic, and genital.

The BDI has 21 items: 15 items evaluating emotional status and behavioral changes, and 6 evaluating somatic symptoms. It has high internal consistency, high content validity, and validity in differentiating between depressed and non-depressed patients, and it is sensitive to change.[3,12] In a study of the prevalence and severity of depression in 161 orthopedic trauma patients,[12] 55% were classified as having minimal depression, 28% moderate, 13% moderate-to-severe, and 4% severe when the somatic elements of the scale were included. Without the somatic elements, the proportion of patients classified as having moderate, moderate-to-severe, or severe depression was 26%.

Measure	Description	Validated for Spinal Trauma	Relevant Dimensions for Spinal Trauma
QOL physical measures			
SIP[5]	pt-oriented 136-item measure of gen health; items: ambulation, mobility, body care & movement, soc interaction, alertness behavior, emot'l behavior, communication, sleep & rest, eating, work, home mgmt, recreation; 2 subscores: phys & psychosoc dysfx range: 0 (perfect health) to 100 (severe disab); 0–3 = little/no disab; 4–9 = mild disab; 10–19 = mod disab; ≥20 = severe disab	no; validated for SCI pts & trauma pts	ambulation, mobility, body care & movement, soc interaction, sleep & rest, eating, work, home mgmt, recreation
SF-36[61]	multipurpose, short-form health survey w/ 36 Qs; generic measurement tool w/ 8 difft domains: phys functioning, role limitations phys, bodily pain, soc functioning, gen mental health, role limitations emot'l, vital- ity, gen health; range 0–100 (perfect health)	no; studied in musculoskel- etal trauma, SCI, head injury, polytrauma	phys functioning, bodily pain, soc functioning, mental health, gen health
EQ-5D[9]	standardized generic non–disease-specific instrument describing & eval- uating health-related QOL; includes 5 dimensions of health w/ a utility index score & a VAS for current health status; range (utility index): 0 (death) to 1 (perfect health)	no	mobility, self-care, usual activi- ties, pain, anxiety
HUI2 & 3[22]	HUI2: 15-item questionnaire about day-to-day health; measures 7 attri- butes of health status (sensation, mobility, emotion, cognition, self- care, pain, fertility); HUI3: 7 items (hearing, speech, ambulation, dexterity, emotion, cognition, pain); range: –0.36 (worse than dead) to 1.00 (perfect health)	no	mobility/ambulation, emotion, self-care, dexterity, pain
QOL mental/psychological measures			
HDRS[26]	17 variables to systematically quantify & use expert clin judgment on severity of illness of pts w/ depr; range: 0 (not serious) to 52 (serious)	no; used in SCI pts	
BDI[3]	21 items: 15 evaluating emot'l status & behavioral changes, & 6 evaluat- ing somatic Sx; each item is scored on a 4-point scale; scores are add- ed; range: 0–63; score <10 min depr, 10–18 mod depr, 19–29 mod- severe depr, >30 severe depr	no; used in pts w/ LBP & orthopedic pts	
HADS[65]	measure of mood, emot'l distress, anxiety, depr, & emot'l disorder in clin populations w/ Sx of clin disease; 14 items answered on a 4-point ver- bal rating scale; anxiety (7 items), depr (7 items), emot'l distress (all 14 items); range: 0–24, if >8 in both subscales indication of clinically relevant anxiety or depr is there	no; validated in pts w/ LBP	
SF-36	see above	no	mental health, role emot'l
HUI	see above	no	

disability indices			
ODI18	self-administered tool consisting of sections assessing ADL in 10 difft categories: pain intensity, personal care, lifting, walking, sitting, stand- ing, sleeping, sex life, soc life, travel- ing; range: 0 (no disab) to 100 (compl disab)	no; validated for pts w/ LBP	all categories
MVAS42	15 subjective variables reflecting severity of back pain, circumstances ex- acerbating Sx, & impact of problem on lifestyle, recorded as VAS scores; range: 0–10 for each item	no	all categories
RMDQ47	health status measure designed to be completed by pts to assess phys disab due to back pain; 24 items relate to phys fx affected by back pain; pts are asked to place a check mark beside statements that apply to them that day; range: 0 (no disab) to 24 (max disab)	no; validated for pts w/ LBP	all categories
LBOS24	developed as a quick, practical outcome score in pts w/ lumbar spine disorders measuring pain & disab; range: 0 (very disabled) to 75 (not at all disabled); 4 outcome catego- ries: ≥65 excellent, ≥50 good, ≥30 fair, <30 poor	no; validated for pts w/ LBP	all categories
QBPDS35	20-item scale of phys disab associated w/ pain; refers to simple activities (sleeping & resting, sitting, standing, ambulation, movement, bending & stooping, handling large or heavy objects); 6-point difficulty scale (0 "not difficult at all" to 5 "unable to do"); range: 0 (no disab) to 100 (compl disab)	no; validated for pts w/ LBP	all categories
NASS-LS13	adaptation of ODI & RMDQ to measure diverse dimensions of impact of lumbar spine problems; 34 items on demo- graphics, medical Hx (14 Qs), body fx (pain, neurogenic Sx; 16 Qs), employment Hx, outcomes of Tx	no	pain, neurogenic Sx
RADL62	measurement of extent to which a person w/ back pain has resumed his or her usual activities; questionnaire focuses on workers w/ LBP due to "soft tissue injuries;" 12 Qs on resumption of sleeping patterns, sexual activity, self-care, household chores, shopping, socializing, traveling, recre- ational activities, & employment are rated from 0% (not at all) to 100% (compl resumption)	no	all categories
LSOQ4	multi-item (56 Qs), self-report questionnaire designed to assess a num- ber of factors in pts w/ LBP: demographics, pain severity, fx'l disab, psych distress, phys Sx, health care utilization, satisfaction w/ Tx; range: 0–100 for pain severity, fx'l disab, psych distress, phys Sx	no; validated for pts w/ LBP	pain severity, fx'l disab, psych dis- tress, phys Sx, health care utilization, satisfaction
functional measures			
RTW39	assessing pts' RTW postinjury: in what time span & often characterizing type of work (blue collar/white collar); subdi- vided into fully employed, fully employed but less-demand- ing occupation, unable to perform full-time employment, not able to work or unemployed despite normal phys fx	no	RTW after spinal trauma
FIM59	data set to assess fx'l independence; divided into FIM motor (self care, sphincter control, transfers, locomotion) & FIM cognitive (communi- cation, soc cognition); further subdivided into 18 items; range: 18 (compl dependence) to 126 (compl independence)	no; studied in mul- tiple limb trauma, head injury, gen trauma, spinal injury	sphincter control: bladder mgmt/ bowel mgmt; trans- fer: bed/ wheelchair; locomotion: walk/ wheelchair/stairs

FCI37	preference-based, multiattribute fx'l outcome measure that provides 10 dimension-specific scores & 1 overall score that summarizes fx across 10 dimensions (eating, excretory fx, sexual fx, ambulation, hand/arm movement, bending/ lifting, vision, hearing, speech, cognitive fx); range 0 (death) to 1 (no limitations)	no; validated for blunt trau- ma pts, lower-extremity Injuries	excretory fx, sexual fx, ambula- tion, hand/arm move- ment, bending/ lifting
HSU50	measuring health service use of pts, mostly comparing post-trauma pts w/ healthy population (hospitalization days, placement in extended care services, home care service, physician claims, etc.	no	health service use after spinal Trauma
SCI measures			
SCIM11	disab scale for pts w/ spinal cord lesions; principal areas of fx: self-care (feeding, bathing, dressing, grooming), respiration & sphincter mgmt (respiration, bladder mgmt, bowel mgmt, use of toilet), mobility (mobil- ity in bed & action to prevent pressure sores, transfers bed/wheelchair/ toilet/tub, mobility for short/moderate/long distances, stair mgmt, transfer wheelchair/car); range: 0 (completely disabled) to 100 (com- pletely independent)	no	feeding, bathing, dressing, grooming, sphincter mgmt, use of toilet, transfers, mobil- ity, stair mgmt
WISCI15	measure of mobility designed for SCI pts; 20-item scale: reflection of walking level, use of devices & phys assis- tance for a distance of 10 m; range 0 (unable to walk) to 20 (walking w/o braces/assistance/de- vices for at least 10 m)	yes; SCI pts	walking

*** clin = clinical; compl = complete; depr = depression; difft = different; disab = disability; dysfx = dysfunction; emot'l = emotional; fx = function; fx'l = functional; gen = general; incl = including; LBP = low-back pain; LSOQ = Lumbar Spine Outcome Questionnaire; mgmt = management; min = minimal; mod = moderate; NASS-LS = NASS Lumbar Spine Outcome Assessment Instrument; phys = physical; psych = psychological; psychosoc = psychosocial; pt = patient; Q = question; RADL = Resumption of Activities of Daily Living Scale; soc = social**

Table 1: Existing outcome measurements for spinal pathology

The HADS is a self-rating instrument for patients with both somatic and mental problems.[65] It was designed in 1983 to identify anxiety or depression disorders in nonpsychiatric hospital clinics. It is divided into an anxiety subscale (HADS-A) and a depression subscale (HADS-D). It performs well in screening for the separate dimensions of anxiety or depression disorders.

All of these depression outcome measurements are mostly used in SCI patients. There has been a lot of research, but none of the measurements are validated for use in spinal trauma patients.

Disability Indices

The Oswestry Disability Index (ODI) and the Roland-Morris Disability Questionnaire (RMDQ) are 2 widely used condition-specific instruments that assess disability in patients with chronic low-back pain.[18,47] These instruments are included here as they are illustrative of the outcome measures commonly employed in this related population. The ODI is a self-administered tool consisting of sections assessing ADL. It seems slightly more sensitive to improvements in condition than the RMDQ. It is one of the most widely used outcome measurements for patients with low-back pain. The RMDQ was derived from the SIP and assesses pain and daily function.

It has been extensively validated and demonstrates good reliability and consistency in patients with low-back pain. The ODI has been directly compared with the RMDQ in several studies. It seems that the ODI is better at detecting change in more seriously disabled persons, whereas the RMDQ may have an advantage in patients with minor disability.[19] Both of these instruments are designed and mainly used to assess disability associated with chronic low-back pain.

The Million Visual Analog Scale (MVAS) consists of 15 questions, with a VAS for each.[42] It is not as widely used as the ODI or RMDQ and is not validated for spinal trauma.

The Low-Back Outcome Score (LBOS) was developed by Greenough and Fraser[24] as a quick score for assessment of disability and pain. It is used in spinal trauma patients, but has not been validated.[52]

The Quebec Back Pain Disability Scale (QBPDS) is validated for patients with low-back pain.[35] In comparison between the QBPDS and the ODI, the ODI showed higher test-retest reliability and responsiveness.[21]

Another interesting questionnaire is the Resumption of Activities of Daily Living Scale (RADLS). As previously mentioned, most questionnaires fail to take into account what was "normal" or "usual" prior to injury. Also, patients' perceptions of readiness to return to work may not agree with clinicians' judgments.[62]

The AAOS/NASS spine questionnaires are adaptations of the ODI and the RMDQ.[13] They differ from the ODI and RMDQ in that they are applied as preoperative and follow-up modules. There are specific questionnaires for the cervical and lumbar spine. In addition to the features found in the ODI and the RMDQ, the AAOS/NASS disease-specific questionnaire contains a comorbidity index (14 questions) and questions on physical health and pain (16 questions), treatment expectations (5), satisfaction with symptoms (1), neurogenic symptoms (6), and pain/ disability (11). Normative values have been published.[33] An analogous instrument, the Neck Disability Index, is used for similar purposes in patients with cervical spine disorders.[58]

Functional Measures

The practice of measuring a single meaningful parameter of outcome has also been implemented as a way to expedite the outcome assessment process. Measures of work status or unidimensional measures of mobility are often included in the commonly used generic and functional outcome measures, but they are also applied independently.

The return to work (RTW) measure represents an interesting construct in that it is a meaningful outcome from a societal as well as an individual point of view and it can serve as an indirect proxy for the domain of participation as well as function. Of course it also reflects the other domains of impairment and activities, making it a complex as well as a useful construct.[8,53] The RTW has been used as a primary outcome in a patient population that underwent surgery for spinal fractures resulting from high-energy trauma.[41] The presence of neurological

symptoms was found to be a major predictor of RTW status at follow-up.

The FCI has been validated in major trauma populations.[37,38] It is a "preference based, multi-attribute functional outcome measure that provides 10 dimension specific scores and one overall score that summarizes function across the 10 dimensions."[37] The dimensions include "eating, excretory function, sexual function, ambulation, hand/arm movement, bending/lifting, vision, hearing, speech and cognitive function."[37] It does not assess psychosocial outcomes and is targeted mainly toward measuring outcome in general trauma patients, reflecting the domains pertinent to that group. It merits inclusion here as it is one of the only outcome measures designed specifically for trauma populations.[51]

More comprehensive functional outcome measures such as the FIM and SCIM incorporate a mobility domain. The FIM is a well-established functional outcome assessment score that also relies on the patient's ability to perform ADL.[56,59] Although initially designed for stroke patients, it has been extensively used in the assessment of trauma outcome and is likely the most widely used functional outcome measure for the SCI population, with demonstrated validity and reliability in that specific population.[48,63]

Although the use of the FIM has merit in spinal trauma patients, it is not free of limitations. Briefly, the FIM scoring aggregation creates a masking effect in that unequal ordinal items are simply summed. Moreover, the FIM measures a narrow set of domains reflecting impairments that are primarily specific to SCI.[30,40] Other limitations are that the FIM is usually administered by a certified therapist, it is largely designed for hospitalized patients, and it is not a patient-directed questionnaire. A "phone FIM" instrument is now also available, however, so direct patient contact is not necessary.

The measure of health service use (HSU) is similar to the RTW in that it is a complex but also useful and important construct for assessing injury outcome. The complexity and multifactorial nature of this measure have been established but not adequately explored in trauma patients, let alone spinal trauma patients.[50] Like the RTW, it is an important measure from an economic and social point of view, but any inferences drawn from it about the efficacy of treatment are tenuous at best.

Spinal Cord Injury Measures

The SCIM is a recently developed disability measure designed and validated for patients with spinal cord lesions.[11] It measures a patient's ability to perform daily tasks and demonstrates good reliability and validity in the SCI population subset. Specifically designed for patients with substantial neurological damage, it would not be suitable for use in a spinal trauma population with no or varying degrees of neurological involvement. Nevertheless, it still warrants mention because it is an excellent tool for assessing the spinal cord–related functional outcome component of spinal injury.[45] The WISCI is also a recently developed measure of mobility designed for SCI, with demonstrated high sensitivity.[43] It is validated for SCI patients with good correlation.[15]

The ICF and Spinal Trauma

The ICF framework reflects a biopsychosocial model of health and functioning. It is generic in nature and not designed with any particular patient group in mind. As such, the practice of selecting a subset of ICF domains and constructs to generate condition-specific "ICF core sets" has emerged to take better advantage of the ICF in specific patient groups.[7] This approach has been applied both to the SCI population and to patients with acute and postacute musculoskeletal conditions.[6,14,49,57] For acute musculoskeletal injuries, a consensus conference selected 47 ICF categories for inclusion into a "core set."[57] The core set emphasizes the integrity and function of musculoskeletal structures.[7,57] By comparison, for longer-term follow-up of these patients, emphasis was placed on activities/participation.

The target patient population did not include SCI trauma patients, who were lumped into a different core set, namely that of acute and postacute neurological patients.[17,55] This population includes patients with head injury, cerebrovascular diseases, and CNS neoplasms. A large number of the categories selected for this population would be tangential to both SCI and non-SCI spine trauma.

The postacute core sets as they are described above seem to be more applicable to outcome assessment, as the acute core sets include assessments of the current state of patients in the acute setting (for example temperature and electrolyte levels). These categories are not meant as outcomes but rather as classification and prognostic aides. The postacute core sets discussed here were generated for trauma and neurological injury populations, and while spinal trauma patients form subsets of both populations, parts of both core sets are not relevant to spinal trauma populations.

Table 2 presents a list of spinal trauma–relevant ICF criteria that we have generated to evaluate how effectively frequently used outcome instruments address spinal trauma outcome domains.

DISCUSSION

This review sought to assess the current state of outcome measurement in spinal trauma patients and to address the question of whether this group is adequately served by current disease-specific and generic health-related QOL instruments. Several overarching trends were found to be significant to the spinal trauma outcome field.

While the SF-36 and other widely used generic instruments provide normative data that allow for demographically adjusted approximations of preinjury scores and comparisons between populations, these instruments were not designed for any specific population. This "one-size-fits-all" approach may be a source of methodological limitations (ceiling and floor effects, large minimally clinically important difference, and low responsiveness) if they are relied on too heavily as primary measures in spinal trauma. Indeed, the applicability of generic health-related QOL assessments as primary outcome measures in SCI studies has come under question recently, as measures of function might be more meaningful

	Instrument										
Domain†	**SF-36**	**EQ-5D**	**HUI**	**ODI**	**RMQD**	**FCI**	**WISCI**	**FIM**	**SCIM**	**RTW**	**HSU**
mental function	+	+	+			+		+			
sensory & pain	+	+	++	+		+					
genitourinary & reproductive						+		+	+		
neuromuscul-oskeletal						+	+		+		
nervous system											
movement-related structures											
mobility	+	+	+	+	+	+	++	++	++		
self-care	+	+		+		+		++	++		
domestic life	+			±	±						
interpersonal relationships	+			+	±	+		+			
major life areas	+	±								++	
community, social, & civic life	+			+	±						
support & relation-ships					±						+
services systems & policies											++

*** This table provides a rough indication of which topics relevant to spinal trauma are addressed in the instruments analyzed. It indicates only their presence or absence and does not address the quality of these items in the measurement instruments. Definition of symbols: + = included in outcome measurement; ++ = emphasized in outcome measurement; ± = mentioned in outcome measurement. † Putative spinal trauma domains**

Table 2: Overview of ICF criteria measured in selected outcome instruments

and sensitive in discriminating between different treatment modalities.[57]

Spine-specific (low-back or cervical) outcome measures, on the other hand, may theoretically have limited applicability to spinal trauma patients in the sense that the domains they measure and the relative weighting of each in the scoring do not correspond to the domains perceived to be important in spinal trauma patients, a fundamentally different population. For instance, chronic pain is probably a lesser issue in trauma patients, including spinal trauma victims, than in chronic cervical or lumbar pain syndromes.[2] Although pain may be relevant for SCI patients,[54] the lower incidence and lower pain scores in spinal trauma patients after the acute phase make the emphasis placed on pain measures in the instruments used in chronic patients potentially misplaced in the spinal trauma population. Additionally, the high psychometric quality of the ODI and RMDQ in patients with low-back pain might not apply to the very different spinal trauma patient group (because of ceiling and floor effects and responsiveness issues, for example).

Concerning the functional outcome measurements, the literature shows that after 12 months of follow-up, the FCI does not correlate with the initial FCI or initial trauma severity, indicating that the FCI may not be a suitable tracker of outcome progression over time.[51] While outcome measures like the WISCI might be useful in populations with substantial SCI, they are too narrow to form the basis of evidence-based decisions in spinal trauma patients with a broad spectrum of neurological involvement and disability.

Our literature search identified only 1 outcome paper dealing directly with acute trauma patients as a single population, and the authors of that paper used the FIM.1 The study showed that FIM scores improved significantly in survivors at 12 months posttrauma. Additionally, no correlation was found between results of the FIM administered on admission and results obtained 1 year post-injury. The authors also did not find a significant correlation between the Injury Severity Score (ISS) and 1-year FIM outcome. Interestingly, a different study performed in spinal trauma patients with polytrauma did find a correlation between the ISS and 1- and 2-year FIM scores.[29] This discrepancy is possibly related to the study populations and/ or the differing statistical methods applied. The study investigating spine patients with polytrauma further found that the severity of spinal injuries was the most important predictor of disability as measured by the FIM, although the authors acknowledge that this finding may relate to the spine bias of the FIM.[29] Nonetheless, this result again suggests the importance of spinal trauma in mid- to long-term disability and emphasizes the selection of outcome measures.

The 2 single-item outcomes, RTW and HSU, are interesting in that they produce a single economically significant outcome measure directly dependent on multiple parameters spread over many health domains. Assessments using these measures are easy to conduct, but one should use caution in analyzing and interpreting the results because they depend on a poorly defined collection of multiple patient-related factors.

The available outcome research was assessed in the trauma and SCI fields

in an attempt to extract and synthesize relevant conclusions about the state of outcome research in spinal trauma, an approach necessitated by the paucity of research specifically involving the spinal trauma patient population. This is a further indication that outcome in spinal trauma patients might not be adequately measured with existing instruments; most of the outcome analysis was being performed in the general (multi)trauma population and/or the SCI population, with spinal trauma patients sometimes constituting a subset in the trauma studies. Conversely, in SCI outcome studies a large proportion of the patients have SCIs of traumatic etiology and therefore represent a subset of spinal trauma patients.

On the other hand, patients with chronic lumbar or cervical pain constitute a distinct population in which much outcome research is being conducted. Although many studies use these outcome measurements for spinal trauma patients, there is actually little overlap between these patients with chronic conditions and the patients with acute spinal trauma.

Identifying those ICF criteria relevant to spinal trauma and evaluating in a summary form which domains were measured in the outcome tools are important issues in the discussion on this subject. The putative ICF domains that we selected as likely to be relevant to spinal trauma demonstrate that no single outcome measure adequately addresses all the major domains likely to be relevant to spinal trauma. After reviewing both the most commonly used outcome measures and current consensus on the underlying domains, we have come to the conclusion that the outcome tools in current use are not adequate for assessing spinal trauma patients as a specific population. It seems that the ICF criteria pertinent to the spinal trauma group could be more or less completely covered by a combination of generic and condition-specific trauma and SCI outcome assessments. The most inclusive approach would be to select a combination of measures, but this would also lead to a significant amount of redundancy, as well as creating the possibility of psychometric issues.

The lack of tools designed for the spinal trauma population and the lack of research into the applicability and validity of existing tools to the spinal trauma population produce a situation in which the efficacy of interventions targeting this group cannot be readily evaluated. Thus the status quo, in which spinal trauma patients are effectively split between trauma patients and SCI patients, is suboptimal, leading to psychometric limitations and redundancy. Additionally it would be desirable to treat spinal trauma patients as a single unified group because of the commonality of SCI and non-SCI patients, and to employ a single psychometrically validated instrument specifically tailored to the unique dynamics of spinal trauma.

A proper spinal trauma injury outcome tool should probably be a combination of already-existing questionnaires and, in our opinion, focus on resumption of activities in comparison with pretrauma level of functioning. Computer adaptive tests or item response theory can be helpful in combining relevant domains and activities for spinal trauma patients to create an outcome instrument for spinal trauma patients with or without SCI.[27,28,34]

CONCLUSIONS

Outcome instruments currently in use fail to capture many of the domains relevant to spinal trauma patients. There is currently very little work that evaluates these patients as a single specific population. In this study, an evidence-based preliminary list of domains has been generated to serve as the basis for further discussion of which domains are pertinent to spinal trauma patients.

REFERENCES

1. **Akmal M, Trivedi R, Sutcliffe J:** Functional outcome in trau-ma patients with spinal injury. *Spine 28:180–185, 2003*

2. **Anagnostis C, Gatchel RJ, Mayer TG:** The Pain Disability Ques tionnaire: a new psychometrically sound measurefor chronic musculoskeletal disorders. *Spine 29:2290–2303,2004*

3. **Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J:** Aninventory for measuring depression. *Arch Gen Psychiatry 4:561–571, 1961*

4. **Bendebba M, Dizerega GS, Long DM:** The Lumbar Spine Outcomes Questionnaire: its development and psychometric properties. *Spine J 7:118 –32, 2007*

5. **Bergner M, Bobbitt RA, Carter WB, Gilson BS:** The Sickness Impact Profile: development and final revision of a health status measure. *Med Care 19:787–805, 1981*

6. **Biering-Sørensen F, Charlifue S, DeVivo M, Noonan V, Post M, Stripling T, et al:** International Spinal Cord Injury Data Sets. *Spinal Cord 44:530–534, 2006*

7. **Biering-Sørensen F, Scheuringer M, Baumberger M, Charlifue SW, Post MW, Montero F, et al:** Developing core sets for persons with spinal cord injuries based on the International Classification of Functioning, Disability and Health as a way to specify functioning. *Spinal Cord 44:541–546, 2006*

8. **Brenneman FD, Redelmeier DA, Boulanger BR, McLellan BA, Culhane JP:** Long-term outcomes in blunt trauma: who goes back to work? *J Trauma 42:778–781, 1997*

9. **Brooks R:** Quality of life measures. *Crit Care Med 24:1769, 1996*

10. **Butler JS, Fitzpatrick P, Ni Mhaolain AM, Synnott K, O'Byrne JM:** The management and functional outcome of isolatedburst fractures of the fifth lumbar vertebra. *Spine 32:443–447, 2007*

11. **Catz A, Itzkovich M, Agranov E, Ring H, Tamir A:** SCIM–Spinal Cord Independence Measure: a new disability scale for patients with spinal cord lesions. *Spinal Cord 35:850–856, 1997*

12. **Crichlow RJ, Andres PL, Morrison SM, Haley SM, Vrahas MS:** Depression in ortho-paedic trauma patients. Prevalence and severity. *J Bone Joint Surg Am 88:1927–1933, 2006*

13. **Daltroy LH, Cats-Baril WL, Katz JN, Fossel AH, Liang MH:**The North American Spine Society Lumbar Spine Outcome Assessment Instrument: reliability and validity tests. *Spine 21:741–749, 1996*

14. **DeVivo M, Biering-Sørensen F, Charlifue S, Noonan V, Post M, Stripling T, et al:** International Spinal Cord Injury Core Data Set. *Spinal Cord 44:535–540, 2006*

15. **Ditunno JF, Scivoletto G, Patrick M, Biering-Sorensen F, Abel R, Marino R:** Validation of the walking index for spinal cord injury in a US and European clinical population. *Spinal Cord 46:181–188, 2008*

16. **EuroQol Group:** EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy 16:199–208, 1990*

17. **Ewert T, Grill E, Bartholomeyczik S, Finger M, Mokrusch T, Kostanjsek N, et al:** ICF Core Set for patients with neurological conditions in the acute hospital. *Disabil Rehabil 27:367–373, 2005*

18. **Fairbank JC, Couper J, Davies JB, O'Brien JP:** The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy 66:271–273, 1980*

19. **Fairbank JC, Pynsent PB:** The Oswestry Disability Index. *Spine 25:2940–2952, 2000*

20. **Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al:** Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 System. *MedCare 40:113–128, 2002*

21. **Fritz JM, Irrgang JJ:** A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther 81:776–788, 2001*

22. **Furlong WJ, Feeny DH, Torrance GW, Barr RD:** The Health Utilities Index (HUI) System for Assessing Health-Related Quality of Life in Clinical Studies, in Centre for Health Economics and Policy Analysis Research Working Paper Series, 01-02. *Hamilton, Canada: McMaster University, 2001*

23. **Gabbe BJ, Williamson OD, Cameron PA, Dowrick AS:** Choosing outcome assessment instruments for trauma registries. *Acad Emerg Med 12:751–758, 2005*

24. **Greenough CG, Fraser RD:** Assessment of outcome in patients with low-back pain. *Spine 17:36–41, 1992*

25. **Hallin P, Sullivan M, Kreuter M:** Spinal cord injury and quality of life measures: a review of instrument psychometric quality. *Spinal Cord 38:509–523, 2000*

26. **Hamilton M:** A rating scale for depression. *J Neurol Neurosurg Psychiatry 23:56–62, 1960*

27. **Hart DL, Mioduski JE, Werneke MW, Stratford PW:** Simulated computerized adaptive test for patients with lumbar spine impairments was efficient and produced valid measures of function. *J Clin Epidemiol 59:947–956, 2006*

28. **Hays RD, Morales LS, Reise SP:** Item response theory and health outcomes measurement in the 21st century. *Med Care 38 (9 Suppl):II28–II42, 2000*

29. **Hebert JS, Burnham RS:** The effect of polytrauma in persons with traumatic spine injury. *A prospective database of spine fractures. Spine 25:55–60, 2000*

30. **Hetherington H, Earlam RJ:** Measurement of disability after multiple injuries: the Functional Independence Measure. *Clinical review. Eur J Surg 161:549–555, 1995*

31. Holtslag HR, Post MW, Lindeman E, Van der Werken C: Longterm functional health status of severely injured patients. *Injury 38:280–289, 2007*

32. Hu R, Mustard CA, Burns C: Epidemiology of incident spinal fracture in a complete population. *Spine 21:492–499, 1996*

33. Hunsaker FG, Cioffi DA, Amadio PC, Wright JG, Caughlin B: The American Academy of Orthopaedic Surgeons outcomes instruments: normative values from the general population. *J Bone Joint Surg Am 84:208–215, 2002*

34. Jette AM, Haley SM, Tao W, Ni P, Moed R, Meyers D, et al: Prospective evaluation of the AM-PAC-CAT in outpatient rehabilitation settings. *Phys Ther 87:385–398, 2007*

35. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, et al: The Quebec Back Pain Disability Scale. Measurement properties. *Spine 20:341–352, 1995*

36. Ku JH: Health-related quality of life in patients with spinal cord injury: review of the Short Form 36-Health Questionnaire Survey. *Yonsei Med J 48:360–370, 2007*

37. MacKenzie EJ, Damiano A, Miller T, Luchter S: The development of the Functional Capacity Index. *J Trauma 41:799–807, 1996*

38. MacKenzie EJ, Sacco WJ, Luchter S, Ditunno JF, Staz CF, Gruen GS, et al: Validating the Functional Capacity Index as a measure of outcome following blunt multiple trauma. *Qual Life Res 11:797–808, 2002*

39. MacKenzie EJ, Shapiro S, Smith RT, Siegel JH, Moody M, Pitt A: Factors influencing return to work following hospitalization for traumatic injury. *Am J Public Health 77:329–334, 1987*

40. Matheson LN: Functional outcome in trauma patients with spinal injury. *Spine 28:105–106, 2003*

41. McLain RF: Functional outcomes after surgery for spinal fractures: return to work and activity. *Spine 29:470–477, discussion Z6, 2004*

42. Million R, Hall W, Nilsen KH, Baker RD, Jayson MI: Assessment of the progress of the back-pain patient 1981 Volvo Award in Clinical Science. *Spine 7:204–212, 1982*

43. Morganti B, Scivoletto G, Ditunno P, Ditunno JF, Molinari M: Walking index for spinal cord injury (WISCI): criterion validation. *Spinal Cord 43:27–33, 2005*

44. Müller U, Röder C, Greenough CG: Back related outcome assessment instruments. *Eur Spine J 15 (Suppl 1):S25–S31, 2006*

45. Németh G: Health related quality of life outcome instruments. *Eur Spine J 15 (Suppl 1):S44–S51, 2006*

46. Neugebauer E, Bouillon B, Bullinger M, Wood-Dauphinée S: Quality of life after multiple trauma–summary and recommendations of the consensus conference. *Restor Neurol Neurosci 20:161–167, 2002*

47. Roland M, Morris R: A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine 8:141–144, 1983*

48. Scheuringer M, Grill E, Boldt C, Mittrach R, Müllner P, Stucki G: Systematic review of measures and their concepts used in published studies focusing on rehabilitation in the acute hospital and in early post-acute rehabilitation facilities. *Disabil Rehabil 27:419–429, 2005*

49. Scheuringer M, Stucki G, Huber EO, Brach M, Schwarzkopf SR, Kostanjsek N, et al: ICF Core Set for patients with musculoskeletal conditions in early post-acute rehabilitation facilities. *Disabil Rehabil 27:405–410, 2005*

50. Schluter PJ, Cameron CM, Purdie DM, Kliewer EV, McClure RJ: How well do anatomical-based injury severity scores predict health service use in the 12 months after injury? *Int J Inj Contr Saf Promot 12:241–246, 2005*

51. Schluter PJ, Neale R, Scott D, Luchter S, McClure RJ: Validating the Functional Capacity Index: a comparison of predicted versus observed total body scores. *J Trauma 58:259–263, 2005*

52. Shen WJ, Liu TJ, Shen YS: Nonoperative treatment versus posterior fixation for thoracolumbar junction burst fractures without neurologic deficit. *Spine 26:1038–1045, 2001*

53. Soberg HL, Finset A, Bautz-Holter E, Sandvik L, Roise O: Return to work after severe multiple injuries: a multidimensional approach on status 1 and 2 years postinjury. *J Trauma 62:471–481, 2007*

54. Steeves JD, Lammertse D, Curt A, Fawcett JW, Tuszynski MH, Ditunno JF, et al: Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP panel: clinical trial outcome measures. *Spinal Cord 45:206– 221, 2007*

55. Stier-Jarmer M, Grill E, Ewert T, Bartholomeyczik S, Finger M, Mokrusch T, et al: ICF Core Set for patients with neurological conditions in early post-acute rehabilitation facilities. *Disabil Rehabil 27:389–395, 2005*

56. Stineman MG, Shea JA, Jette A, Tassoni CJ, Ottenbacher KJ, Fiedler R, et al: The Functional Independence Measure: tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. *Arch Phys Med Rehabil 77: 1101–1108, 1996*

57. Stoll T, Brach M, Huber EO, Scheuringer M, Schwarzkopf SR, Konstanjsek N, et al: ICF Core Set for patients with musculoskeletal conditions in the acute hospital. *Disabil Rehabil 27: 381–387, 2005*

58. Stoll T, Huber E, Bachmann S, Baumeler HR, Mariacher S, Rutz M, et al: Validity and sensitivity to change of the NASS questionnaire for patients with cervical spine disorders. *Spine 29:2851–2855, 2004*

59. UDS Data Management Service: Guide for Use of the Uniform Data Set for Medical Rehabilitation Including the Functional Independence Measure, Version 4. *Buffalo: State University of New York at Buffalo, 1993*

60. Van Beeck EF, Larsen CF, Lyons RA, Meerding WJ, Mulder S, Essink-Bot ML: Guidelines for the conduction of follow-up studies measuring injury-related disability. *J Trauma 62: 534–550, 2007*

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

62. Williams RM, Myers AM: A new approach to measuring recovery in injured workers with acute low back pain: Resumption of Activities of Daily Living Scale. *Phys Ther 78:613–623, 1998*

63. Wood-Dauphinée S, Exner G, Bostanci B, Exner G, Glass C, Jochheim KA, et al: Quality of life in patients with spinal cord injury—basic issues, assessment, and recommendations. *Restor Neurol Neurosci 20:135–149, 2002*

64. World Health Organization: International Classification of Functioning, Disabilities and Health (ICF). *Geneva: World Health Organization, 2001*

65. Zigmond AS, Snaith RP: The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand 67:361–370, 1983*

5

Surgeon equipoise as an inclusion criterion for the evaluation of nonoperative versus operative treatment of thoracolumbar spinal injuries

Agnita Stadhouder
F. Cumhur Öner
Kevin W. Wilson
Alex R. Vaccaro
Owen D. Williamson
Abraham J. Verbout
Jan A. Verhaar
Luuk W.L. de Klerk
Erik Buskens

The Spine Journal, Volume 8, Number 6, 975-981 March 2008.

ABSTRACT

Object. Valid outcome assessment tools specific for spinal trauma patients

Background Context: Prospective studies have failed to demonstrate the superiority of either operative or nonoperative treatment of thoracolumbar fractures. Similar to other surgical fields, research has been limited by the variability in surgical interventions, difficult recruitment, infrequent pathology, and the urgency of interventions.

Purpose: To outline factors precluding randomized controlled trials in spinal fractures research, and describe a novel methodology that seeks to improve on the design of observational studies.

Study Design/Setting: A preliminary report describing an observational study design with clinical equipoise as an inclusion criterion. The proposed methodology is a cohort study with head- to-head comparison of operative and nonoperative treatment regimens in an expertise-based trial fashion. Patients are selected retrospectively by an expert panel and clinical outcomes are assessed to compare competing treatment regimens. Surgeon equipoise served as an inclusion criterion.

Patient Sample: Patients with closed or open thoracolumbar spinal fracture with or without neurological impairment, presenting to one of two different trauma centers between 1991 and 2005 (N=760).

Outcome Measures: Homogeneity of baseline clinical and demographic data and distribution of prognostic risk factors between the operative and the nonoperative cohort.

Methods: Patients treated for spine fractures at two University hospitals practicing opposing methods of fracture intervention were identified by medical diagnosis code searches (n=760). A panel of spine treatment experts, blinded to the treatment received clinically has assessed each case retrospectively. Patients were included in the study when there was disagreement on the preferred treatment, that is, operative or nonoperative treatment of the injury. Baseline and initial data of a study evaluating nonoperative versus operative spinal fracture treatment are presented.

Results: One hundred and ninety patients were included in the study accounting for a panel discordance rate of 29%. The distribution of baseline characteristics and demographics of the study populations were equal across the parallel cohorts enrolled in the study, that is, no differences in prognostic factors were observed.

Conclusions: The use of clinical equipoise as an inclusion criterion in comparative studies may be used to avoid selection bias. Using multivariate analysis of retrospectively assembled parallel cohorts, a valid comparison of operative and nonoperative spine fracture treatment strategies and their outcomes is possible.

INTRODUCTION

Although the randomized controlled trial (RCT) is widely accepted as the paradigm for evaluating the effect of therapeutic interventions,[1-3] practical and technical barriers to proper RCTs in surgical fields have been identified. The inherent variability of interventions, infrequent pathology, lack of recruitment, and urgent clinical decision making create problems with standardization and timing in RCTs.[4] These factors are particularly problematic when comparing surgical treatment with medical management.[5] The barriers to surgical RCTs are particularly prohibitive in clinical spinal fracture research, resulting in a paucity of valid evidence supporting either nonoperative or surgical management.

Proponents of nonoperative management claim that almost all types of injury can be successfully treated with nonoperative measures.[6-10] However, long-term complications after nonoperative treatment such as persistent pain and progressive deformity are well known and recognized.[11] Operative treatment has gained popularity since the 1980s, after the introduction of relatively simple and effective fixation techniques aimed at improving long-term results. As usual in surgical practice, operative treatment has been gradually introduced to treat spine fractures without valid comparison. Numerous studies have been published with a wide variety of clinical and radiological outcomes, which often contradict each other.[6-17] Evidence of treatment superiority is lacking even in critical outcomes such as neurologic recovery.[18]

The only large-scale prospective multicenter survey, conducted by the Scoliosis Research Society between 1986 and 1991, was limited by several shortcomings.[18] Various other attempts at prospective studies have found that randomization was not feasible. [6,9] Recently, Wood et al. described a prospective randomized study comparing operative and nonoperative treatment for burst fractures in patients without neurological deficit.[19] It is the first study of this kind regarding treatment of spinal fractures, but multiple operative strategies limited the interpretation and generalizability of the results.[20]

Clinical equipoise, as described by Freedman, exists when there is a genuine uncertainty within the expert medical community about the optimal treatment of a certain disease.[21] In the presence of equipoise, it is common to form ''schools'' based on convictions and obduracy of clinical superiority. In the treatment of spinal fractures, both nonoperative and operative schools have become well established in different hospitals, with vast resources and clinical experience in support of both interventional modalities. Surgeons in the operative ''school'' are reluctant to treat neurologically impaired patients in a nonoperative manner, whereas nonoperative ''schools'' argue against the expense and invasive nature of surgery. With polarized opinions and sincere concern for patients receiving inferior treatment, it has become difficult for surgeons to agree on an RCT design.

Moreover, even if an RCT design is approved with sufficient surgeon participation, a properly conducted RCT may remain difficult from a logistical standpoint. To begin with, acquiring informed consent from patients present-

ing to a trauma center with spinal fractures and neurologic impairment is often impractical because of the patient's general condition or the need for immediate intervention. The power and generalizability of a study are often jeopardized by limited recruitment. Preliminary power analysis usually demonstrates a need for multicenter studies, which creates another host of problems with center effects or surgeon effects.

Another major preclusion to RCTs in spine surgery remains the complexity of intervention. RCTs are best applied for problems where the intervention is straightforward and easily defined and implemented such as when comparing pharmaceutical interventions. If the intervention is a complex procedure, such as in surgery, or if patients themselves are complex, such as with multiple injuries, it is not clear whether randomization would yield valid results.[22,23] In spinal fractures, both nonoperative treatment and operative treatment require specialists, supported by specific infrastructure and experience, to provide optimal care in a routine manner. For an ideal conventional RCT, the ''nonoperative'' clinic would have to switch to operating on these patients on a regular basis, and the ''operative'' clinic would have to switch to providing adequate nonoperative care to patients with unstable fractures. Suboptimal treatments can be expected in such a nonoptimized setting. In fact, ethical concerns may rise in a study design if patients do not receive optimal treatment.[24] The Declaration of Helsinki states, ''In any medical study, every patient -including those of a control group, if anyshould be assured of the best proven diagnostic and therapeutic method''.[1] Randomization of all known and unknown confounders with regard to fracture healing and functional outcome is the optimum method of assuring validity in treatment outcomes, especially when physicians have strong biases on perceived treatment alternatives. The logistics and cost of setting up a prospective randomized trial in a spinal trauma population evaluating two opposed treatment regimens may be prohibitive. Moreover, randomization, of course, could not be blinded, as surgery and nonoperative care are obviously distinctively different methods of treatment. Obviously patients, surgeons, and staff may be influenced by knowledge of competing treatment regimens. Another hypothetical option would be to initially randomize patients during in-field assessment and then transfer them to the clinic offering the allotted treatment strategy. One easily recognizes the potential risk and unacceptable burden to the patients and emergency care providers because of confusion in treatment allocation during an emergency, potential delayed or prolonged transportation times, all leading to limited recruitment and considerable dropout.

Where an RCT cannot be adequately performed, as often is the case in surgery, observational studies are frequently used as the alternative.[26] Black stated that randomized trials and observational studies each have their strengths and weaknesses and should be seen as complementary.[2] As a general rule, observational studies such as cohort studies, quasi-cohort studies, or case-control studies are all considered to be less valid because of the possible bias created by intentional treatment allocation by physicians or patients.[27]

Considering the difficulties of RCTS in spinal surgery and the limitations of observational studies, we propose a different method for comparing operative and nonoperative interventions, clinical equipoise. What is unique about the proposed methodology is the presentation of a protocol ensuring that through a blinded assessment of eligibility by a panel of experts, uncertainty with respect to optimal treatment is an inclusion criterion. To the best of our knowledge, this is the first time a retrospectively assembled cohort design that uses surgeon equipoise as an inclusion criterion is described.

METHODS

The design is a cohort study in which patients are retrospectively identified and prospectively followed to gather outcome data. The clinical question, critical sample size, and measurable outcomes must be established first. The feasibility of the study can be determined using historical data on the number of patients with the targeted condition treated at selected hospitals. Through a medical record diagnosis code search, a cohort of patients are assembled from all patients with a clinical condition of interest presenting at the two participating hospitals practicing opposite "schools" of intervention. The treatment strategies need to be clearly defined and consistently provided to the patients with the specific diagnosis.

Key prognostic characteristics within the subgroup of patients must be identified to account for possible confounding variables. Comparison of outcomes can subsequently be adjusted for significant confounders. Baseline characteristics and data of each patient's condition at presentation should be collected into a comprehensive database and made anonymous. All eligible patients must be considered to prevent selection bias as the result of dropouts before the unique selection process is applied.

The anonymous data sets are then presented to a panel of experts blinded to the actual treatment and the origin of the case. The expert panel should consist of an equal distribution of representatives from the differing "schools" of intervention. The panel decides the preferred treatment strategy for that particular patient as if the patient were brought to their hospital at that moment. All relevant data to reach a "clinical" decision should be available, including the basic clinical, laboratory, and radiological information. Patients are selected for inclusion when there is disagreement on the treatment of choice between the "schools" of intervention. Patients are excluded when the panel agrees on the preferred treatment (see Fig. 1 for details). The group of eligible patients should comprise individuals who would have received treatment A in school A, but were in fact seen and treated according to the preferences of center B, and vice versa.

After obtaining informed consent from eligible patients, follow-up data of interest may be gathered from patients and their physicians. Prespecified protocols for follow-up and standardized measures of outcomes are required to avoid detection bias. Prospective data can include a combination of standardized clinic

follow-up evaluations, questionnaires, laboratory or radio-diagnostic studies, and specific criteria for measuring outcomes of interest. It is vital to collect contact information on the eligible patients at baseline to successfully complete the prospective aspect of the study and avoid transfer bias as the result of the differential loss to follow-up.

As in all retrospective studies, completeness of medical record baseline information is of utmost importance. This highlights the importance of developing a sound research question and appropriate outcome parameters before commencing the study. Otherwise, missing data might create bias and diminish power because of the inevitable exclusion of patients. It is vital to collect baseline data in a systematic manner so that the analysis can detect any systematic difference in baseline characteristics of dropouts and those who complete the study. The influence of the dropouts on the measured effect for each treatment can then be determined.

During the analysis phase, statistical adjustment must be made to account for any unequal distribution of baseline characteristics and known confounders through a multivariate analysis. Methods to test and account for known confounders and effect modification, or interactions between potential known confounders, are well described and can be implemented in the analysis phase of studies of this design.[25]

RESULTS

The proposed research methodology was tested in a study of thoracolumbar fractures in patients presenting to two different trauma centers, with opposing preferences in the management of thoracolumbar spinal injuries. Medical diagnosis codes were searched for closed or open thoracolumbar spinal fracture with or without neurological impairment, identifying 760 possible cases between 1991 and 2005. Because of incomplete data, 124 patients were excluded, leaving 636 patients to be evaluated by the expert panel. Other exclusion criteria were osteoporosis, corticosteroid use, and pathologic fracture.

The expert panel consisted of one orthopedic surgeon from each hospital. Annually, each clinic treats approximately 60 patients with traumatic spinal fractures. The 636 cases were collected from hospital records and placed in a standardized format for presentation to the panel. The package, blinded to the hospital of presentation and treatment received, included digitized plain films and a digitalized database with history of present injury, neurological impairments, coincident injuries, medical and surgical histories, medications, social history, and level of care provided until arrival at the treatment center. Computed tomography was available in 70% of cases. Data were presented as if the patient had just arrived at the treatment center. The attending surgeons made the decisions following their experience and routine according to the practices of their ''schools'' of intervention. The reviewers were allowed to independently review each case, with no discussion

among panel members. The decision to operate or not operate was submitted for each case to an independent moderator who rejected cases where consensus was made. In 446 patients, there was agreement on the treatment choice and these cases were rejected for an exclusion rate of 71%. Disagreement was found in 190 cases, and these patients were selected for further study for an inclusion rate of 29%.

The treatment protocols in both centers were well defined and consistent during the study period. The operative treatment protocol in one hospital consisted of short-segment posterior stabilization, bridging one or two disc spaces as needed, followed by a protective brace for 12 weeks. In multiple-vertebral fractures, treatment consisted of posterior long-segment fixation. The treatment at the

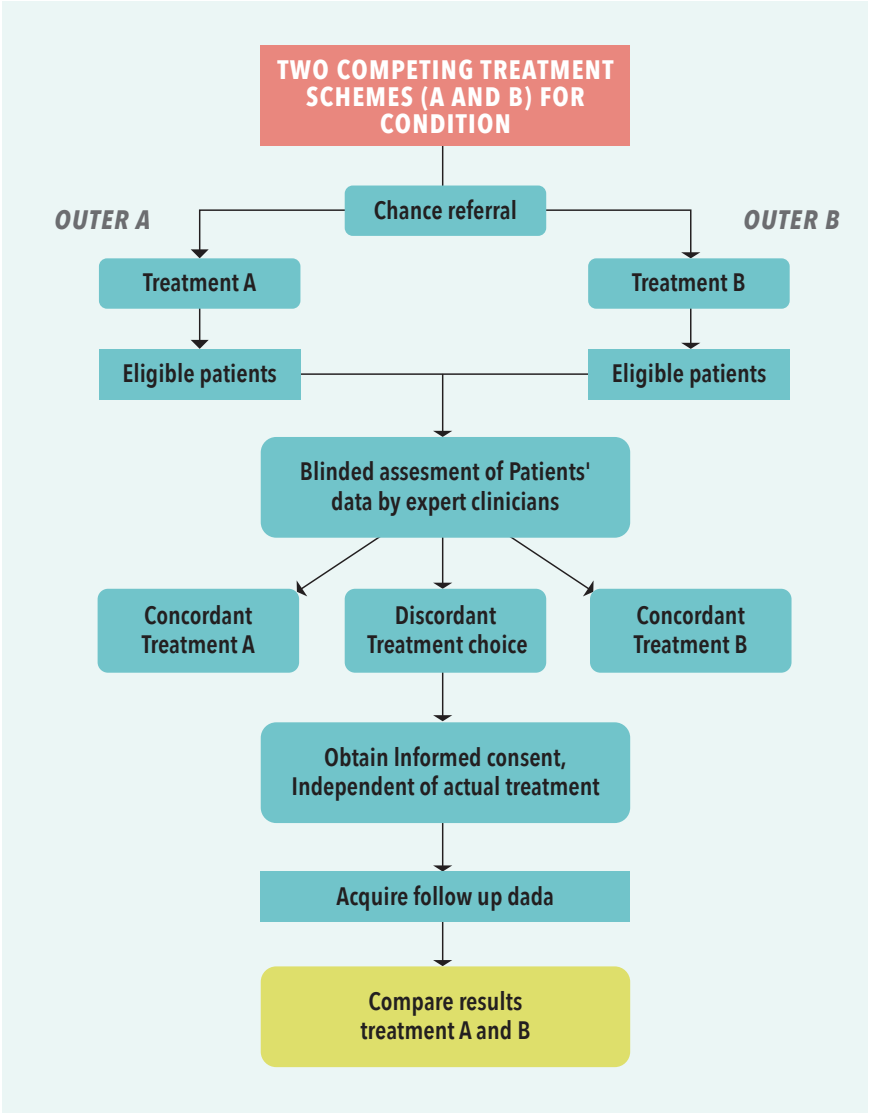

Fig. 1. Flow-diagram patient inclusion.

other hospital consisted of bed rest for several weeks (4– 6 wk depending on the type of fracture) combined with a plaster orthosis afterward for 6 weeks.

Table 1 shows the baseline characteristics and demographics of our study populations. The distribution of both is equal across the parallel cohorts enrolled in the study, that is, no differences in prognostic factors were observed.

General outcomes measurements include Short Form-36, Eq5D, Oswestry Disability Index, and Denis Pain and Work Scale. Specific outcomes include neurologic deterioration, reoperation rates, complication rates, and instances of change in treatment regimens. Final results of our study will be published in the near future.

	Nonoperative Total N=95 (N=50%)	Operative Total N=95 (N=50%)	p Value
Patients			
Male	50 (44)	65 (59)	.19
Female	45 (56)	31 (41)	.11
Age (y)			
Mean	38.5 (18-84)	36.6 (18-79)	.48
Male	38.9 (19–81)	36.6 (18–79)	.37
Female	38.1 (18–84)	38.0 (19–77)	1.0
Polytrauma	32 (34)	35 (37)	.6
No. of fractures	131	124	.4
Trauma cause			
Fall	46 (48)	42 (44)	.7
Traffic	27 (28)	26 (27)	.9
Jump	13 (14)	19 (20)	.3
Other	9 (10)	8 (9)	.8
ASIA scale			
A	3 (3)	8 (9)	.2
B	4 (4)	5 (5)	.7
C	2 (2)	6 (6)	.2
D	14 (15)	18 (19)	.5
E	71 (76)	58 (61)	.3
Unknown	1 (1)		

Table 1. Clinical and demographic data

DISCUSSION

Regardless of the difficulties with conducting RCTs, the proper treatment of thoracolumbar fractures is still a matter of debate, deserving attention from the spine research community. We must be prepared to accept guidance from alternative methods when the gold standard is not achievable. Clearly, efforts should also be directed at improving and inventing alternative methods.

The proposed methodology is a cohort study with head-to-head comparison of two treatment regimens in an expertise-based trial (EBT) fashion. The US Preventative Services Task Force stated, ''The potential for bias is much greater in cohort and case-control studies than in RCTs so recommendations from overviews combining observational studies will be much weaker''.[27] Observational studies with significant bias systematically overestimate or underestimate the magnitude of treatment effect. In 2000, Concato et al. challenged the current consensus in clinical research about the superiority of RCTs and argued that data from weaker forms of observational studies are mistakenly used to criticize all forms. The ''weaker forms'' apparently involved historical, unblinded cohorts or those without randomly assigned control subjects. After an evaluation of 99 reports in major medical journals covering five clinical topics, Concato et al. suggested that observational studies designed with the principles of strict exclusion criteria and control of prognostic factors can demonstrate magnitudes of treatment effects similar to RCTs.[28]

Horwitz et al. described the design of a restricted cohort study, which incorporates many design principles and cohort assembly procedures of RCTs. The use of exclusion criteria, standardized disease-defining criteria, a specific zero time for determining patient eligibility, adjustment for prognostic risk, and intention-to-treat statistical strategies enabled the design of a cohort study that estimated a nearly identical treatment difference to a ''gold standard'' multicenter, randomized, double-blind clinical trial.[29]

The strength of the ''golden standard'' is predicated on the a priori randomization procedure to ensuring that known and unknown confounders are equally distributed between the study groups in an RCT. In the proposed methodology, there is no a priori randomization procedure, rather there is a post hoc presumption that the patients are representative of the same population and the hospital of presentation is selected through a random process, that is, the location at which the incident occurred. This is not true randomization, leaving a susceptibility to confounding variables, which will need to be accounted for in the final analysis. One must consider the proximity of certain hospitals to higher-risk populations and the logistical factors of the Emergency Medical Services (EMS) system. To address the lack of true a priori randomization, there must be tight adherence to the principles of a restricted cohort study to ensure that both treatment groups are essentially equal in their prognosis. That is, both groups have a reasonably similar chance at having a favorable (or unfavorable) result of the therapeutic

intervention. On the basis of the principles of restricted cohort studies, we used strict exclusion criteria, a zero time for eligibility, standardized injury definitions, and adjustment for prognostic risk factors.

Eligibility criteria are critical to limit bias. The target group should be restricted to only those patients who would be eligible to receive either treatment modality.[29] Otherwise, the treatment effects observed can be distorted by selection bias, because patients may have received treatment partially indicated or even contra-indicated. In our methodology, patients are selected by a panel of experts from both ''schools'' of treatment modalities blinded to the treatment actually received. Patients were selected from a comprehensive cohort of patients with a specified diagnosis. Inclusion only occurs when a clear and genuine uncertainty exists regarding the appropriate treatment among the expert panel of surgeons treating spinal fractures,that is,the definition of clinical equipoise. In this manner, selection bias is addressed. No patient should receive a contrain dicated treatment, and no patient should be included with an injury that has characteristics lending itself to a (preferred) modality already agreed upon.

Because of the observational nature of the study, the most hazardous threat to the validity of the study remains confounding. RCTs prospectively identify, measure, and record known and measurable prognostic variables. To demonstrate homogeneity and avoid susceptibility bias in a retrospective cohort study, it is important to identify measurable prognostic factors and possible confounding variables within the available data that may affect the outcomes. We attempted to stratify the injury severity by factors including the presence of polytrauma, cause of trauma, number of fractures, fracture classifications, and the American Spinal Injury Association scale. All prognostic factors should be subjected to multivariate analysis. Statistical adjustments can be made for any prognostic factor found to differ between groups. Clearly, the baseline characteristics of the cases enrolled need to be scrutinized and possible differences across the treatment strategies compared should be adjusted for in the analysis phase. In our specific study, gender appeared to show a slight imbalance, which we will adjust for when presenting the final results.

An unavoidable limitation of the design is the inability to ensure that unknown and unmeasured prognostic factors are equally distributed between the treatment groups, because of the lack of a priori randomization. The allocation of treatment is dictated by factors affecting the distribution of patients by ambulance. It is reasonable to assume, however, that an unknown confounder will likely be revealed if it significantly affects the results of the study. Newly revealed confounders, such as socioeconomic factors or comorbidities not already excluded, can be accounted for in multivariate analysis of final outcomes.

Rosenbaum and Rubin introduced the propensity scores method in 1983, a means for ensuring equal distribution of known confounders across groups in observational studies.[30] The propensity score represents the probability of receiving treatment A rather than B for patients in a non- randomized study, based on observed baseline characteristics. This method is used increasingly in surgical

research, particularly in cardiothoracic surgery.[31] In a recent study by Orosz et al. involving 1,200 hip fractures in an emergency setting, propensity scores were shown to be of restricted use in studies with limited samples sizes.[32] Luellen et al. also mentioned that it is not clear how large a sample size should be to prevent imbalances in covariates when using propensity scores.[33] Thus, propensity scores may not solve the potential problem of confounding especially in our typical limited sample size study. However, the method might prove valid in future large-scale studies of similar design.

Another consideration in the design is the use of separate schools to administer the two different interventions, similar to an EBT. Devereux recently made the case for the need of more EBTs in surgery and in fact corroborates our suggestions. EBTs aim to eliminate the impact of differential procedure performance by randomizing patients to a surgeon rather than a procedure. Surgeons perform only their preferred procedure, thereby limiting the influence from preferences for one of the interventions. EBTs are applicable to any study involving the skill set of the clinician, especially when the clinician cannot be blinded to the intervention. Devereux points out that EBTs are more feasible than conventional randomization and more ethically sound because clinicians are performing the procedures in which they have established experience.[34]

In our design, all patients were treated by surgeons experienced in their particular school of expertise in a clinical setting optimized to the specific intervention. No ethical concerns should exist for patients receiving an inferior treatment, as clinical equipoise is demonstrated by the expert panel disagreement. The surgeons practice in a clinically realistic manner based on their established protocols. In many RCTs, treatments are standardized to rigid rules to ensure treatments are always uniform, though unresponsive to the changing clinical scenario. In the proposed design, well-established protocols are conducted, but some variation from the protocols must be expected and appropriately reported.

Freedman originally argued that equipoise (general uncertainty) of treatment preference must exist for the ethical randomization of patients to treatment arms in clinical studies[21], but much literature has discussed the complexity, merit, and limitations of equipoise in clinical trials.[35,36] Veatch recently claimed that the concept of equipoise is irrelevant, and proper informed consent is the guard against ethical dilemmas in clinical research.[37] This methodology simply uses clinical equipoise as a tool to narrow the spectrum of spine fracture cases to the heart of the real clinical dilemma. It is believed that the use of surgeon equipoise as an inclusion criterion can achieve a homogenous sample of the patient conditions that exist in a controversial area of clinical decision making, where the use of either treatment regimen can be clinically and ethically defended.

As outcome data continues to be collected in our study, we eagerly await the conclusion of follow-up and are excited at the prospect that this design may help answer the difficult question of appropriate treatment in spinal fractures. There are inherent drawbacks in the design because of its retrospective aspect, but the

unique method of patient selection should address the influence of selection bias and allow for a valid comparison of prospective outcomes. Although we can critically appraise evidence, we cannot be critical of it unless we can propose a solution that can practically realize a better solution. The proposed model, using equipoise as an inclusion criterion, represents an advance on what is currently achievable, given the precluding factors of RCTs. It is hoped that the proposed methodology can also be applied in other controversial fields where RCTs have been ineffective or infeasible.

REFERENCES

1 Angell M. The ethics of clinical research in the Third World. *N Engl J Med 1997;337: 847-9.*

2 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ 1996;312:1215-8.*

3 Chene G, Morlat P, Leport C, et al. Intention-to-treat vs. on-treatment analyses of clinical trial data: experience from a study of pyrimethamine in the primary prophylaxis of toxoplasmosis in HIV-infected patients. *ANRS 005/ACTG 154 Trial Group. Control Clin Trials 1998;19:233-48.*

4 McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ 2002;324:1448-51.*

5 Solomon MJ, McLeod RS. Should we be performing more randomized controlled trials evaluating surgical operations? *Surgery 1995;118:459-67.*

6 Laursen M, Eiskjaer SP, Christensen FB, Thomsen K, Bunger CE. Results after surgical treatment of unstable thoracolumbar fractures. *Ugeskr Laeger 1999;161:1910-4.*

7 Parker JW, Lane JR, Karaikovic EE, Gaines RW. Successful short segment instrumentation and fusion for thoracolumbar spine fractures: a consecutive 41/2-year series. *Spine 2000;25:1157-70.*

8 Resch H, Rabl M, Klampfer H, Ritter E, Povacz P. Surgical vs. conservative treatment of fractures of the thoracolumbar transition. *Unf=-allchirurg 2000;103:281-8. 980 A. Stadhouder et al. / The Spine Journal 8 (2008) 975-981*

9 Shen WJ, Liu TJ, Shen YS. Nonoperative treatment versus posterior fixation for thoracolumbar junction burst fractures without neurologic deficit. *Spine 2001;26:1038-45.*

10 Aligizakis A, Katonis P, Stergiopoulos K, Galanakis I, Karabekios S, Hadjipavlou A. Functional outcome of burst fractures of the thoracolumbar spine managed non-operatively, with early ambulation, evaluated using the load sharing classification. *Acta Orthop Belg 2002;68:279-87.*

11 Speth MJ, Öner FC, Kadic MA, de Klerk LW, Verbout AJ. Recurrent kyphosis after posterior stabilization of thoracolumbar fractures. 24 cases treated with a Dick internal fixator followed for 1.5-4 years. *Acta Orthop Scand 1995;66:406-10.*

12 D'Ariano GD, Errico TD. Posttraumatic kyphosis of the thoracic and lumbar spine. *Curr Opin Orthop 1995;6:95-103.*

13 Knop C, Blauth M, Buhren V, et al. Surgical treatment of injuries of the thoraco-lumbar transition-3: follow-up examination. Results of a prospective multi-center study by the "Spinal" Study Group of the German Society of Trauma Surgery. *Unfallchirurg 2001;104:583-600.*

14 McCormack T, Karaikovic E, Gaines RW. The load sharing classification of spine fractures. *Spine 1994;19:1741-4.*

15 McLain RF, Sparling E, Benson DR. Early failure of short-segment pedicle instrumentation for thoracolumbar fractures. A preliminary report. *J Bone Joint Surg Am 1993;75:162–7.*

16 Shen WJ, Shen YS. Nonsurgical treatment of three-column thoracolumbar junction burst fractures without neurologic deficit. *Spine 1999;24:412–5.*

17 Stromsoe K, Hem ES, Aunan E. Unstable vertebral fractures in the lower third of the spine treated with closed reduction and transpedicular posterior fixation: a retrospective analysis of 82 fractures in 78 patients. *Eur Spine J 1997;6:239–44.*

1**8 Gertzbein SD. Scoliosis Research Society.** Multicenter spine fracture study. *Spine 1992;17:528–40.*

19 Wood K, Butterman G, Mehbod A, Garvey T, Jhanjee R, Sechriest V. Operative compared with nonoperative treatment of a thoracolumbar burst fracture without neurological deficit: a prospective, randomized study. *J Bone Joint Surg Am 2003;85-A(5):773–81.*

20 Verlaan JJ, Öner FC. Operative compared with nonoperative treatment of a thoracolumbar burst fracture without neurological deficit. *J Bone Joint Surg Am 2004;86-A:649–50.*

21 Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med 1987;317: 141–5.*

22 Gartland JJ. Orthopaedic clinical research. Deficiencies in experimental design and determinations of outcome. *J Bone Joint Surg Am 1988;70:1357–64.*

23 Wehbe MA. The prospective, randomized, double-blind clinical trial in orthopaedic surgery. *J Bone Joint Surg Am 1998;80:1395.*

24 Doyal L. Informed consent in medical research. Journals should not publish research to which patients have not given fully informed consentdwith three exceptions. *BMJ 1997;314:1107–11.*

25 Localio AR, Belrin AJ, Ten Have TR, Kimmel SE. Adjustments for center in multi-center studies: an overview. *Ann Intern Med 2001;135:112–23.*

26 Solomon MJ, Laxamana A, Devore L, McLeod RS. Randomized controlled trials in surgery. *Surgery 1994;115:707–12.*

27 Houston TP, Elster AB, Davis RM, Deitchman SD. The U.S. Preventive Services Task Force Guide to clinical preventive services, second edition.*AMACouncilon-ScientificAffairs.AmJPrevMed1998;14:374–6.*

28 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med 2000;342:1887–92.*

29 Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med 1990;89:630–8.*

30 Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in observational studies for causal effects. *Biometrika 1983;41–55*.

31 Chukwuemeka A, Weisel A, Maganti M, et al. Renal dysfunction in high-risk patients after on-pump and off-pump coronary artery bypass surgery: a propensity score analysis. *Ann Thorac Surg 2005;80:2148–53*.

32 Orosz GM, Magaziner J, Hannan EL, et al. Association of timing of surgery for hip fracture and patient outcomes. *JAMA 2004;291: 1738–43*.

33 Luellen JK, Shadish WR, Clark MH. Propensity scores: an introduction and experimental test. *Eval Rev 2005;29:530–58*.

34 Devereaux PJ, Bhandari M, Clarke M, et al. Need for expertise based randomised controlled trials. *BMJ 2005;330:88*.

35 Chiong W. The real problem with equipoise. *Am J Bioeth 2006;6:37–47*.

36 Chard JA, Lilford RJ. The use of equipoise in clinical trials. *Soc Sci Med 1998;47:891–8*.

37 Veatch RM. The irrevelance of equipoise. *J Med Philos 2007;32: 167–83*.

6

Traumatic Thoracic and Lumbar Spinal Fractures: Operative or Nonoperative Treatment: Comparison of Two Treatment Strategies by Means of Surgeon Equipoise

Agnita Stadhouder
Erik Buskens
Luuk W. de Klerk
Jan A. Verhaar
Wouter A. Dhert
Abraham J. Verbout
Alex R. Vaccaro
F. Cumhur Öner

Spine 2008;33:1006-1017

ABSTRACT

Study Design. A center parallel cohort study with blinded inclusion based on clinical equipoise.

Objective. To compare outcomes of nonoperative and operative treatment strategies in terms of quality of life and neurologic and functional status.

Summary of Background Data. Despite a considerable body of literature, sound evidence regarding the optimal treatment for traumatic thoracic and lumbar spine fractures is lacking.

Methods. Medical records of patients hospitalized for traumatic spinal fractures between 1991 and 2002 were identified in 2 trauma centers in the same country with established and different treatment strategies. Eligibility was retrospectively assessed for each case by a panel of orthopaedic surgeons who were representative of the 2 medical centers, and who were blinded to the treatment actually administered. Patients were included in the study when there was disagreement on the suggested treatment method. Thus, 2 comparable groups were identified undergoing nonoperative or operative treatment. Outcome assessment and comparison across groups focused on quality of life, residual pain, neurologic recovery, and employment in the middle-long-term follow-up.

Results. Discordance in regards to choice of treatment was identified in 190 (95 treated nonoperative, 95 operative) of 636 potentially eligible patients. Patients were comparable regarding baseline characteristics, except for a somewhat higher proportion of males and neurologic impairment in the operative group. Seventeen percent of the nonoperative and 21% of the operative group developed complications and 3 patients displayed neurologic deterioration for which a treatment change was considered necessary. Follow-up was complete in 79%; mean follow-up time was 6.2 years with a minimum of 2 years. Pain scores, disability indexes, and general health outcome were comparable at follow-up. Compared with matched population norms, outcomes were poorer regardless of treatment method. Neurologic recovery was better in the operative group, but this difference did not reach statistical significance. Multivariate regression analyses revealed that female gender and neurologic impairment were independent predictors of poor functional outcome. Eighty-eight and 83% of the nonoperatively and operatively treated patients were employed at some point after a rehabilitation period.

Conclusion. Overall outcome of nonoperative and operative treatment in middle-long-term follow up is comparable, although there seems to be a difference in neurologic recovery patterns. Studies on the cost-effectiveness of treatment options and the patterns of recovery within 2 years after injury would assist in guideline development and stimulate interest for future research.

INTRODUCTION

Comparative studies on treatment of traumatic thoracic and lumbar fractures are generally retrospective in nature, with some prospective series and 2 recent small-scale randomized clinical trials.[1-8] The major problem with nonrandomized series in spinal fractures is that patient selection bias often occurs with more serious injuries generally treated operatively (OP). OP management outcomes are then frequently compared with nonoperative (NON-OP) management methods in retrospective series often reporting biased results reflecting surgeon preference. Because of this apparent selection bias, valid conclusions cannot be attained and results from former nonrandomized studies should be interpreted cautiously.

Research on the topic of spinal trauma is fraught with several difficulties. Variations in classification systems and treatment methods as well as the diversity of the trauma populations make it extremely difficult to compare treatment outcomes.

Since the 1980s, operative treatment has gradually gained popularity because of its presumed clinical benefit and relatively simple and effective fixation techniques.[9-11] Although these commonly used posterior reduction and fixation techniques are associated with complication rates generally considered acceptable and with high patient satisfaction, it is unknown whether nonoperative management would have possibly yielded similar or better outcomes. Lack of evidence as to optimal treatment method is attributable to the lack of controlled randomized trials in the general trauma population. Because nonoperative treatment alternatives comprise a range of different treatment methods, not always accurately described, valid comparisons to surgical intervention is difficult and practically nonexisting.

Another source of confusion has been the classification of injuries, which has posed considerable difficulty because of the complexity of injury patterns and variations between centers on imaging selection.[12,13] Several classification systems have been used during the last 50 years, which makes a comparison of reported results difficult. The AO classification system by Magerl is comprehensive in nature but inter- and intraobserver variability is considerable. The practical and even theoretical use of this system therefore remains limited.[12,14] The Spine Study Trauma Group introduced a new classification system based on morphology, neurologic impairment and integrity of the posterior ligamentous complex, which may minimize interobserver reliability because of its simplicity and ease of application.[15,16]

Finally, randomization in surgical- and emergency situations poses considerable practical difficulties. Two systematic reviews of the literature on nonoperative and operative treatment of spinal fractures were recently published. Both reviews concluded that there was a significant need for valid clinical research on this subject.[17,18] Only 2 randomized clinical studies comparing nonoperative and operative treatment were performed.[5,8] Both studies were restricted to neurologically intact patients. Polytrauma patients were not included and very small numbers of patients (23 *vs.* 24 and 16 *vs.* 18) could be recruited over a relatively long period

of time, thus rendering these papers limited in their applicability and validity. Not surprisingly, the conclusions of these 2 studies were completely discordant.

Thus, "alternative" study designs that avert the potential of selection bias have been proposed to obtain valid estimates of the magnitude of effects of treatment.[19,20] Expertise based trials, the use of propensity scores, case control studies, and (restricted) cohort designs may solve some of these problems.[19,21-22a]

Yet, today, the debate about optimal management of thoracic and lumbar fractures continues among spine surgeons. In a time of rapid technologic developments in spine surgery and increasing emphasis on the importance of evidence based medicine, a novel method of cohort comparison in the trauma population, *i.e.*, clinical surgeon equipoise, has been proposed. Surgeon equipoise is introduced as an inclusion criterion for a new blinded parallel group observational design (Stadhouder, *et al,* unpublished data).[23b] This has enabled the acquisition of balanced groups with unbiased evidence regarding the treatment outcomes of spine fractures. In this report, we present the results of a study using this method and discuss the merits of this approach and applicability in prospective research.

MATERIALS AND METHODS

The objective of this study was to compare differences in outcome between operative (OP) and nonoperative (NON-OP) treated patients with a traumatic thoracic or lumbar spine fracture with respect to neurologic (ASIA score) improvement, residual pain, patient-outcome (patient burden and quality of life as perceived by the patient), and functional (rehabilitation and final residual disability) outcome. We also investigated the work status of patients.

Details of this new methodology are published elsewhere (Stadhouder, *et al,* unpublished data)[23b]. In brief: all patients admitted with a traumatic thoracic or lumbar spine fracture from 1991 to 2002, and treated at 2 comparable university hospitals with historically different treatment strategies (OP *vs.* NON- OP) in the same country, were analyzed in terms of treatment outcomes. One of these centers has a long established treatment strategy of nonoperative care and performs surgery only in exceptional cases. The other center has a more aggressive approach and performs surgery in cases of fractures with neurologic deficits or those deemed mechanically "unstable." A search on diagnosis code(s) was performed, and patient charts and radiology files were extracted to obtain information on trauma mechanism, fracture type, and neurologic status. At trauma and follow-up times, treatment, hospital admittance, and complications were determined. The clinical and radiologic data of patients admitted in this period were blinded to the actual treatment they received and the clinical outcome. The 2 treatment teams representing either "school" independently assessed each case to make a decision for NON-OP or OP treatment as if patients were presenting to their hospital at that moment. The treatment protocols for fractures labeled as "unstable" in both centers were well defined and consistent during the study period. Center A NON-OP: bed rest

for several weeks (4–6 weeks depending on the type of fracture) combined with a plaster orthosis afterwards for 6 weeks. Center B OP: short segment posterior stabilization and fusion (PS), bridging 1 or 2 disc-spaces as needed, followed by a protective TLSO for 12 weeks or in multiple-vertebral fractures posterior long segment fixation. Patients with discordant treatment advice were identified and constituted the final study groups, which means that each of these patients would have a different treatment if they had been presented to the other center. After obtaining informed consent, questionnaires focusing on residual pain, disability, work history, duration of absence from work, and employability were sent out to the eligible patients.

To evaluate outcome after treatment, various pain, disability, and health score outcomes were used:

The visual analog score (VAS) is a well-known instrument to let patients score their daily pain. This instrument is validated and compared with other self-administered pain scales with good correlation.[23a] We used a 0 to 10 scale with 0 = no pain and 10 = unbearable pain.

The Denis Pain and work scale was used to assess pain and work score on a 1 to 5 ordinal scale (0 = no pain, 5 = constant or severe incapacitating pain, 0 = back to heavy labor, 5 = not able to perform any labor).[7]

The Oswestry disability index (ODI) is a functional, disease specific instrument on low back pain. It contains 10 questions on limitations of daily activities of living. The index (%) is calculated by adding up the points per question (0–5), and multiplying the score by 2. The index therefore ranges between 0 (best health state) to 100 (worst health state). The index was validated for the Dutch population.[24,25]

We used 2 questionnaires concerning health-related quality of life: the EQ5D, which is a standardized generic nondisease-specific instrument describing and evaluating health-related quality of life. It includes 5 dimensions of health, enabling a utility index score to be calculated, and it comprises a VAS EQ5D from 0 to 100 on current health status (0 = worst health state, 100 = best health state). The utility index score ranges from 1 (perfect health) to 0 (death), with the possibility of a negative score for a worse than death determination.[26]

The SF-36 is a widely used measure for general health status as defined by the WHO.[27] It contains 36 multiple choice questions on 8 subscales of health. The scores range from 0 to 100, with 100 denoting a perfect score. The SF-36 was also translated and validated for use for the Dutch population.[27,28]

Results were analyzed with SPSS 12.0 SAS Business Unit (2003). Univariate *t* tests and χ^2 tests for noncontinuous variables were used to compare baseline characteristics of the 2 treatments. Results were considered significant at a $P < 0.05$. Comparability between the 2 groups was thus evaluated. Further multivariate-linear and logistic regression analyses were performed on 5 parameters. For these parameters, treatment, gender, polytrauma patients, neurologic impairment at admission, and the integrity of the posterior ligamentous complex relevant effects on clinical outcomes were assessed. These were introduced in the regression

model to adjust for possible differences in baseline characteristics and to identify possible prognostic factors for better outcome in patients with >2 years follow-up after a thoracolumbar spinal fracture. Beta coefficients and odds ratios inclusive of 95% confidence intervals were presented to express the association between treatment and outcome, i.e., recovery after a spine fracture.

RESULTS

Six hundred and thirty-six patients were assessed in a blinded fashion by representative experts of the 2 treatment schools. Of these, 190 (30%) patients were included in the study group because of discordance in proposed treatment by the 2 experts. Figure 1 shows the assessment process of the 636 patients included in the study as well as the prospective data acquired. Baseline characteristics are

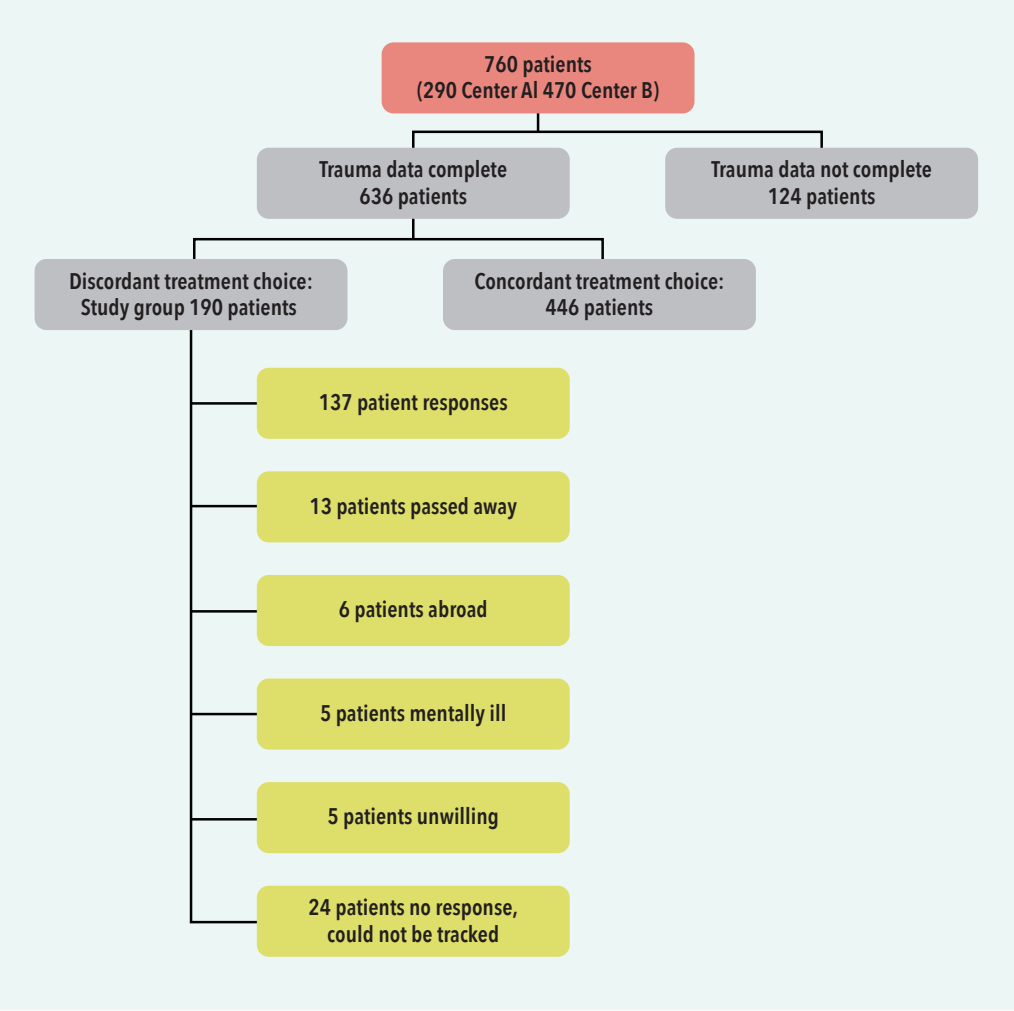

Fig 1. Assessment and follow-up response of patients with a traumatic thoracolumbar spine fracture.

shown in Table 1. OP and NON-OP groups were apparently comparable except for gender, with a significant predominance of male patients in the operative group (P = 0.009). Treatment location and treatment itself did not have a recognizable effect on ICU admittance. Out of the 190 patients, 143 had 1 fractured vertebra, 32 had 2 fractured vertebrae, 12 patients 3, and 3 patients 4 fractured vertebrae. There was an even distribution between NON-OP and OP treatment and fracture level. Sixty-eight percent of the patients fractured a vertebra at the thoracolumbar junction (T11–L1), 49% were treated NON-OP, and 51% OP. Trauma causes were comparable in the 2 groups and no differences in pretrauma psychiatric disorders were shown (12% of the NON-OP and 19% of the OP treated group had attempted to commit suicide P = 0.16). In terms of neurologic status, the ASIA scale appeared comparable between the 2 groups. However, when all patients with neurologic impairment (ASIA A-D) were combined and compared, the proportions across treatment groups demonstrated that significantly (P = 0.03) more patients with neurologic involvement received operative treatment when compared with nonoperative treatment among the study population. The percentage of males and females with neurologic impairment was not significantly different (Figure 2).

The mean admission time was comparable between treatments (NON-OP: 20.3 days, SD: 21; OP: 18.3 days, SD: 12). When patients admitted to the ICU were excluded, repeated analysis for NON-OP treated showed a mean admission time of 18.9 days (SD: 20.2) and for OP treated 16.5 days (SD: 11.3) (mean difference, 2.3; confidence interval, −2.6 to 7.4). The mean ICU admittance was also comparable between groups.

The majority (86%) of NON-OP treatment consisted of bed-rest followed by a plaster orthosis and mobilization, with or without an attempt for closed fracture reduction.

The vast majority of OP treated patients underwent posterior short segment fixation (76%) except in cases of multiple fractures where longer segment internal fixation was chosen (19%). Three patients, initially treated NON-OP, deteriorated neurologically and were treated OP (1x A, 2x PS). Two other patients were operated on because of progressive kyphosis at 4 months (A) and 2 years (PS) after the initial trauma. One patient died of concomitant cerebral injury within a week after admission.

The AO classification of the most severely fractured vertebra demonstrated a variety in injury patterns ranging from A1.2 to C3.2 (Figure 3). Of the 18 patients with A1 and A2 fractures, 4 had neurologic impairment, 4 had multiple fractures, and the remaining had a localized kyphosis angle greater than 30 degrees or a substantial vertebral body collapse. Because magnetic resonance imaging was not used in all cases, some injuries may have been unrecognized distraction injuries (type-B). When type A, B, and C injury patterns were compared, there were no significant differences between treatments (P = 0.3, 0.2, and 0.7). Twenty-five patients in the NON-OP group and 17 in the OP group had A3.1 fractures (P = 0.22; confidence interval, −0.22 to 1.0).

	NON-OP Total N = 95 (50%)	OP Total N = 95 (50%)	P
Patients			
Male	50(44)	64 (56)	0.19
Female	38.5 (18-84)	37.1 (18-79)	0.48
Age (y)			
Mean	8	15	0.08
Male	131	124	0.4
Female			
Polytrauma	46 (48)	42 (44)	0.7
No. of fractures	27 (28)	26 (27)	0.9
Trauma cause			
Fall	9 (10)	8 (9)	0.8
Traffic			
Jump	3 (3)	8 (9)	0.2
Other	4 (4)	5 (5)	0.7
ASIA scale			
A	14 (15)	18 (19)	0.5
B	71 (76)	58 (61)	0.3
C	1 (1)		
D	14 (15)	18 (19)	0.03
E	71 (76)	58 (61)	0.4
Unknown	1 (1)		0.5
A-D combined	23 (25)	37 (39)	0.4
Mean admittance (d)	20.3 range 1-109	18.3 range 1-63	0.5
ICU admittance (d)	14.4 SD 10	11.1 SD 11	
Treatment			
PS		73 (77)	
PL		18 (19)	
A†		1 (1)	
PS + BAER*		2 (2)	
2x PS		1 (1)	
Bed rest	7 (7)		
Bed rest and plaster	44 (47)		
Brace	1 (1)		
Plaster	38 (40)		
Change treatment	4 (4)		
Deceased admittance	1 (1)		
Hospital			
A	50 (88)	7 (12)	0.0003
B	45 (34)	88 (66)	

*A indicates anterior fixation; †PS + BAER, posterior short segment fixation; BAER, Balloon-assisted endplate reduction.

Table 1. Baseline Characteristics

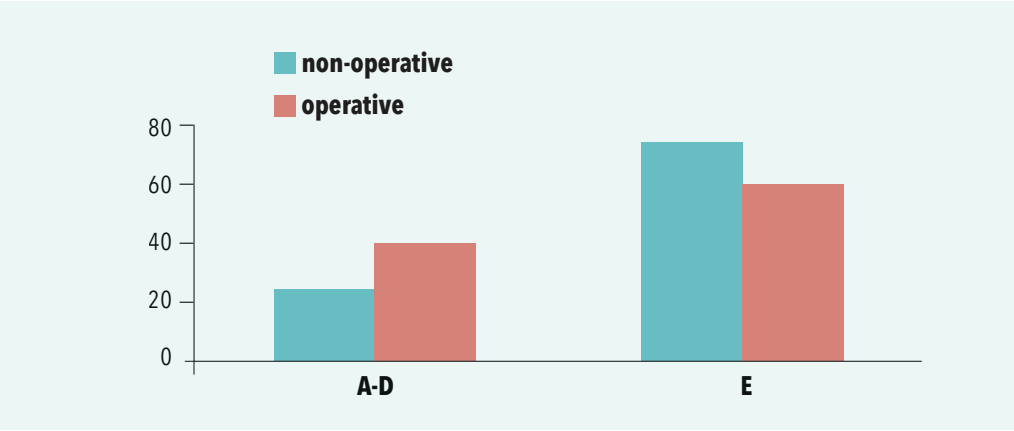

Figure 2. ASIA scale combined (%) at admittance.

There was an expected and significant difference in patient populations in both hospitals and treatments; in center A, the vast majority was treated NON-OP, and in center B, OP.

Despite the fact that NON-OP treated patients were mostly bedridden, the percentage of patients that returned home after admittance (67% and 63% NON-OP and OP respectively) was comparable (P = 0.7). Other discharge possibilities were rehabilitation centers, psychiatric wards, other hospitals, or nursing-homes. Complications (defined as pathologic processes or events occurring during a disease that are not essential parts of the disease; that may result from the disease or from independent causes) were recorded from the medical records. Sixteen (17%) NON-OP treated patients had a complication during their treatment. Three patients, because of neurologic deterioration, were switched to OP treatment, 6 patients developed a urinary tract infection for which antibiotics were prescribed, 2 patients developed delirium, 1 had a deep venous thrombosis, 1 patient developed a pressure ulcer under their plaster orthosis, 1 developed a neuropathy of the nervous peroneus, and 1 had an allergic reaction to their plaster orthosis. One patient suffered neurologic deterioration but did not worsen beyond an ASIA D category, and as a result NON-OP treatment was continued.

Twenty patients (21%) in the OP group experienced complications 2 patients developed a paralytic ileus for which conservative treatment was applied, 2 patients experienced bladder retention,1 patient an intravenous line sepsis, 1 a urinary tract infection, 1 patient had multiple pneumonias, and 1 had pulmonary atelectasis. Seven patients developed a complication directly related to OP treatment: 1 patient had a superficial infection and 1 patient developed a draining fistula for which eventual implant removal 6 months after surgery was undertaken. Four patients experienced problems related to spinal internal fixation: 3 patients developed a deep infection for which multiple reoperations were necessary and intravenous antibiotics were applied. In 1 of these patients, the instrumentation was removed 4 weeks after implantation because of a deep infection. Finally, 1 patient

developed an infection of the iliac crest, which was treated by surgical debridement and intravenous antibiotics. Excluding polytrauma patients and patients admitted to the ICU, complication percentages of 8.9% and 8.3% were noted respectively (P = 1.0) for NON-OP and OP groups. The complication rate of patients with neurologic impairment was 30% compared with 15% in the non-neurologic group. This difference was significant with an even distribution in NON-OP/OP treated groups. When comparing treatment of polytrauma patients admitted to the ICU, no difference in complication percentage was noted between treatments.

The complication rate in patients where the implants were operatively removed was 7%. In four cases, there was a large postoperative hematoma for which 1 patient received a blood transfusion, and 2 had an incisional drainage. Two patients had a superficial wound infection treated with antibiotics and 1 patient experienced decreased sensation on the lateral side of the knee after surgery. Two patients went on to a non- or delayed fusion, which was managed with repeat bone grafting.

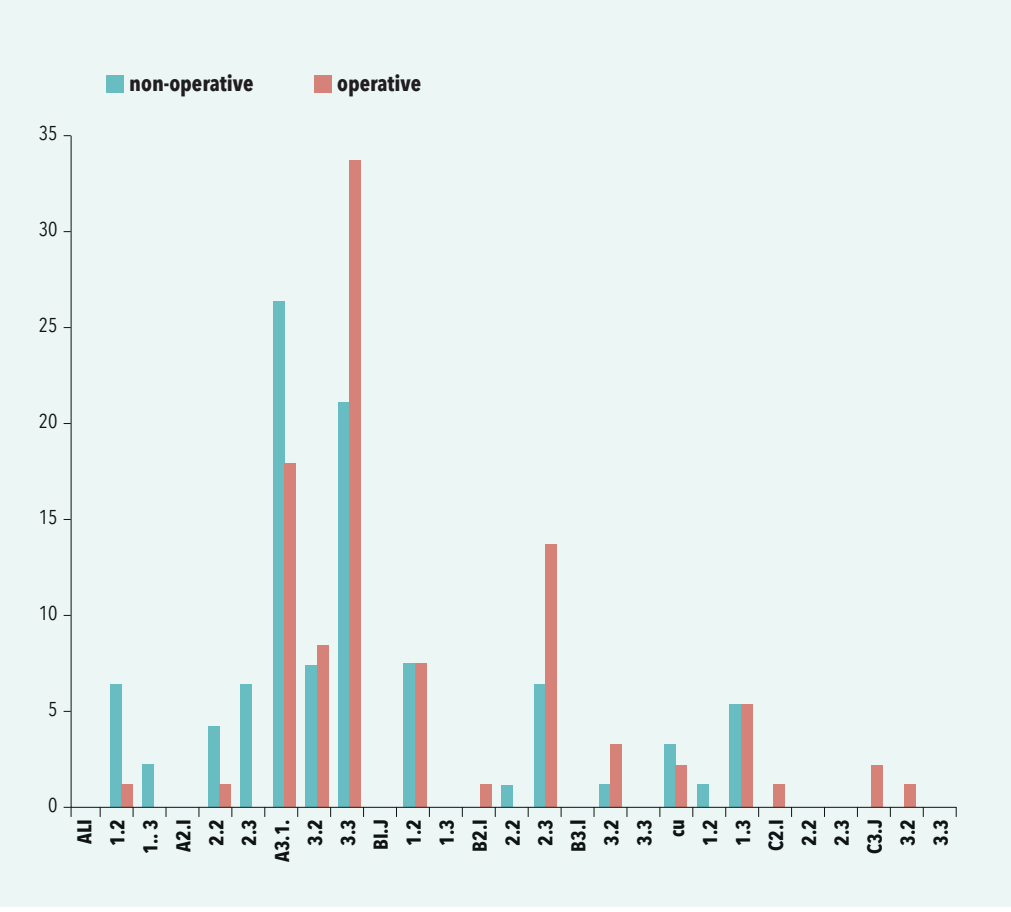

Fig 3. Treatment (%) and AO classification of most severe fractured vertebra.

Follow-up

One hundred and thirty-seven patients were available for long-term follow-up. Thirteen patients had passed away accounting for a follow-up percentage of 78.9% of which 50% were in the OP treatment group (Figure 1). The mean follow-up period for the total population was 6.2 years with a minimum of 2 and a maximum of 12 years.

Clinical Outcome

According to the intention to treat principle, the 5 patients with a change of treatment were analyzed in the original NON-OP group. Three were operated within 24 hours and 2 patients respectively after 4 months and 2 years.

Pain

The mean VAS score at the last follow-up was 2.5 for NON-OP patients (SD 2.8) and 2.5 for OP treated patients (SD: 2.7) (dif, 0.018; −0.9 to 0.95). The Denis pain-scale comparing treatments showed no significant difference between NON-OP and OP treated patients (P = 0.37) with scales 4 and 5 combined.

Disability

The mean ODI for NON-OP treated patients was 16.3 (SD: 19.9) and 15.7 for OP treated patients (SD: 17.7) (dif, 5.2; −5.7 to 7.0). Some authors have used a cut-off point of 40 to differentiate between moderate and severe disability.[29,30] Using this determination, sixteen patients (8.4%) had an ODI of 40 or more, with a distribution of 9 NON-OP and 7 OP treated (not significant).

Health-Related Quality of Life

The mean EQ5Dindex was 0.80 (SD: 0.24) and 0.74 (SD: 0.3) for NON-OP and OP treated patients (dif, 0.059; −0.03 to 0.16), the VAS EQ5D 76.5 (SD: 20.2) and 73.0 (SD: 19.6) (dif, 3.5; −3.3 to 10.3) respectively (Figures 4 and 5). The SF-36 mean values for physical function (NON-OP 69.2 SD 30.5/OP 63.9 SD 32.0) (dif, 5.3; −5.4 to 15.9), role limitations due to physical problems (NON-OP 72.3 SD 37.8/OP 65.2 SD 40.9) (dif, 7.1; −6.4 to 20.6), bodily pain (NON-OP 70.0 SD 25.4/OP 66.4 SD 25.1) (dif, 3.6; −5.0 to 12.1), general health (NON-OP 66.0 SD 26.5/OP 62.8 SD 25.2) (dif, 3.2; −5.6 to 12.0), vitality (NON-OP 67.1 SD 21.2/OP 63.9 SD 21.0) (dif, 3.1; −4.1 to 10.3), social functioning (NON-OP 80.4 SD 22.7/OP 79.4 SD 24.2) (dif, 4.0; −7.0 to 8.9), role limitations due to emotional problems (NON-OP 83.3 SD 32.3/OP 76.8 SD39.3) (dif, 6.5; −5.7 to 18.7), mental health (NON-OP 75.8 SD 18.2/OP 73.1 SD 19.9) (dif, 2.7; −3.8 to 9.2) were all comparable across the 2 treatments. Overall, patients with complications scored worse on the clinical outcome scales, but there was no difference in outcome between NON-OP and OP treated patients in this group.

The 2 patients with a change of treatment (NON-OP to OP) at long-term did worse on the residual pain, disability, and quality of life scales than the average NON-OP or OP population.

Concerning the neurologic status at 1 year and at longest follow up, there was in general a better recovery pattern for the OP group (Tables 2 and 3). But after combining all patients who experienced neurologic improvement, proportion scores for the number of patients with neurologic impairment, and therefore the ability to improve, showed (neurologic improvement prepost treatment CI −0.06 to 0.37; pretreatment-1-year follow-up CI −0.14 to 0.40; pretreatment–long-term follow-up CI −0.15 to 0.38) no significant differences.

Further, univariate analyses was performed for all clinical outcome scores and 5 possible clinically relevant prognostic parameters such as treatment, gender, polytrauma, neurologic impairment, and injury to the PLC. In all clinical outcome scores, neurologic impairment at admission had a significant influence; women had significantly lower scores on the SF role physical function, vitality, and mental health. Other parameters did not have a significant effect.

Multivariate analyses (Table 4), which correct for various parameters besides treatment, showed that gender (women had worse outcome than men) and neurologic impairment (neurologically impaired patients at trauma-moment) had a significant impact on clinical outcome. Treatment, polytrauma status, and possible injury to the PLC did not. Analyses of patients who were neurologically intact at follow-up did not demonstrate significant differences in outcome measures and treatment.

In both treatment groups, 73% of the patients were employed before the traumatic event. Thirty and 28% performed heavy labor, 17% and 22% moderate labor, and 31 and 27% had a desk-job (NON-OP *vs.* OP respectively).

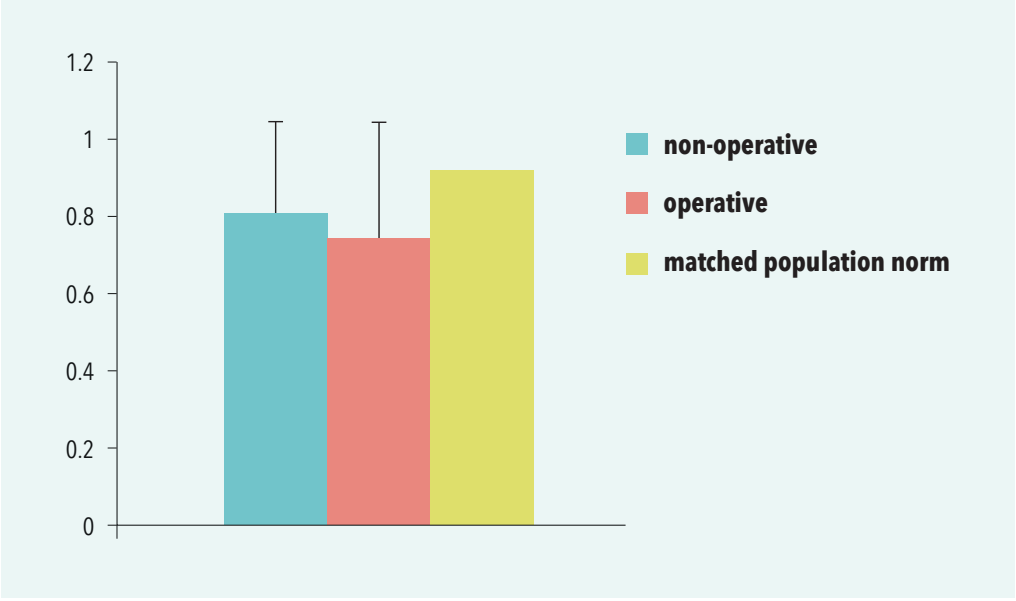

Fig 4. Mean EQ5Dindex and treatment at follow up

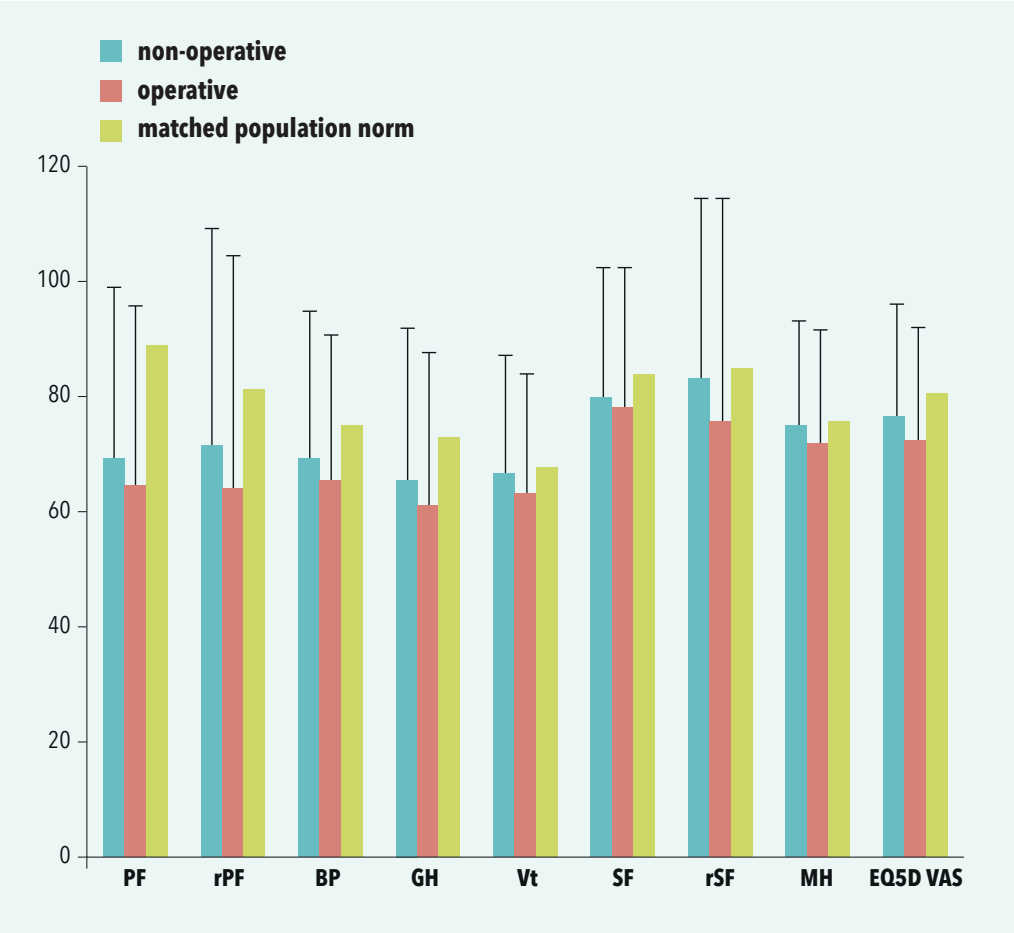

Fig 5. Mean SF-36-scores, $EQ5D_{vas}$ and treatment at follow-up

Five and 7% already received workers' compensation before the traumatic event. The remaining patients (17% and 16%) were either unemployed, housewives, students, or retired. Of the NON-OP treated 67% returned to their former job (34% heavy labor, 18% moderate labor, and 36% desk-job), of the OP this percentage was 52% (33% heavy labor, 24% moderate, and 30% desk-job). Thirty-five percent of the NON-OP treated patients received some form of workers' compensation after the traumatic event, for OP treated patients this was 39%. Overall, 88% of the NON-OP treated and 83% of the OP treated performed some form of labor (P = 0.59). The Denis work scale showed no significant differences between treatments (P = 0.42).

Seventy-three percent of the NON-OP treated patients with neurologic impairment posttrauma were employed, for the OP treated this was 75%.

Thirteen (7%) patients were deceased at follow-up, which is a relatively high percentage in this population with a median age of 36 years. Investigation did not show any relation with treatment performed: 2 patients died of complications at

admission (pneumonia and cerebral damage), 1 elderly patient died 5 days after discharge of unknown causes (NON-OP treatment), 3 patients committed suicide (pre-existent psychiatric disorder), and 7 patients died of other unrelated causes years after the traumatic event.

DISCUSSION

This study has shown that a best effort comparison using clinical equipoise can be used to validly compare treatment outcomes while at the same time respecting the treatment wishes of the treating surgeon. Although this study design is meth-

		A	B	C	D	E
OP	A	6	2			
NON-OP		3				
OP	B		2	1	2	
NON-OP			2		2	
OP	C				3	3
NON-OP				1		1
OP	D				7	11
NON-OP					5	7

Table 2. Neurologic Improvement Pretrauma and 1-Year Follow-up (ASIA Scale) and Treatment

		A	B	C	D	E
OP	A	6	2			
NON-OP		3				
OP	B		2	1	2	
NON-OP			2		2	
OP	C				2	3
NON-OP				1		1
OP	D				7	11
NON-OP					4	8

Table 3. Neurologic Improvement Pretrauma and Final Follow-up (ASIA Scale) and Treatment

Table 4. Multivariate Analyses Possible Prognostic Parameters and Clinical Outcome

	VAS		ODI		EQ5D		EQ5DIndex	
	β	CI	β	CI	β	CI	β	CI
Parameters								
Polytrauma	-0.78	-0.99 to 0.83	-1.8	-8.1 to 4.4	-3.6	-10.6 to 3.4	-0.04	-1.3 to 0.5
Polytrauma	-1.3	-2.3 to -0.39*	-8.2	-14.8 to -1.7*	8.2	0.87 to 15.6*	0.09	-0.004 to 0.19
Polytrauma	0.12	-0.85 to 1.1	1.6	-5.1 to 8.3	-2.2	-9.6 to 5.2	-0.03	-0.13 to 0.065
Neurologic imp	2.3	1.3 to 3.3	14.7	7.8 to 21.6*	-8.3	-16.1 to 0.54*	0.24	-0.34 to -0.14*
Post disruption	-0.45	-1.4 to 0.54	-0.17	-7.0 to 6.7	-0.44	-8.1 to 7.2	0.052	-0.47 to 0.15
	SF P. F		SF role P.F		SF B.P.		SF G.H.	
	β	CI	β	CI	β	CI	β	CI
Multivariate								
Treatment	-2.0	-11.8 to 7.8	-7.1	-20.5 to 6.4	-3.4	-12.2 to 5.3	-2.6	-11.8 to 6.5
Gender	13.9	3.7 to 24.1*	19.7	5.5 to 33.8*	10.5	1.4 to 19.7*	6.4	-3.2 to 16.0
Polytrauma	-7.7	-18.1 to 2.7	-2.8	-17.2 to 11.6	-2.8	-12.1 to 6.5	1.2	-8.6 to 11.0
Neurologic imp	-32.9	-43.7 to -22.1*	-22.2	-37.2 to -7.3*	-11.8	-21.4 to -2.1*	-13.0	-23.1 to -2.8*
Post disruption	-1.4	-12.1 to 9.3	-2.6	-17.3 to 12.2	0.89	-8.7 to 10.4	-5.0	-15.0 to 5.0
	SF Vit		SF S.F.		SF rS.F.		SF MH	
	β	CI	β	CI	β	CI	β	CI
Multivariate								
Treatment	-4.2	-11.3 to 2.8	-1.1	-9.1 to 6.9	-6.3	-18.7 to 6.0	-4.3	-10.7 to 2.1
Gender	13.1	5.7 to 20.5*	9.7	1.3 to 18.1*	14.2	1.3 to 27.2*	12.5	5.7 to 19.2*
Polytrauma	-0.61	-8.1 to 6.9	-1.4	-10.0 to 7.2	-6.8	-20.0 to 6.4	-3.1	-9.9 to 3.7
Neurologic imp	-12.3	-20.2 to -4.5*	-13.5	-22.4 to -4.6*	-17.2	-31.0 to -3.5*	-6.9	-14.1 to 0.20*
Post disruption	6.2	-1.5 to 13.9	4.0	-4.7 to 12.8	5.6	-8.0 to 19.2	6.6	-0.48 to 13.6*

*Indicates significant.

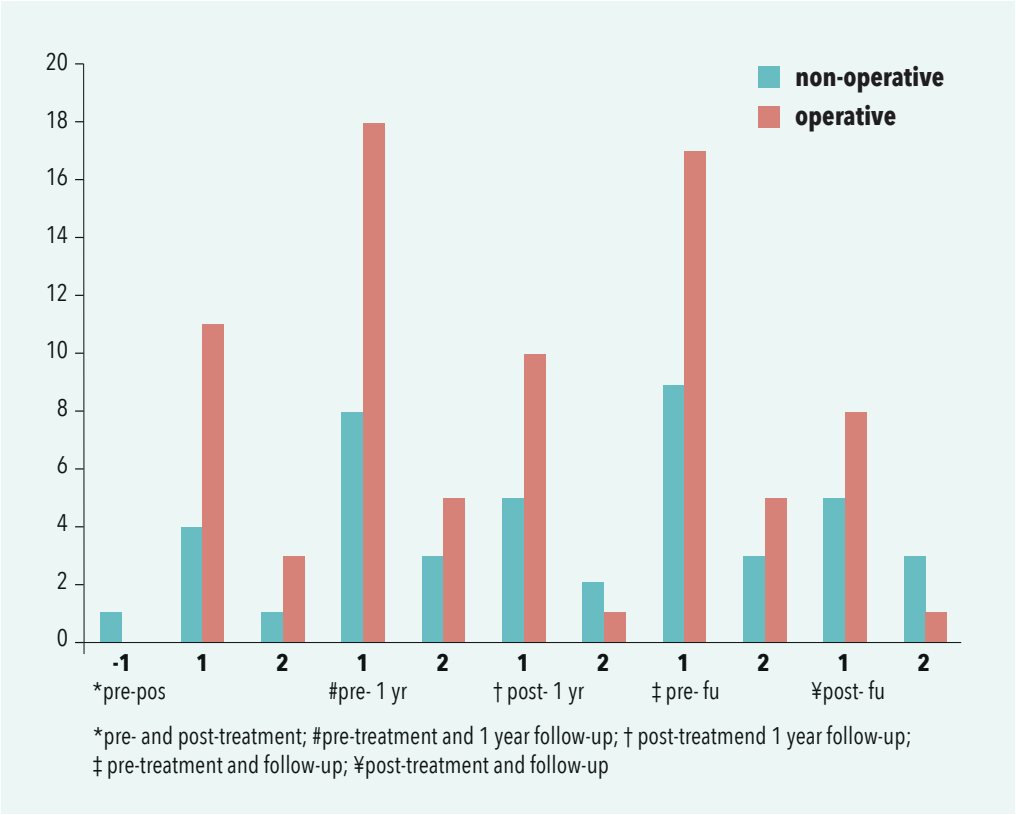

Fig 6. Neurological status (deterioration or improvement) as ASIA scale-1, 1 and 2 steps (% patients) and treatment.

odologically new, we are convinced that the conclusions are valid because the 2 groups identified were largely comparable and every patient received the treatment considered optimal by his or her treating surgeon. From the 636 patients identified, 190 patients would have been eligible for inclusion in a prospective randomized trial if that study design was so chosen. This agrees with our estimations that in one-third of the spine fracture population there is genuine disagreement on the optimum treatment: NON-OP or OP. We had a unique opportunity to create 2 historically comparable groups of treatment populations based on historical treatment differences and clinical equipoise regarding the best treatment.

The retrospective nature of the design inevitably leads to unavoidable missing data. It took considerable effort to obtain the relevant trauma-data, i.e., gathering trauma radiographs from the hospital of patient origin before transfer and collecting complete data from patient charts retrospectively. Considering the relative success of this method, we suggest that it can also be used for prospective studies or in databases where data gathering is conducted prospectively. The final follow-up percentage of 79 is in our opinion satisfactory for a study which covers a time span of 10 years with minimum of 2-year follow-up.

There is a discrepancy in baseline characteristics regarding gender with a high

male/female ratio in both groups, more pronounced in the OP group. From large trauma series, this gender effect is well known varying from a two- to four-fold male–female ratio.[1,8,31–34]

One third of patients were classified as polytrauma patients, comparable with other large spine trauma series.[1] Initially, 60 patients were neurologically impaired, with a significant higher ratio of impaired patients in the OP group. This illustrates the difficulties one can expect when a retrospective study of this nature is performed. Most spinal surgeons nowadays are reluctant to treat neurologically impaired patients NON-OP, although various authors have demonstrated that recovery in NON-OP treatment may be considerable, too.[1,35–41]

The mean admittance period did not differ between the 2 treatment regimens, even when excluding patients admitted to the ICU. This is remarkable since NON-OP treatment (usually with some period of bed rest) is predestined to a longer admittance time. A possible explanation for this may be the existence of an adequate home care system and the relatively short distances to hospitals in our country. Further, because treatment in center A is historically NON-OP, it has well-defined and optimized routine admission and treatment protocols and the necessary infrastructure and expertise for this treatment, which provides for a shorter hospital stay.

The group of patients needing intensive care left the ICU in a similar time span, if we compare NON-OP and OP treated patients. The treatment policy in center B follows an expedited (within 24 hours) stabilization protocol in (severe) polytrauma or neurologically impaired patients to prevent complications of immobilization and hopefully provide for earlier rehabilitation. The literature shows that this policy is effective and safe for this patient group with high ISS scores who are at risk.[34,42] We did not, however, notice differences in complication rates between patients admitted to the ICU or polytrauma patients and patients less severely injured, but this might be due to insufficient numbers.

The overall complication rate was 19.5% for the total population and both treatments showed comparable complication rates, even when complications directly linked to treatment were scored (both 7.3%). Five patients in the NON-OP group had to be operated on later due to deteriorating neurology or persistent pain and residual deformity. A total of 4 OP treated patients needed 7 debridement operations including intravenous antibiotics because of a deep infection. This is a considerable burden for both the patient and the surgical ward and has to be included in the total cost aspect of treatments. In addition, removal of the implants occurred in 17% of patients. Routine removal of instrumentation was usually performed in the study period, but this practice was abandoned after the introduction of titanium implants.

Failure of treatment was noted in 6 (3.2%) patients, 5 were treated NON-OP, 1 OP. Neurologic deterioration of NON-OP treated patients was noted in 3 patients in our series and resulted in operative decompression and stabilization. Two of these patients deteriorated from ASIA E to D and recovered to E at the final fol-

low-up. One patient deteriorated from an ASIA C to B and recovered to D after surgery. Denis reported in his series of neurologically intact patients, a neurologic deterioration rate of 17% in the NON-OP treated group. Gertzbein reported a deterioration rate of 3.4% after admission and Kinoshita 4.3%.[7,38,43] In a recently published randomized trial, 1 (6.3%) NON-OP treated patient developed a conus medullaris syndrome. Others did not notice any deterioration.[35,36,44-47] The neurologic recovery patterns showed differences between OP and NON-OP. But after correction for possible improvement by means of proportion comparison, these differences were not statistically significant. One should also note that this was not a comparison between all operatively or conservatively treated patients with neurologic impairment. Only the patients deemed by the NON-OP school eligible for conservative treatment were included. Gertzbein noticed that Motor scores of OP treated patients were significantly better at 1 and 2-year follow-up. Similarly, Dendrinos also observed a significant improvement in OP treated patients compared with NON-OP, although this was in a nonrandomized retrospective study.[45] Other authors saw neurologic improvement in NON-OP treated[36,48] as well as in OP treated patients.[49-51] There are no randomized trials comparing neurologically impaired patients with NON-OP or OP treatment. In a review article, Boerger stated that surgical treatment for neurologically impaired patients is not justified.[41] Verlaan in a review article on OP treatment of thoracic and lumbar fractures noticed that neurologic recovery depends on the ASIA scale at admission. Of a total of 5748 patients included in this review, only 2 deteriorated after surgery (ASIA E to D). In our study, a definite answer to this question is still lacking because the sample of neurologically impaired patients was probably too small, although we did observe different patterns of recovery showing consistently better neurologic outcome in the OP group (Figure 6).

Comparing functional and health outcomes between treatments showed comparable results in all fields with a slightly poorer result for OP treated patients probably because more neurologically involved patients were in this group.

By means of uni- and multivariate analysis, we tried to identify prognostic parameters for a poor outcome of treated patients. Univariate analysis showed that of the SF-36 questionnaire role physical function, vitality, and mental health were significantly different between the sexes in favor of males. In multivariate analysis, the VAS, ODI, EQ5D, SF Physical Function, Bodily Pain, Vitality, Social Function, role Social Function, and Mental Health were all influenced by gender and neurologic impairment. We consider the minor differences in gender in baseline characteristics as not clinically relevant when comparing the 2 treatments and conclude that women overall did worse than men. As far as we know, this gender effect was not reported before in the literature on spinal fractures. Rath noticed that elderly female patients had less neurologic recovery, possibly related to an osteoporotic and stiffer spine.[52] A study by McGeary et al showed that in patients who suffer from chronically disabling spinal disorders after work injury, women fare worse than men on biopsychological outcomes, as can also be concluded from our results.[53]

Siebenga did not mention a gender difference.[8] Holbrook did notice a significant gender effect on functional and psychological outcome after major trauma, which is in accordance with our results.[54,55] However, they do not have an explanation for the differences between sexes, either.

Further analysis showed that neurologic impairment (ASIA A-D) at admission is, as expected, of substantial influence on the functional outcome. Literature on incomplete spinal cord lesions is scarce, but patients with complete spinal cord lesions show significantly lower quality of life, which slightly improves with longer follow-up.[56] Overall, the mean ODI of our population[16] is slightly higher than the ODI (10.2) mentioned by Fairbanks in a review article of a "normal population". This is, however, far better than patients with chronic low back pain who have a mean score of 43.3 on their ODI.[25] But one can argue about what constitutes a normal population, and since the mean age of our population is 38 years the "normal" ODI for this age group should probably be lower. It also raises the question whether a measurement tool designed for chronic back patients is an appropriate instrument for this group of patients. The normal values of the EQ5D and SF scores in Figures 4 and 5 are all age adjusted, and our patient population scores are worse in all items. It means that, as a result of traumatic spinal fractures, patients experience deterioration of their quality of life. Siebenga noticed significant differences in favor of OP treated patients on VAS pain, VAS spine scores, and RMDQ-24. The VAS pain scores for OP as well as NON-OP treated patients were better than our OP group, but this is probably a result of selection bias as they excluded neurologically impaired and polytrauma patients.

Concerning employability after trauma, in the long run respectively 88% and 83% of the NON-OP and OP treated patients were performing some labor, including patients who receive a percentage of workers' compensation. In our follow-up questionnaire we also asked patients to indicate the period of absence of work in relation to their injury. Half of the patients were not able to answer how many months they were on sick leave, and therefore the results were not representative. This period, however, is of major importance when we consider the cost-aspect of these injuries. Shen showed in his comparative study of neurologically intact patients that in the first year, OP treated patients had less pain and better low back outcome scores compared with NON-OP.[6] This is affirmed by Gertzbein with a significant severe pain score ratio in the NON-OP group at 1 and 2 years follow-up.[1] The percentage of patients who returned to heavy work at 2-year follow-up in the Shen group was higher in the OP group but there was no mention of when patients returned to their work in either treatment groups. The higher rate of employability in the first year might become more and more important when we consider the cost-aspect of spine fractures. NON-OP treated patients in the study of Siebenga returned to work at an average of 13.8 months, OP patients at 6.8 months, although the difference was not significant.[8] Wood found no significant differences in return to work percentage within 6 months, but 1 can doubt the power of his study with regards to this aspect as we are not informed about the physical burden of the

patients' jobs.[5] Finally, Burnham observed that if patients worked 1 year previous to injury, or were employed at time of injury or had OP therapy for their spine fracture, these were all positive predictors for employment postinjury.[57] Also, unemployment in the first year was common (46% in a cohort of 489 patients) in that study.

We only found 3 studies in which cost aspects of NON-OP and OP treatment were compared. For the neurologically unimpaired patient, the ratios NON-OP/OP treatment were 0.55, 0.63, and 0.23.[5,35,58] Obviously, OP treatment entails higher costs related to surgery and implant costs. But because of faster mobilization, shorter rehabilitation period and earlier return to work the overall costs at 1 year postinjury might turn out in favor of OP therapy.

Although relatively uncommon, spinal injuries have the lowest functional outcomes and the lowest rates of return to work after injury of all major organ systems.[59] Considering that the majority of these patients are young and working persons, the long-term economic impact of residual impairment is substantial, especially if accompanied by neurologic involvement. Optimal treatment of these patients, even if expensive, may yield considerable benefits to their communities and cost-effectiveness of treatment should be the goal of future research.

In short, OP and NON-OP treatment strategies of thoracic and lumbar spine fractures yield comparable results on the middle-long-term. The method we describe to compare treatments is potentially very useful for the study of this type of patients where there is a genuine disagreement among specialists in the field and randomized controlled trials are not feasible or yield unreliable and conflicting results because of patient selection forced by the design of randomized controlled trials. We propose to use this method for multicenter prospectively collected databases to get valid answers especially to the important questions of patient satisfaction, burden of treatment, and socioeconomic consequences within the 2-year period after injury as well as on the longer-term. This study design will be acceptable to the spine surgery community, because we can get valid answers while each surgical school may continue to perform its preferred treatment.

REFERENCES

1. **Gertzbein SD**. Scoliosis Research Society. Multicenter spine fracture study. *Spine 1992;17:528–40*.
2. **Knight RQ, Stornelli DP, Chan DP, et al**. Comparison of operative versus nonoperative treatment of lumbar burst fractures. Clin Orthop 1993; 112–21.
3. **Öner FC, van Gils AP, Faber JA, et al**. Some complications of common treatment schemes of thoracolumbar spine fractures can be predicted with magnetic resonance imaging: prospective study of 53 patients with 71 fractures. Spine 2002;27:629–36.
4. **Resch H, Rabl M, Klampfer H, et al**. [Surgical vs. conservative treatment of fractures of the thoracolumbar transition]. *Unfallchirurg 2000;103:281–8*.
5. **Wood K, Butterman G, Mehbod A, et al**. Operative compared with nonoperative treatment of a thoracolumbar burst fracture without neurological deficit: A prospective, randomized study. J Bone Joint SurgAm 2003; 85:773- 81.
6. **Shen WJ, Liu TJ, Shen YS**. Nonoperative treatment versus posterior fixation for thoracolumbar junction burst fractures without neurologic deficit. Spine 2001;26:1038–45.
7. **Denis F, Armstrong GW, Searls K, et al**. Acute thoracolumbar burst fractures in the absence of neurologic deficit. A comparison between operative and nonoperative treatment. *Clin Orthop 1984;142–9*.
8. **Siebenga J, Leferink VJ, Segers MJ, et al**. Treatment of traumatic thoracolumbar spine fractures: a multicenter prospective randomized study of operative versus nonsurgical treatment. Spine 2006;31:2881–90.
9. **Dick W, Kluger P, Magerl F, et al**. A new device for internal fixation of thoracolumbar and lumbar spine fractures: the 'fixateur interne'. Paraplegia 1985;23:225–32.
10. **Rea GL, Zerick WR**. The treatment of thoracolumbar fractures: one point of view. J Spinal Disord 1995;8:368–82.
11. **Verlaan JJ, Diekerhof CH, Buskens E, et al**. Surgical treatment of traumatic fractures of the thoracic and lumbar spine: a systematic review of the literature on techniques, complications, and outcome. Spine 2004;29:803–14.
12. **Öner FC, Ramos LM, Simmermacher RK, et al**. Classification of thoracic and lumbar spine fractures: problems of reproducibility. A study of 53 patients using CT and MRI. *Eur Spine J 2002;11:235–45*.
13. **Wood KB, Khanna G, Vaccaro AR, et al**. Assessment of two thoracolumbar fracture classification systems as used by multiple surgeons. J Bone Joint Surg Am 2005;87:1423–9.
14. **Kriek JJ, Govender S**. AO-classification of thoracic and lumbar fractures– reproducibility utilizing radiographs and clinical information. Eur Spine J 2006;15:1239–46.
15. **Vaccaro AR, Baron EM, Sanfilippo J, et al**. Reliability of a novel classification system for thoracolumbar injuries: the thoracolumbar injury severity score. Spine 2006;31:S62–S69.
16. **Vaccaro AR, Lehman RA Jr, Hurlbert RJ, et al**. A new classification of thoracolumbar injuries: the importance of injury morphology, the integrity of the posterior ligamentous complex, and neurologic status. Spine 2005;30: 2325–33.

17. Thomas KC, Bailey CS, Dvorak MF, et al. Comparison of operative and nonoperative treatment for thoracolumbar burst fractures in patients without neurological deficit: a systematic review. J NeurosurgSpine 2006;4: 351–8.

18. van der RN, de Lange ES, Bakker FC, et al. Management of traumatic thoracolumbar fractures: a systematic review of the literature. Eur Spine J 2005;14:527–34.

19. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000; 342:1887–92.

20. Devereaux PJ, Bhandari M, Clarke M, et al. Need for expertise based randomised controlled trials. BMJ 2005;330:88.

21. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996;312:1215–8.

22. Solomon MJ, Laxamana A, Devore L, et al. Randomized controlled trials in surgery. Surgery 1994;115:707–12.

22a.Stadhouder A, Öner FC, Wilson KW, et al. Surgeon equipoise as an inclusion criterion for the evaluation of nonoperative treatment of thoracolumbar spinal injuries. Spine J 2008;[Epub ahead of print].

23a. Price DD, McGrath PA, Rafii A, et al. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain 1983; 17:45–56.

23b.Stadhouder A, Oner FC, Wilson KW, Vaccaro AR, Williamson OD, Verbout AJ, Verhaar JA, de Klerk LW, Buskens E. Surgeon equipoise as an inclusion criterion for the evaluation of nonoperative versus operative treatment of thoracolumbar spinal injuries. Spine J. *2008 Nov-Dec;8(6):975-81.*

24. Fairbank JC, Couper J, Davies JB, et al. The Oswestry low back pain disability questionnaire. Physiotherapy 1980;66:271–3.

25. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine 2000;25: 2940–52.

26. Brooks R. EuroQol: the current state of play. Health Policy 1996;37:53–72.

27. Ostendorf M. Outcome assessment of total hip arthroplasty in the Netherlands and Sweden. Thesis/dissertation. *Utrecht: Utrecht University; 2004*.

28. Aaronson NK, Muller M, Cohen PD, et al. Translation, validation, and norming of the Dutch language version of the SF-36 health survey in 1016 Spine • Volume 33 • Number 9 • 2008 community and chronic disease populations. J Clin Epidemiol 1998;51: 1055–68.

29. Farasyn A, Meeusen R. The influence of non-specific low back pain on pressure pain thresholds and disability. Eur J Pain 2005;9:375–81.

30. Wood KB, Bohn D, Mehbod A. Anterior versus posterior treatment of stable thoracolumbar burst fractures without neurologic deficit: a prospective, randomized study. J Spinal Disord Tech 2005;18(Suppl):S15–S23.

31. Vaccaro AR, Daugherty RJ, Sheehan TP, et al. Neurologic outcome of early versus late surgery for cervical spinal cord injury. Spine 1997;22:2609–13.

32. Knop C, Blauth M, Buhren V, et al. [Surgical treatment of injuries of the thoracolumbar transition–3: follow-up examination. Results of a prospective multi-center study by the "Spinal" study group of the German society of trauma surgery]. *Unfallchirurg 2001;104:583–600.*

33. Sekhon LH, Fehlings MG. Epidemiology, demographics, and pathophysiology of acute spinal cord injury. Spine 2001;26:S2–S12.

34. Kerwin AJ, Frykberg ER, Schinco MA, et al. The effect of early spine fixation on non-neurologic outcome. J Trauma 2005;58:15–21.

35. Hitchon PW, Torner JC, Haddad SF, et al. Management options in thoracolumbar burst fractures. Surg Neurol 1998;49:619–26.

36. Weinstein JN, Collalto P, Lehmann TR. Long-term follow-up of nonoperatively treated thoracolumbar spine fractures. J Orthop Trauma 1987;1: 152–9.

37. Mohanty SP, Venkatram N. Does neurological recovery in thoracolumbar and lumbar burst fractures depend on the extent of canal compromise? Spinal Cord 2002;40: 295–9.

38. Kinoshita H, Nagata Y, Ueda H, et al. Conservative treatment of burst fractures of the thoracolumbar and lumbar spine. Paraplegia 1993;31: 58– 67.

39. Burke DC, Murray DD. The management of thoracic and thoraco-lumbar injuries of the spine with neurological involvement. J Bone Joint Surg Br 1976;58:72–8.

40. Jones RF, Snowdon E, Coan J, et al. Bracing of thoracic and lumbar spine fractures. Paraplegia 1987;25:386–93.

41. Boerger TO, Limb D, Dickson RA. Does 'canal clearance' affect neurological outcome after thoracolumbar burst fractures? J Bone Joint Surg Br 2000;82: 629–35.

42. McLain RF, Benson DR. Urgent surgical stabilization of spinal fractures in polytrauma patients. Spine 1999;24:1646–54.

43. Gertzbein SD. Neurologic deterioration in patients with thoracic and lumbar fractures after admission to the hospital. Spine 1994;19:1723–5.

44. Andreychik DA, Alander DH, Senica KM, et al. Burst fractures of the second through fifth lumbar vertebrae. Clinical and radiographic results. *J Bone Joint Surg Am 1996;78:1156–66*.

45. Dendrinos GK, Halikias JG, Krallis PN, et al. Factors influencing neurological recovery in burst thoracolumbar fractures. Acta Orthop Belg 1995;61: 226–34.

46. Krompinger WJ, Fredrickson BE, Mino DE, et al. Conservative treatment of fractures of the thoracic and lumbar spine. Orthop Clin North Am 1986;17: 161–70.

47. Mumford J, Weinstein JN, Spratt KF, et al. Thoracolumbar burst fractures. The clinical efficacy and outcome of nonoperative management. *Spine 1993; 18:955–70*.

48. de Klerk LW, Fontijne WP, Stijnen T, et al. Spontaneous remodeling of the spinal canal after conservative management of thoracolumbar burst fractures. Spine 1998;23:1057–60.

49. Bradford DS, McBride GG. Surgical management of thoracolumbar spine fractures with incomplete neurologic deficits. Clin Orthop Relat Res 1987; 201–16.

50. Clohisy JC, Akbarnia BA, Bucholz RD, et al. Neurologic recovery associated with anterior decompression of spine fractures at the thoracolumbar junction (T12–L1). Spine 1992;17:S325–S330.

51. Tasdemiroglu E, Tibbs PA. Long-term follow-up results of thoracolumbar fractures after posterior instrumentation. Spine 1995;20:1704–8.

52. Rath SA, Kahamba JF, Kretschmer T, et al. Neurological recovery and its influencing factors in thoracic and lumbar spine fractures after surgical decompression and stabilization. Neurosurg Rev 2005;28:44–52.

53. McGeary DD, Mayer TG, Gatchel RJ, et al. Gender-related differences in treatment outcomes for patients with musculoskeletal disorders. Spine J 2003;3:197–203.

54. Holbrook TL, Hoyt DB, Anderson JP. The importance of gender on outcome after major trauma: functional and psychologic outcomes in women versus men. J Trauma 2001;50:270–3.

55. Holbrook TL, Hoyt DB. The impact of major trauma: quality-of-life outcomes are worse in women than in men, independent of mechanism and injury severity. J Trauma 2004;56:284–90.

56. Westgren N, Levi R. Quality of life and traumatic spinal cord injury. Arch Phys Med Rehabil 1998;79:1433–9.

57. Burnham RS, Warren SA, Saboe LA, et al. Factors predicting employment 1 year after traumatic spine fracture. Spine 1996;21:1066–71.

58. van der RN, de Bruyne MC, Bakker FC, et al. Direct medical costs of traumatic thoracolumbar spine fractures. Acta Orthop 2005;76:662–6.

59. Hu R, Mustard CA, Burns C. Epidemiology of incident spinal fracture in a complete population. Spine 1996;21:492–9.

7

Natural experiments for orthopaedic trauma research: An introduction

Bryan J.M. van de Wall
Agnita Stadhouder
R. Marijn Houwert
F. Cumhur Öner
Frank J.P. Beeres
Rolf H.H. Groenwold

Injury, Nov 2022, Volume 54, Number 2

ABSTRACT

Natural experiments are observational studies of medical treatments in which treatment allocation is determined by factors outside the control of the investigators, arguably resembling experimental ran- domisation. Natural experiments in the field of orthopaedic trauma research are scarce. However, they have great potential due to the process governing treatment allocation and the existence of opposing treatment strategies between hospitals or between regions as a result of local education, conviction, or cultural and socio-economic factors. Here, the possibilities and opportunities of natural experiments in the orthopaedic trauma field are discussed. Potential solutions are presented to improve the validity of natural experiments and how to assess the credibility of such studies. Above all, it is meant to spark a discussion about its role within the field of orthopaedic trauma research.

INTRODUCTION

Randomised clinical trials (RCT) are widely accepted as the highest level of evidence for evaluating and comparing effectiveness of therapeutic interventions.[1] Nevertheless, results of traditional RCTs may have limited generalizability; the field of orthopaedic trauma research is no exception to that. The potential limited generalizability stems from the highly artificial conditions that are usually imposed on surgical practice to fit the randomised study design.[2] Surgeons frequently have a strong personal preference for a certain treatment due to technical skills, personal experience, and local culture and infrastructure.[3] These aspects play an important role in the surgeons' decision whether or not to include patients into an RCT.[2,4] In addition, patients frequently have a strong opinion as well about treatment options, particularly when it comes down to fundamentally different treatments, such as non-operative care and surgical treatment. This also contributes to selective inclusion of participants in surgical trials. What is more, the time from presentation of concept at a congress to publication of RCTs in orthopaedic trauma research is on average 10 years, which is highly undesirable in a fast-developing field like orthopaedic trauma surgery.[5]

Observational studies are increasingly regarded to provide evidence that is complementary to that from RCTs, provided the observational studies are of sufficient quality.[6,7] In contrast to RCTs, observational studies are often more representative of daily clinical practice. In addition, they are less costly and generate evidence much faster than traditional RCTs. However, due to the absence of randomisation, incomparability of treatment groups may occur leading to confounding bias. Natural experiments, a particular type of observational study, might provide a solution. In this paper, we describe different aspects of natural experiment studies, with a focus on natural experiments in orthopaedic trauma research. We discuss issues that need to be considered when conducting, reporting, or reading about natural experiment studies.

NATURAL EXPERIMENTS

In natural experiment studies, the potential for confounding might be substantially less than in other observational studies.[8,9] Natural experiments are observational studies in which patients are exposed to either the experimental or the control condition, whereby treatment allocation is determined by factors outside the control of the investigators. The process governing treatment allocation arguably resembles the random assignment in an experimental setting, hence the name natural experiment .

An example of a natural experiment could be a comparison of treatment strategies, where differences in strategies exist between hospitals or regions as a result of local education and conviction, cultural, and socio-economic differences. Generally, trauma patients will receive acute care from the nearest hospital able to

facilitate adequate treatment, which is determined by the geographical location of the incident: "what you get, depends on where you live and who you see."(10, 11) The exact location of their accident – and thus the hospital they are referred to – is, to a large extent, considered independent of the characteristics of those patients.[10-12] Hence, different trauma care facilities are expected to treat similar groups of patients. This is an ideal starting point to compare treatment strategies across hospitals or regions in a natural experiment setting.[12] For the remainder, we will refer to these hospitals or regions with opposing treatment regimens simply as "schools".

EXAMPLES OF NATURAL EXPERIMENTS IN ORTHOPAEDIC TRAUMA

An illustrative example of a natural experiment is the study by Hauschild et al.[13] They compared non-operative with operative treatment for proximal humerus fractures by comparing patients across four hospitals in Switzerland. One of the participating centres consistently offered non-operative treatment to all their patients while the other three performed surgery on all their patients with proximal humerus fractures. Patients were very similar across the treatment groups, providing the possibility to make valid comparisons.

Another example of a natural experiment is a study by Stadhouder et al.,[12] who used information available in medical records to compare patients hospitalised for traumatic spinal fractures in two university trauma centres in the Netherlands. One of these centres had a long-established treatment strategy of non-operative care and performed surgery on rare occasions. In the other centre, patients more often received operative care in case of traumatic spinal fractures. Since the patient groups that were admitted to either of the two hospitals were very similar, this allowed for a comparison between treatment strategies.

A third example is a natural experiment in the form of a pre-post design (or before-after design). Schoenfeld et al. compared 14 040 patients with femoral neck fractures prior to implementation of a new healthcare reform in Massachusetts with 9445 patients after implementation with regard to cost-effectiveness and complications.[14] Again, baseline characteristics between the two groups of patients were remarkably similar.

METHODOLOGICAL CHALLENGES OF A NATURAL EXPERIMENT

Defining the research question

When designing a natural experiment, the first step, as in all research, is to define a clear research question. It should be articulated which treatment regimens are compared (intervention as well as comparator), in what patient population

with which clinical condition (study population) and clearly define the outcome of interest (primary and secondary outcomes). A frequently used structure to articulate a research question, is the so-called PICO (Table 1).

In orthopaedic trauma research the intervention and comparator are often defined only by the nature of the treatment itself (non-operative or operative) and, in case of operative treatment, the surgical technique. It is important to also incorporate aftercare into these elements of the research questions for several reasons. Aftercare strategies may differ between hospitals or regions. Additionally, they are part of the treatment strategies patients receive and may impact clinical and functional outcomes. In the context of a natural experiment, they are part of the 'school' that patients are exposed to and thus should be clearly defined.

Clearly defining clinical outcomes is also important. In orthopaedic trauma, many clinical outcomes can be measured objectively and are frequently based on events requiring (operative or medical) interventions, radiological, biochemical or microbiological outcome data. A clear outcome definition should include a time component (*when* is it measured) and manner in which it is measured, which is frequently neglected in current literature.[15]

Design

For natural experiments in orthopaedic trauma there are certain design elements that should be considered to maximise its potential. The backbone of this design is formed by a treatment allocation process that is (to a large extent) independent of patient characteristics.[10,11] The archetype of a natural experiment in orthopaedic trauma is a comparison between hospitals where different treatment protocols are implemented, while referrals to the different hospitals are independent of patient characteristics (i.e., similar patient "case-mix" across hospitals). Preferably these "schools" consistently provide one of the treatment options to all (or the majority) of their patients with the clinical condition of interest. When performing a natural experiment, it is important that researchers convincingly argue that treatment allocation is indeed independent of individual patient characteristics, rather than trying to find convincing arguments in the comparison of baseline characteristics between treatment groups.

Important to note is that the setting of the "schools" should be similar in order to prevent relevant case-mix differences (i.e., difference in characteristics of their treated patients) possibly leading to confounding. For natural experiments in orthopaedic trauma care this means that the "schools" should provide the same level of trauma care (level I, II or III) and be located in regions with the same socio-economic development. Essentially, both "schools" should be comparable to such a degree that it is plausible to assume that school A could have provided the treatment from school B, and vice versa, if their conviction on optimal treatment had been different. Once these "schools" have been identified, they can be used to compare the treatments of interest.

All eligible patients should be registered, including patients that are excluded in order to gain insight on possible selection mechanisms by comparing patient

PICO	Minimal set of items to report/asses (if applicable):
Population	**Anatomical location fracture**
	Type of fracture (open/closed, simple/multifragmentary or combination/all)
	Age group
Intervention	**In case of surgical treatment:**
	Osteosynthesis material
	Surgical approach
	Postoperative treatment (type&duration)
	In case of conservative treatment:
	Type of conservative treatment (including duration)
Comparator	**In case of surgical treatment:**
	Osteosynthesis material
	Surgical approach
	Postoperative treatment (type&duration)
	In case of conservative treatment:
	Type of conservative treatment (including duration)
Outcome	**What is the outcome?**
	When is the outcome assessed?
	How is the outcome assessed?

Table 1. PICO for orthopaedic trauma research.

Pros	Cons
Inexpensive	Dependant on natural variation for feasibility of study
High patient accrual rate	No true randomization (does not rule out unmeasured confounding)
Generalisability of results	
Good internal validity due to quasi-randomisation	
Conduct research that is otherwise unethical.	

Table 2. Pros and cons of natural experiments in orthopaedic trauma research.

characteristics between included and excluded patients. Patients should be treated according to the local preference and conviction of the "schools" with regard to the optimal treatment for their clinical condition.

A natural experiment study can be performed retrospectively as well as prospectively. The advantages of a prospective design are that follow-up and measurement of baseline characteristics and outcomes can be pre-specified and standardised across the "schools", thus reducing the potential for information bias. Fig. 1 illustrates the necessary steps. Patients with the clinical condition of interest are identified through a hospital records search (retrospective design) or during their visit at the emergency department/outpatient clinic (prospective design) in participating hospitals representing different "schools" of intervention (school A and B). Data on baseline characteristics and the clinical condition of each patient should be collected.

Comparability of treatment groups

Even though natural experiments aim to compare different "schools" across, for example, different hospitals, patient groups may differ between schools in more respects than only the treatment strategies under study. To the extent that such differences in potential confounding variables are measured, this can be controlled for in the analysis, such as in any observational study of treatment effects. It is therefore of the utmost importance to collect data on key prognostic patient characteristics, as these will be needed in the statistical analysis of the study to correct for possible confounding (Fig. 1). This advice holds irrespective of whether data are collected retrospectively or prospectively, be it that in prospective studies it may be possible to ensure that information about confounding variables is collected in a standardized manner and possibly the proportion of data being missing is smaller than in retrospective studies using routinely collected data for example based on electronic patient records.[16]

Conventional methods to correct for (measured) confounding include stratification, regression adjustment, and matching. Another approach to reduce the amount of confounding is to restrict the study population by using clinical equipoise as an inclusion criterion.[12] In practise, this could be achieved by presenting all relevant data of eligible patients to an independent expert panel blinded for actual treatment received. The expert panel should consist of representatives from both "schools" of intervention. The panel is asked to decide independently on the preferred treatment for the eligible patient as if patients were presented to them in clinical practice. Patients are included if there is disagreement on treatment choice between the "schools"; they are excluded in case of agreement (Fig. 1). This ensures that the included study population consists of patients that would have received treatment A in "school A" but were in fact seen and treated according to the conviction of "school B", and vice versa (exchangeability). This way, the patients for whom the panel agrees regarding preferred treatment strategy,

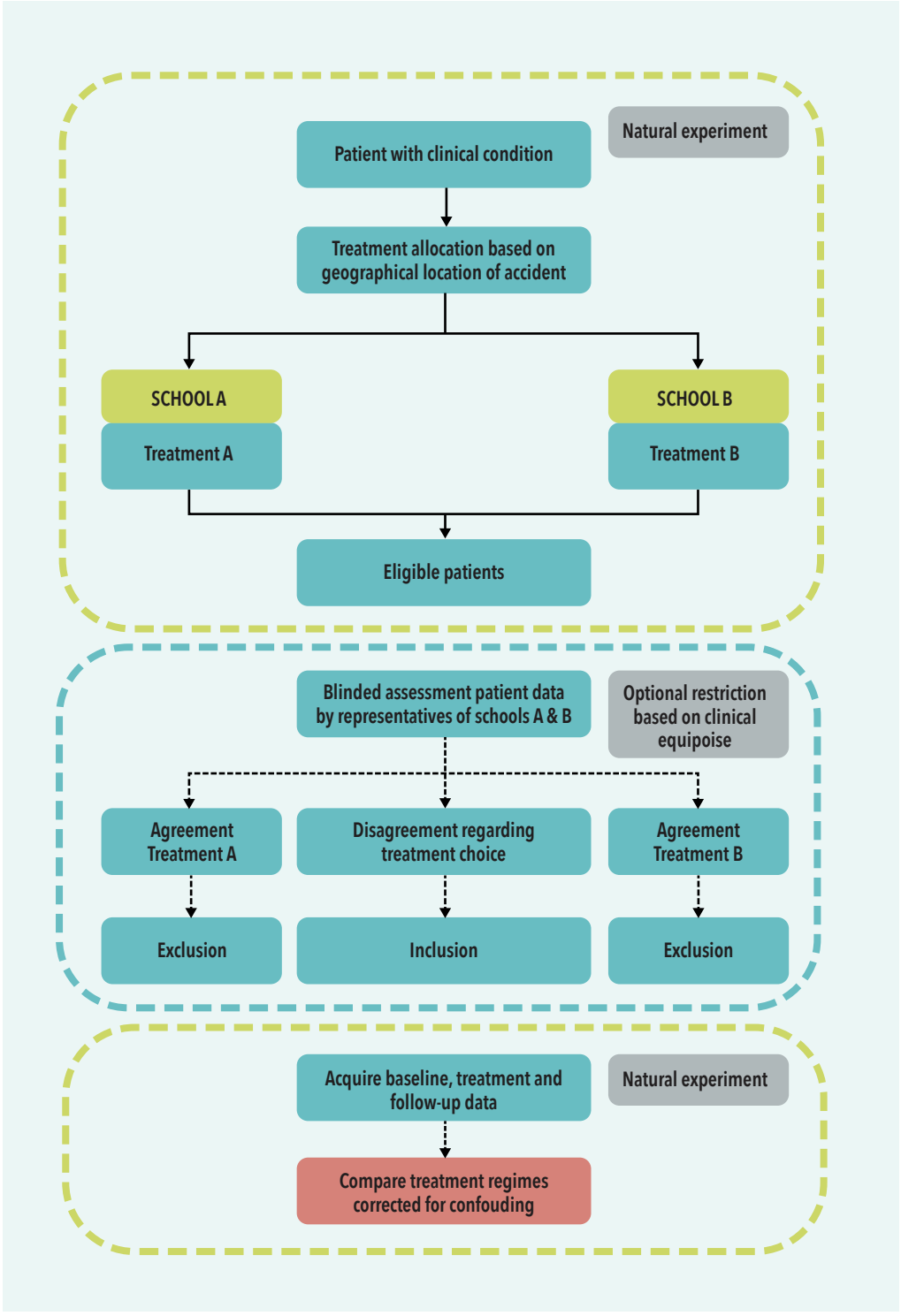

Fig 1. Flowchart of patient inclusion for the natural experiment including the restriction method using clinical equipoise as criterion).

which are generally patients with very distinct disease or patient characteristics driving treatment preference unanimously in one direction, will be excluded. By restricting the study population to those for whom there is clinical equipoise, the potential impact of confounding is reduced. Clinical equipoise as inclusion criterion can be used both in prospective and retrospective natural experiment studies.

Regarding the use of an expert panel and clinical equipoise for inclusion of patients, there are different options to implement this in a study. For example, patients could be included if at least 20% of panel members disagree with the other 80% of panel members. (1:4 distribution amongst experts). This means that patients are eligible for inclusion if (in a panel consisting of, for example, five experts) 4 (or fewer) experts prefer treatment A, while 1 (or more) prefer treatment B for a given patient. Such a threshold could be based on a study that assessed at which proportion of agreement on the merit of a new treatment amongst ethical committee members, the members perceived the conduct of a trial investigating the new treatment as ethically responsible (the level of collective equipoise).[17]

Importantly, the use of clinical equipoise as inclusion criterion is expected to reduce the number of patients included in the study. In the aforementioned study by Stadhouder et al., 190 of the 636 patients (30%) could be included based on this criterion.[12] The addition of clinical equipoise as inclusion criteria should therefore not be seen as a necessity but rather an extension of the natural experiment design to further improve comparability of treatment groups.

Irrespective of whether this restriction method is used or not, it is important to assess the distributions of baseline characteristics across the treatment groups. This provides insight into the apparent comparability of "schools" and whether clinical equipoise as inclusion criterion has proved successful in creating comparable treatment groups. Nevertheless, known confounders could still be accounted for, for example through a multivariable regression analysis or propensity score analysis.[18]

Reporting of natural experiments

The STROBE statement is a checklist of items that should be reported on in papers about observational studies.[19] The RECORD statement is a reporting guideline for studies using routinely collected data.[19] Many of the items mentioned in these reporting guidelines are also applicable to natural experiments. Some items, however, require specific attention when reporting on natural experiments for orthopaedic trauma. In case of a comparison between ''schools'', it is essential to give insight whether participating schools offer only one treatment (''pure school''), or both treatment modalities under investigation, but with a distinct preference of one treatment over the other (''majority school''). In the latter situation, proportions of applied treatments within schools should be reported on. What is more, arguments should be provided to support the assumption of comparability of patient groups across different schools. In addition, details about the compared strategies should be reported, including peri-operative care and after-treatment, except

perhaps in case these are according to (international) standards, in which case a reference to those standards would be sufficient.

DISCUSSION

Natural experiments in the field of orthopaedic trauma are still uncommon.[20-23] Nevertheless, this study design has great potential in this field compared to traditional observational study designs. Under the conditions outlined above (specifically regarding comparability of treatment groups), evidence obtained through natural experiments may be complementary to the evidence obtained through randomised trials. In particular, in orthopaedic trauma, patients are exposed to high variability of surgical decision making caused by strong convictions by surgeons as "surgeons agree mostly with themselves, and not so much with each other".[24] In natural experiments, this variability is turned into an advantage, by using it as the basis of a comparison between treatment strategies.[3] All pros and cons of natural experiments are described in Table 2 .

According the Oxford level of evidence natural experiments are categorized as observational cohort studies, thus traditionally considered level 2b.[25] It should, however, be acknowledged that natural experiments differ from traditional observational studies by the fact that confounding is addressed in both the study design (school comparison) and analysis stage (correction for confounders) in contrast to traditional observational studies that generally only perform the latter. By incorporating a measure to limit confounding in the design, they share more similarities with randomised clinical trials than traditional observational studies; hence also the similarity in nomenclature between "natural experiment" and the alternative name for randomised clinical trial, "randomised experiment". The most pronounced differences between randomised clinical trials, natural experiments and traditional observational studies are described in Table 3 .

We would like to stress the importance of a proper sample size calculation as integral part in conducting a natural experiment.[26] One can only draw a precise and accurate conclusion with a sufficiently large sample size. A smaller sample will give a result which may not be sufficiently powered to detect a difference between the groups and the study may turn out to be falsely negative leading to a type II error. Natural experiments follow the same standard approach to sample size calculation as any other empirical study.[8] Also in natural experiments, the sample size calculation is based on the primary endpoint of interest.

Both the annual incidence of the clinical condition of interest and estimated proportion that are expected to be included by using clinical equipoise as inclusion criterion, play a vital role in evaluating feasibility of the planned natural experiment. In order to estimate the proportion that may be included when using clinical equipoise as inclusion criterion, one can measure the amount of disagreement in the expert panel prior to conducting the study. Basically, this can be done by subjecting clinical data of, for example, 12 random historical patients with the clinical

condition of interest to the expert panel from the opposing "schools". The amount of disagreement reflects the proportion of all patients with the clinical condition of interest that can be included in the study. As described previously, the addition of clinical equipoise as inclusion criteria should not be seen as a necessity but rather an extension of the natural experiment design if conditions allow the inclusion of this design-element into the study.

In recent years several prospective natural experiments have been initiated by the Natural Experiments (NEXT) Study Group. The NEXT Study Group is an international non-profit collaboration of clinical researchers in the field of emergency and (orthopaedic) trauma surgery. The ambition of the NEXT Study Group is to contribute to the improvement of patient care by collecting relevant evidence through international natural experiments. Ongoing studies include the OPVENT study comparing non-operative care to surgical treatment for multiple rib fractures and the LADON proximal humerus study also comparing non-operative and operative

	Randomised clinical trials	Traditional observational studies	Natural experiments
Exposure (intervention)	Intervention that may differ from clinical practice. Usually two interventions included in trial.	Standard clinical practice.	Standard clinical practice.
Population	Often restricted to younger and relatively healthy patients.	Can include entire population.	Can include entire population.
Confounding control	Control for both measured and unmeasured confounding through randomisation.	Control for measured confounding through statistical correction. No control for unmeasured confounding.	Control for measured confounding through statistical correction. Control (to unknown extent) for unmeasured confounding through school comparison.
Costs	Expensive	Often inexpensive	Often inexpensive
Time frame	Time consuming due to partial inclusion of patients.	Often fast as most patients are included.	Often fast as most patients are included.
Outcome	Standardised measurement of endpoints.	Measurement of endpoints restricted by routine clinical practice.	Standardised measurement of endpoints.
Blinding patient	Possible	Not possible	Not possible
Blinding outcome assessor	Possible	Unusual	Possible

Table 3. Differences between randomised clinical trials, traditional observational studies and natural experiments in orthopaedic trauma research.

treatment strategies. In the LADON study clinical equipoise is used as an additional inclusion criterion.[27,28]

Orthopaedic trauma is a fast-developing field requiring study designs that deliver high quality evidence and, most of all, can keep up with ongoing developments within the field. This manuscript discusses the possibilities of natural experiments as a means to provide valuable evidence and how to assess the credibility of such studies within the orthopaedic trauma field. Above all, it is meant to spark a discussion about its role within our research field.

REFERENCES

1 Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med 1982;72(2):233–40* .

2 Beks RB, Bhashyam AR, Houwert RM, van der Velde D, van Heijl M, Smeeing DPJ, *et al.* When observational studies are as helpful as randomized trials: examples from orthopedic trauma. *J Trauma Acute Care Surg 2019;87(3):730–2* .

3 Houwert RM, Beks RB, Dijkgraaf MGW, Roes KCB, Öner FC, Hietbrink F, et al. Study methodology in trauma care: towards question-based study designs. *Eur J Trauma Emerg Surg 2021;47(2):479–84* .

4 McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ 2002;324(7351):1448–51*

5 Axelrod D, Trask K, Buckley RE, Johal H. The Canadian Orthopaedic Trauma Society. *Bone Joint J 2021;103-B(5):898–901*

6 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med 20 0 0;342(25):1887–92*

7 Abraham NS, Byrne CJ, Young JM, Solomon MJ. Meta-analysis of well-designed non-randomized comparative studies of surgical procedures is as good as randomized controlled trials. *J Clin Epidemiol 2010;63(3):238–45*

8 Craig P, Katikireddi SV, Leyland A, Popham F. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. *Annu Rev Public Health 2017;38:39–56*

9 Uddin MJ, Groenwold RH, Ali MS, de Boer A, Roes KC, Chowdhury MA, *et al.*Methods to control for unmeasured confounding in pharmacoepidemiology: an overview. *Int J Clin Pharm 2016;38(3):714–23*

10 Fanuele J, Koval KJ, Lurie J, Zhou W, Tosteson A, Ring D. Distal radial fracture treatment: what you get may depend on your age and address. *J Bone Joint Surg Am 2009;91(6):1313–19*

11 Koval KJ, Lurie J, Zhou W, Sparks MB, Cantu RV, Sporer SM, *et al.* Ankle fractures in the elderly: what you get depends on where you live and who you see. *J Orthop Trauma 2005;19(9):635–9* .

12 Stadhouder A, Buskens E, de Klerk LW, Verhaar JA, Dhert WA, Verbout AJ, *et al.* Traumatic thoracic and lumbar spinal fractures: operative or nonoperative treatment: comparison of two treatment strategies by means of surgeon equipoise. *Spine (Phila Pa 1976) 20 08;33(9):10 06–17*

13 Hauschild O, Konrad G, Audige L, de Boer P, Lambert SM, Hertel R, et al. Operative versus non-operative treatment for two-part surgical neck fractures of the proximal humerus. *Arch Orthop Trauma Surg 2013;133(10):1385–93*

14 Schoenfeld AJ, Weaver MJ, Power RK, Harris MB. Does Health Reform Change Femoral Neck Fracture Care? A Natural Experiment in the State of Massachusetts. *J Orthop Trauma 2015;29(11):494–9*

15 Paul Pynsent JF, Carr Andrew. Outcome measures in orthopaedics and orthopaedic trauma. *CRC Press; 2004*

16 Bots SH, Groenwold RHH, Dekkers OM. Using electronic health record data for clinical research: a quick guide. *Eur J Endocrinol 2022*

17 Mhaskar R, BB B, Djulbegovic B. At what level of collective equipoise does a randomized clinical trial become ethical for the members of institutional review board/ethical committees? *Acta Inform Med 2013;21(3):156–9*

18 Hernán M.A. *RJCIWIBRCHC, 2006.*

19 von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting of observational studies. *Internist 2008;4 9(6):6 88–93*

20 Beks RB, Ochen Y, Frima H, Smeeing DPJ, van der Meijden O, Timmers TK, et al. Operative versus nonoperative treatment of proximal humeral fractures: a systematic review, meta-analysis, and comparison of observational studies and randomized controlled trials. *J Shoulder Elbow Surg 2018;27(8):1526–34*

21 Smeeing DPJ, van der Ven DJC, Hietbrink F, Timmers TK, van Heijl M, Kruyt MC, *et al.* Surgical versus nonsurgical treatment for midshaft clavicle fractures in patients aged 16 years and older: a systematic review, meta-analysis, and comparison of randomized controlled trials and observational studies. *Am J Sports Med 2017;45(8):1937–45*

22 van de Wall BJM, Ochen Y, Beeres FJP, Babst R, Link BC, Heng M, *et al.* Conservative vs. operative treatment for humeral shaft fractures: a meta-analysis and systematic review of randomized clinical trials and observational studies. *J Shoulder Elbow Surg 2020*

23 Ochen Y, Beks RB, van Heijl M, Hietbrink F, Leenen LPH, van der Velde D, *et al.* Operative treatment versus nonoperative treatment of Achilles tendon ruptures: systematic review and meta-analysis. *BMJ 2019;364:k5120*

24 Doornberg JN, Rademakers MV, van den Bekerom MP, Kerkhoffs GM, Ahn J, Steller EP, *et al.* Two-dimensional and three-dimensional computed tomography for the classification and characterisation of tibial plateau fractures. *Injury 2011;42(12): 1416–25*

25 Durieux N, Vandenput S, Pasleau F. [OCEBM levels of evidence system]. *Rev Med Liege 2013;68(12):644–9*

26 Nayak BK. Understanding the relevance of sample size calculation. *Indian J Ophthalmol 2010;58(6):469–70*

27 Beks RB, de Jong MB, Sweet A, Peek J, van Wageningen B, Tromp T, *et al.* Multicentre prospective cohort study of nonoperative versus operative treatment for flail chest and multiple rib fractures after blunt thoracic trauma: study protocol.*BMJ Open 2019;9(8):e023660*

28 Nonoperative versus Operative Treatment for proximal humeral fractures. An International multicentre prospective cohort study (LADON). *Netherlands Trial Register Trial NL9357, 2022.*

8

Natural Experiments as a Study Method in Spinal Trauma Surgery: A Systematic Review

Agnita Stadhouder
Luke X. van Rossenberg
Charlotte Kik
Sander P. J. Muijs
F. Cumhur Öner
R. Marijn Houwert

Global Spine J . 2024 Jun;14(5):1640-1649

ABSTRACT

Study Design: Systematic review.
Objectives: To determine if the natural experiment design is a useful research methodology concept in spinal trauma care, and to determine if this methodology can be a viable alternative when randomized controlled trials are either infeasible or unethical.
Methods: A Medline, Embase and Cochrane database search was performed between 2004 and 2023 for studies comparing different treatment modalities of spinal trauma. All observational studies with a natural experiment design comparing different treatment modalities of spinal trauma were included. Data extraction and quality assessment with the MINORS criteria was performed.
Results: Four studies with a natural experiment design regarding patients with traumatic spinal fractures were included. All studies were retrospective, one study collected follow-up data prospectively. Three studies compared different operative treatment modalities, whereas one study compared different antibiotic treatment strategies. Two studies compared preferred treatment modalities between expertise centers, one study between departments (neuro- and orthopedic surgery) and one amongst surgeons. For the included retrospective studies, MINORS scores (maximum score 18) were high ranging from 12- 17 and with a mean (SD) of 14.6 (1.63).
Conclusions: Since 2004 only four studies using a natural experiment design have been conducted in spinal trauma. In the included studies, comparability of patient groups was high emphasizing the potential of natural experiments in spinal trauma research. Natural experiments design should be considered more frequently in future research in spinal trauma as they may help to address difficult clinical problems when RCT's are infeasible or unethical.

INTRODUCTION

In current evidence based medicine practice, randomized controlled trials are considered the gold standard, as this methodology is particularly effective in preventing selection bias, information bias and confounding.[1-3] However, randomized controlled trials in surgical fields may encounter certain difficulties which reduce reliability of results and allow introduction of bias.[4-6] Practical difficulties such as a learning curve for new procedures, variation in quality of surgical performance or clinician and patient equipoise are common in general surgical studies.[7-11] Methodical difficulties include challenges in acquiring informed consent, blinding of patients and randomization of patients.[9,11,12] In acute surgery fields, where urgent lifesaving treatment is often involved, randomized controlled trials are difficult to conduct properly due these challenges. This is also the case for spinal trauma care.[10,13] Other study designs might pose a more viable solution.[9]

Observational studies have historically been used to demonstrate credible results in situations where a randomized controlled trial is either unethical or unfeasible.[14] However, observational studies are more prone to bias and confounding.[14-16] To minimize confounding, observational studies must be carefully and rigorously designed.[17] In therapeutic studies a randomized design has greater value and credibility of results compared to observational studies and Vandenbroucke[17] states that observational studies will be credible only in exceptional circumstances. To ensure similar credibility of observational studies compared to randomized studies, three essential restrictions have been proposed by Vandenbroucke[17] in the Lancet in 2004.[17] The first restriction pertains to the selection of research topics where allocation of exposure is minimally associated with the outcome of interest. This is the most easily applicable in studies on adverse events as these are always unintended and their risk unknown or unpredictable. The second restriction involves that a study design is required to have at least a quasi-random allocation of exposure to treatment. Quasi-random allocation is a method of allocating participants which is not fully random or blinded but prevents researcher/clinician biased allocation of treatment based on patients characteristics or prognosis.[17,18] Examples of quasi random allocation include allocation by date of birth or geographical location. The third is restriction to topics where potential confounding variables can be identified, accurately measured, and appropriately adjusted for in statistical models.

Among the different types of observational studies, the natural experiment is a promising method that mimics the design of an RCT without the need for randomization. As described by van de Wall et al "A natural experiment is a quasi-experimental study in which patients are exposed to either the experimental or control condition, whereby treatment allocation is determined by factors outside the control of the investigators."[18] To ensure adequate comparability, it is crucial that a genuine state of clinical equipoise is present, where both treatment strategies are considered equally viable options.[19] Clinical equipoise is "a state of genuine

uncertainty on the part of the clinical investigator regarding the comparative therapeutic merits of different treatment options"[20] Clinical equipoise resulting in different treatment strategies can occur on various levels, e.g. amongst surgeons, hospitals, expertise centers, or so called "schools", as well as internationally.[21] A natural experiment becomes feasible when clinical equipoise is present and allocation of treatment is dependent on external factors.[21] This is especially true for trauma patients. Generally, trauma patients will receive acute care at the nearest hospital able to facilitate adequate treatment.[22,23] In this case allocation of treatment is determined by the geographical location of the incident, rather than by patient characteristics or any manipulation of the researcher. The hospital where the patient is treated determines the exposure to either the control or experimental condition and utilizing natural variation of treatment allocation increases validity of results as it emulates randomization.[21] Multiple natural experiments in trauma surgery have been conducted reporting results matching the credibility of randomized controlled trials.[24–28] However, natural experiment design is a relatively new study method in surgical research. This is also the case in spinal trauma and it is currently unknown to what degree natural experiment designs are utilized and to what extent they provide credible evidence.

Therefore, this systematic review aims to investigate to what extent natural experiment design has been conducted in all types of spinal trauma, and if they pose a viable alternative for randomized controlled trials in this field.

METHODS

Search Strategy

This study was conducted in line with the PRISMA guidelines. We systematically searched literature on primary intervention studies reporting on natural experiments in spinal trauma patients. The systematic search was performed from 2004 until 2023 and updated on the 30st of March using the search terms 'spinal trauma', 'spinal fractures', 'vertebrae' and synonyms in the Medline, Embase and the Central databases. Full text, English or Dutch written articles were reviewed for inclusion. The full search string is provided in Appendix A.

Study Selection and Eligibility Criteria

Three reviewers (AS, SC, LXR) independently assessed the titles and abstracts to identify cohort studies with a natural experiment design in adult spinal trauma patients. A study was considered a natural experiment if there was evident geographical (pseudo)randomization of treatment allocation, either amongst schools, departments or surgeons. (e.g., surgeon A always performs a certain type of treatment, whilst surgeon B always performs a different type of treatment for similar injuries). Historic comparison studies were excluded since in a certain time span of the research period also other factors can be changed.

Subsequently, full texts were independently evaluated for eligibility following in- and exclusion criteria, which are displayed in Table 1. Disagreement was resolved through consensus. Non-English or Dutch reports, randomized controlled trials, systematic reviews, and meta-analyses, reviews, cohorts with a historical control, case-control studies, cross-sectional studies, case reports, case series, conference abstracts, editorials, letters and comments and animal studies were excluded. EndNote X8 was used to manage the screening and reviewing process. Finally, the reference lists of included articles and relevant reviews were screened for eligible studies.

Data Extraction

Two investigators (AS and LXR) extracted data independently of all included studies. From each eligible study, the following data were collected: first author, year of publication, country of conduct, study design, number of included spinal trauma patients, number of patients in the intervention group, number of patients in the control group, mean age of participants, gender and the mean Injury Severity Score (ISS) if available.

Quality Assessment

The methodologies of the included studies were critically appraised using the validated Methodological Index for Non-Randomized Studies (MINORS) criteria, which assesses articles on the presence of various forms of bias including selection, performance, detection, attrition, reporting and other bias (scored as 'not reported', 'reported but inadequate' or 'reported adequately').[29] Two authors (AS and LXR) scored all articles independently. When in disagreement, a third reviewer (CK) was asked to make an additional assessment and the majority vote was counted.

Inclusion Criteria		Exclusion Criteria	
Population	All patients with a traumatic spinal fracture (total spine)	Population	Malignant or osteoporotic fractures
Intervention/ comparison	Comparison between any two treatments in spinal trauma surgery	Intervention/ comparison	No spinal treatment
Outcome	Not applicable	Outcome	Not applicable
Study design	Observational study with natural experiment as study design	Study design	Historic comparison
			Other study design[a]

NE, natural experiment.
[a](RCT, case series, case control, case report, observational cohort studies without NE* design).

Table 1. Summary of Inclusion and Exclusion Criteria.

A maximum of 24 points could be given to prospective comparative studies, and 18 points for retrospective comparative studies, as MINORS criteria "prospective collection of data", "loss to follow up less than 5%" and "prospective calculation of the study size" are not applicable for retrospective designs. Retrospective comparative studies with MINORS scores ranging between 12 and 18 are considered high quality.[30] Further information on the assessment of methodological quality is provided in Appendix B.

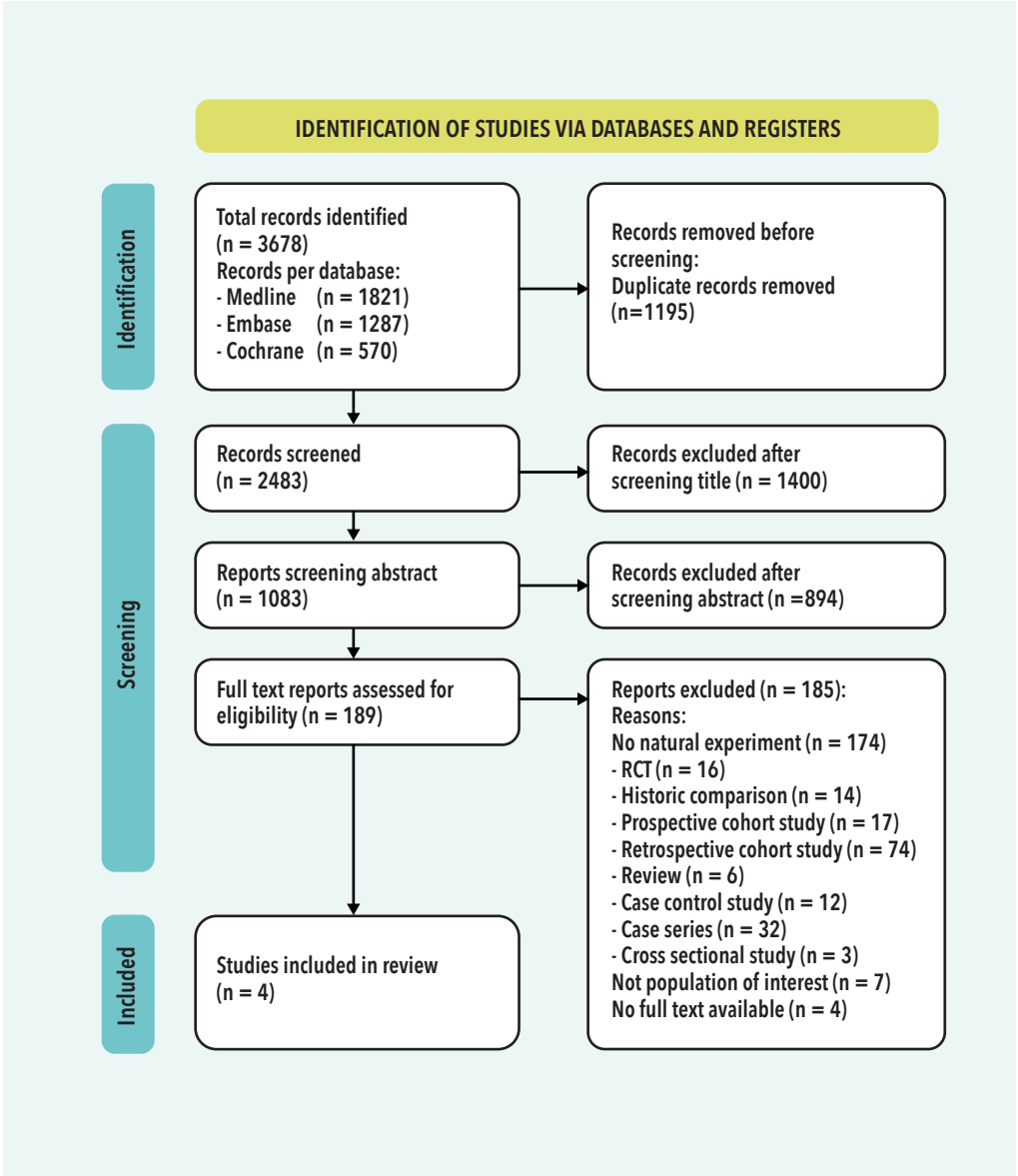

Fig 1. Flowchart of patient inclusion for the natural experiment including the restriction method using clinical equipoise as criterion).

RESULTS

Identification of Studies

The systematic search yielded 3678 articles. After removal of duplicates, 2483 articles were screened on title and abstract for eligibility. One-hundred and eighty-nine citations were retrieved for full-text assessment and evaluated for inclusion. One-hundred and eighty-five studies were excluded because they did not adhere to the desired "natural-experiment" design, outlined in Figure 1.

Study Characteristics

The overall characteristics of the included studies are summarized in Table 2. Studies were published between 2008 and 2021.[13,31-33] All four studies were designed retrospectively,[13,31-33] of which one study retrieved participant retrospectively but collected patient reported outcomes actively of included participants.[13] Three studies were performed in Western Europe,[13,31,32] while one was conducted in the United States of America.[33] Two studies included a multi-center setting.[13,31] All participating hospitals in the studies were level-1 trauma centers.[13,31-33] Three studies included a comparison between different surgical treatment modalities for acute spinal fractures,[13,31,32] while one study focused on the rate of infectious complications by the use of vancomycin powder in posterior spinal stabilization of traumatic injuries.[33] Of the 852 participants in the included studies 481 (56%) were male, and the mean age ranged from 37 to 69 years.[13,31-33] Mean follow-up ranged from 6 months to 74 months.[13,31-33] Two studies reported the injury severity score (ISS) of which all participants had mean ISS scores ≤ 16.[31,32] Three studies reported trauma mechanisms and of the 742 participants the majority was injured due to a fall (78.8%), the minority by traffic accident (11.5%) or other specific causes (9.2%) (e.g., paragliding, horse riding or skiing).[13,31,32] Of all spinal trauma injuries, the majority of patients had either a thoracic and/or lumbar fracture (66.2%), whereas cervical fractures were less common (33.8%).[13,31-33] Three studies reported presence of neurological impairment, of all 765 patients 157 (20.5%) were partially or completely neurologically impaired.[13,31,32] One study excluded patients with cervical fractures and/or neurological impairment.[33]

Quality Assessment

On average total MINORS score for the four retrospective studies were high ranging from 12 to 17 with a mean (SD) of 14.6 (1.63). Stadhouder et al scored highest with a score of 17,[13] followed by O'Neill et al and Erichsen et al both with score of 16.[31-33] Myers et al scored lowest with a score of 12.[32] Average scores in the MINORS section "additional scores for comparative studies" (range 0-8) were high, ranging from 6 to 8 with a mean (SD) of 7.25 (.75).[13,31-33] See Table 3 for an overview of individual scores.

The Natural Experiment Design

O'Neill et al used a natural experiment design to evaluate the effectiveness of perioperative intrawound vancomycin powder use in patients who underwent posterior spine fusion to prevent infections. Retrospective identification of patients over a 2-year period at a single academic center resulted in two groups: those who received vancomycin powder in their surgical wound during their initial surgery and those who did not, following the standard of care at the time. Patients were (pseudo)randomized by surgeon preference and only one surgeon always treated patient with intrawound vancomycin, whereas other surgeons did not. The study found that the use of intrawound vancomycin significantly reduced the incidence of infections in patients with traumatic spine injuries. Data retrieval began in 2004, and the study was published in 2011.[33] The study demonstrated the possibility to use a natural experiment design in pharmaceutical studies.

Similarly, Stadhouder et al conducted a study on the operative vs nonoperative treatment of traumatic thoracic and lumbar spinal fractures using a natural experiment design. A blinded panel of orthopedic surgeons from two University Medical Centers retrospectively reviewed cases where there was disagreement on the suggested treatment modality, creating two comparable groups of patients who either underwent nonoperative or operative treatment. After an average follow-up period of 6 years, patients' clinical outcomes were compared, and it was found that operative and nonoperative treatments were comparable. Start of data retrieval was 2004, publication year 2008. A limitation of the study was its retrospective design and longer follow up between treatment and outcome, which led to probable missing data. The authors suggested that a natural experiment design could also be used in prospective series.[13] The study demonstrated the possibility of using clinical equipoise to create two comparable groups and compare treatment outcomes without influencing the treatment preferences of the surgeon/school.

In subsequent years, the natural experiment design has been utilized in two further studies related to the management of spinal fractures.[31,32] Erichsen and colleagues conducted a retrospective review of cases involving patients with a traumatic AO spine type A3 fracture of the thoracolumbar spine who received different types treatment depending on which hospital they were treated. In one hospital all patients received open posterior stabilization, while in the other hospital all patients underwent percutaneous posterior stabilization. The treatment effects were evaluated after a 2-year follow-up period using the Owestry Disability Index, Visual Analog Scale, and a 36-item Short Form Health Survey. The trial was registered in the German Clinical Trial Registry in 2018, publication was in 2020.[31] Similarly, Myers and colleagues conducted a retrospective evaluation of the difference in direct treatment outcomes between patients with spinal fractures who were treated by neurosurgical teams vs those treated by orthopedic teams in weekly shifts. The end of data retrieval period was December 2016, publication was in 2021.[32] Both research groups had similar baseline characteristics, admit-

											Trauma Mechanism, n (%)			Fracture Location, n (%)		
					Number of Patients		Age in Years, Mean (SD/ Range)	Gender Male, n (%)	Follow Up in Months, Mean (SD/Range)	Mean (SD) ISS[b]	Fall	Traffic Accident	Other	Cervical	Thoraco-lumbar	Neurological Impairment, n(%)
Study and Year	Study Period	Design	Country	Natural Experiment Comparison[a]	Total	Case/ Control	Case/ Control	Case/ Control	Case/ Control	Case/ Control	Case/ Control	Case/ Control	Case/ Control	Case/ Control	Case/ Control	Case/Control
Myers, 2021	2012-2016	Retrospective	United Kingdom	Departments, SS	465	266/199	60.2 (21.3)/61/1 (22)	146 (54.9%)/ 106 (53.3%)	*na*	8.7 (4.6)/ 8.9 (6.9)	446 (95.9%)[c]	15 (3.2%)[c]	4 (0.8%)[c]	128 (20.8%)/ 106 (17.2%)	229 (37.2%)/152 (24.7%)	27 (10.2%)/20 (10%)
Erichsen, 2020	2013-2015	Retrospective	Switzerland/ Germany	Schools, MS	87	44/43	43.5 (14.3)/ 48.4 (12.2)	19 (43.2%)/ 26 (60.5%)	24 (no SD)[c]	ISS ≤ 16 (no SD)[c]	21 (47.7%)/30 (69.8%)	9 (20.4%)/11 (25.5%)	14 (31.8%)/2 (4.6%)	Excluded	87 (100%)[c]	Excluded
O'Neill, 2011	2008-2009	Retrospective	United States	Surgeons, SS	110	56/54	43 (17)/ 45 (18)	35 (63%)/35 (65%)	6 (no SD)/7 (no SD)	*na*	*na*	*na*	*na*	21 (38%)/ 23 (43%)	35 (62%)/31 (57%)	22 (40%)/28 (52%)
Stad-houder, 2008	1991-2002	Retrospective	The Nether-lands	Schools, MS	190	95/95	38.5 (18-84)/ 37.1 (18-79)	50 (52%)/64 (67%)	74.4 (no SD)[c]	*na*	46 (48%)/ 42 (44%)	27 (28%)/ 26 (27%)	22 (24%)/ 27 (29%)	61 (32.1%)[c]	63 (33.2%)/66 (34.7%)	23 (25%)/37 (39%)

Na, not reported/not measured; Excluded, patients with this characteristic were excluded for the study.
[a]What level of comparison (high to low = schools/expertise centers, departments, surgeons), SS single center, MS, multicenter.
[b]Injury severity score.
[c]Only overall value available, no distinction between case and control groups.

Table 2. Baseline Characteristics of Included Studies With a Natural Experiment Design.

MINORS Quality Assessment of Included Studies in a Systematic Review of Natural Experiments in Spinal Trauma Surgery				
Citeria	**Stadhouder et al 2008**	**O'Neill et al 2011**	**Erichsen et al 2020**	**Myers et al 2021**
A clearly stated aim	2	2	2	1
Inclusion of consecutive patients	1	2	2	2
Prospective collection of data[a]	0	0	0	0
Endpoints appropriate to the aim of the study	2	2	2	2
Unbiased assessment of the study end-point	2	0	0	0
Follow-up period appropriate to the aim of the study	2	2	2	0
Loss to follow-up less than 5%[a]	0	0	0	0
Prospective calculation of the study size[a]	0	0	0	0
Additional criteria for comparative studies				
An adequate control group	2	2	2	2
Contemporary groups	2	2	2	2
Baseline equivalence of groups	2	2	1	2
Adequate statistical analyses	2	2	1	1
Total MINORS score	17	16	16	12
All items are scored 0 (not reported/not applicable), 1 (reported but inadequate) or 2 (reported and adequate) [a]All included studies are retrospectively designed, scores range from 0-24 for comparative studies and 0-18 for retrospective comparative studies.				

Table 3. MINORS Score.

tance practice strategies and exclusion rates.[31,32] The authors conclude that the study demonstrates that the natural experiment design is suitable for comparing patient outcome between two different surgical specialties (schools) in the same hospital.[31,32]

DISCUSSION

In this systematic review on the methodology of natural experiments in spinal trauma, only four papers were found that used this methodology in 18 years of spinal trauma research.[13,31-33] Topics of the four included papers differed: open vs percutaneous placement of pedicle screws in A3 fractures, differences in management of isolated spinal fractures between neurosurgeons and orthopedic surgeons on call, the use of intrawound vancomycin powder to reduce surgical site infections in spinal trauma posterior fixation and operative vs non-operative treatment in thoracolumbar spinal fractures.[13,31-33] These are all relevant topics but in the spinal trauma community one can think of several other issues where clinical equipoise exists. Examples include conservative or operative treatment of AO classification A3 or A4 fractures,[34] treatment strategy of C2 fractures in the elderly[35] and timing of intervention in patients with spinal cord injury.[36] For this matter natural experiments can be of value since within spinal trauma treatment, the different schools and treatments are common practice already.[37-40] Therefore, with this paper we aim to increase the knowledge within the spinal community about natural experiments design and its promising potential in clinically meaningful research.

The development of prospective trauma databases can be an added value in performing natural experiments.[41] As are the common practice of Electronic Medical Records (EMR) in hospitals,[42,43] and Patient Reported Outcome Measurements (PROMS) prospectively gathered in specific patient groups.[44] In the included retrospective natural experiment study of Stadhouder et al[13] demographic and clinical data were not up to date. This was mostly due to the longer follow up period of 2-12 years. Gathering clinical and follow up data in a retrospective manner required a huge effort leading to a follow-up percentage of 79%. The longer follow up period can lead to attrition bias when the number of drop outs differ between the two groups. With longer follow up the number of dropouts will increase but there is no recognized dropout rate that is considered acceptable.[45] For analysis of results of natural experiments, as in RCT data, there is no accepted specific strategy that deals with drop outs or loss to follow up.[45] Results therefore should be carefully interpreted when there is a high and difference between groups number of drop outs.[45,46]

A study performed by the Canadian Orthopedic Trauma Society showed that the average time of presentation of concept to presentation of an RCT took almost 10 years.[47] A review by Leatherdale et al on natural experiments in the public health domain, where natural experiments are more common, mentioned that one of the three core strengths of natural experiments is 'creating timely evidence'.[48] Van de Wall also noted that one of the differences between randomized clinical trials and natural experiments/traditional observational studies is that the latter are often fast in their time frame since most patients are already included.[18] We observed in the included papers that the average time from data gathering to publication in the four studies included was 3.75 years (3-5 years).[13,31-33] A shorter duration of study time can be a contributing factor to conducting research

in quickly developing specialties as orthopedic trauma and spine surgery.[18,49]

Two included papers were published more than 10 years ago,[13,33] two papers more recently.[31,32] We think that natural experiments in clinical situations where equipoise is present have a promising future in trauma research. In this sense the total amount of four papers published utilizing some form of natural experiment in spinal trauma is disappointing. A possible explanation can be that this concept is not well known yet among spinal trauma researchers/surgeons. Another explanation might be that authors describe the method of a natural experiment inadequately, contributing to the difficulty of identifying a natural experiment. The Natural Experiments Study Group (NEXT Study Group) is an international nonprofit collaboration of clinical researchers in the field of emergency and (orthopedic) trauma surgery.[50] They so far published four relevant papers with a natural experiment as methodology and more studies are being conducted and soon to be published.[27,28,51,52] One study showed that with a natural experiment design on rib fixation there was no difference in outcome between nonoperative and operative treated patients.[27] The inclusion was finished one year earlier than predicted and took three years.[27] Before this publication, a RCTwas conducted in Australia which took four years and where almost half of the eligible patients refused to participate in this study.[53] It shows the difficulties of conducting RCT's in a trauma/surgery patient population. Also, the result showed no difference in outcome between operative and nonoperative patients,[53] comparable with the natural experiment paper.[27] Both articles impacted the current clinical practice in our hospital and resulted in an 80% decrease of surgical rib fixations. Currently surgical rib fixation is only performed in case of traumatic flail chest injuries and/or when difficulties in the weaning process of mechanical ventilation are present.

The MINORS criteria were developed as a methodological index for non-randomized studies to assess the quality of studies.[29] It comprises twelve items with a maximum score of 24, that applies to meticulously designed RCT's.[29] The studies included in our paper had a score of respectively 12, 16, 16 and 17 points.[13,31-33] Since all studies were retrospective, 3 of 4 studies were not blinded for outcome, and loss to follow up <5% is difficult to achieve in a trauma population, weconsider the quality of the natural experiment studies high as compared to other non-natural experiment retrospective comparative studies.

A systematic review of 12 comparative studies published by Phan et al[54] in 2015 on percutaneous vs open procedures in spinal fractures concluded that percutaneous screws were associated with shorter operative time and hospital stay, reduced intraoperative blood loss and reduced infection rates. They also stated that: "given the lack of robust clinical evidence, these findings warrant verification in large prospective registries and randomized trials."[54] Another more recent systematic review by Sathish et al evaluated 96 systematic reviews published in spine surgery.[55] Reviews were scored by the AMSTAR score (A measurement Tool to Assess systematic Reviews),[56] PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)[57] and MOOSE (Meta-analyses Of Observational Studies in

Epidemiology).[58] The authors concluded that there is improvement in methodological quality of reviews and meta-analysis but a substantial proportion of critical flaws remain. To our opinion, this shows the difficulties in interpreting results of comparative studies and reviews in trauma and spinal surgery and one can argue if results of these studies have additive scientific value. Natural experiments are more susceptible to confounding and bias, but when designed appropriately, it is possible to have robust internal and external validity and evidence.[48] As stated in a previous published paper on natural experiments we suggest to collect data on key prognostic patient factors, either prospective or retrospective.[18] Further, it is important to correct for confounding by stratification, regression adjustment or matching.[18] Another solution is to use clinical equipoise as an inclusion criterion.[10,18–20,45,51] Eligible data is presented to an independent expert panel, blinded for the actual treatment and the expert panel should be representative of the two schools that are compared.[10,18,51] In this review one paper used an expert panel,[13] the other three studies did not.[31–33]

When reviewing the 189 full text papers for inclusion in our review we noticed a high number of papers with a historical comparison group (see Figure 1). One of the MINORS criterion (No 10: Contemporary groups: control and study group should be managed during the same time period) considers a historical control group as less valid. This was also reported in a publication by Agabegi et al.[45] They describe that historical controls should be used with caution because of differences in in- and exclusion criteria. Treatment techniques may have improved over time and results might be a reflection of this improvement instead of a treatment effect.[45] Also it is unknown if patient and treatment factors study of controls and research subjects were similar in the time span of the study. We therefore excluded these studies.

To conclude, of the 2483 papers published on spinal trauma in the last 14 years only four papers had a natural experiment design. These papers were of high quality according to the MINORS criteria. This methodology has, to our opinion, a high potential in trauma and spinal trauma research to address difficult clinical problems in a relative short time span. We hope this systematic review will improve the attention for natural experiment designs in spinal trauma and trauma surgery.

REFERENCES

1. **Altman DG, Schulz KF, Moher D, et al.** CONSORT GROUP Consolidated Standards of Reporting Trials. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med. 2001;134(8):663-694.*

2. **Kaptchuk TJ.** The double-blind, randomized, placebo-controlled trial: gold standard or golden calf? *J Clin Epidemiol. 2001;54(6):541-549.*

3. **Groenwold RH.** [Three types of bias: distortion of research results and how that can be prevented]. *Ned Tijdschr Geneeskd. 2013;157(40):A6497.*

4. **Meakins JL.** Innovation in surgery: the rules of evidence. *Am J Surg. 2002;183(4):399-405.*

5. **Pollock AV.** Surgical evaluation at the crossroads. *Br J Surg. 1993;80(8):964-966.*

6. **Stirrat GM, Farrow SC, Farndon J, Dwyer N.** The challenge of evaluating surgical procedures. *Ann R Coll Surg Engl. 1992; 74(2):80-84.*

7. **Bonenkamp JJ, Songun I, Hermans J, et al.** Randomised comparison of morbidity after D1 and D2 dissection for gastric cancer in 996 Dutch patients. *Lancet. 1995;345(8952):745-748.*

8. **Chalmers TC.** Randomization of the first patient. *Med Clin. 1975;59(4):1035-1038.*

9. **McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D.** Randomised trials in surgery: problems and possible solutions. *BMJ. 2002;324(7351):1448-1451.*

10. **Stadhouder A, Öner FC, Wilson KW, et al.** Surgeon equipoise as an inclusion criterion for the evaluation of nonoperative versus operative treatment of thoracolumbar spinal injuries. *Spine J. 2008;8(6):975-981.*

11. **Sibai T, Carlisle H, Tornetta P 3rd.** The darker side of randomized trials: recruitment challenges. *J Bone Joint Surg Am. 2012;94(Suppl 1):49-55.*

12. **Robinson NB, Fremes S, Hameed I, et al.** Characteristics of randomized clinical trials in surgery from 2008 to 2020: a systematic review. *JAMA Netw Open. 2021;4(6):e2114494.*

13. **Stadhouder A, Buskens E, de Klerk LW, et al.** Traumatic thoracic and lumbar spinal fractures: operative or nonoperative treatment: comparison of two treatment strategies by means of surgeon equipoise. *Spine. 2008;33(9):1006-1017.*

14. **Benson K, Hartz AJ.** A comparison of observational studies and randomized, controlled trials. *N Engl J Med. 2000;342(25): 1878-1886.*

15. **Concato J, Shah N, Horwitz RI.** Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med. 2000;342(25):1887-1892.*

16. **Ioannidis JP, Haidich AB, Pappa M, et al.** Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA. 2001;286(7):821-830.*

17. Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet. 2004;363(9422):1728-1731.*

1**8. van de Wall BJM, Stadhouder A, Houwert RM, Öner FC, Beeres FJP, Groenwold RHH**, NEXT Study Group. Natural experiments for orthopaedic trauma research: an introduction. *Injury. 2023;54(2):429-434.*

19. Khullar D, Jena AB. "Natural experiments" in health care research. *JAMA Heal For. 2021;2(6):e210290.*

20. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med. 1987;317(3):141-145.*

21. Houwert RM, Beks RB, Dijkgraaf MGW, et al. Study methodology in trauma care: towards question-based study designs. *Eur J Trauma Emerg Surg. 2021;47(2): 479-484.*

22. Fanuele J, Koval KJ, Lurie J, Zhou W, Tosteson A, Ring D. Distal radial fracture treatment: what you get may depend on your age and address. *J Bone Joint Surg Am. 2009;91(6): 1313-1319.*

23. Koval KJ, Lurie J, Zhou W, et al. Ankle fractures in the elderly: what you get depends on where you live and who you see. *J Orthop Trauma. 2005;19(9):635-639.*

24. Cremer OL, van Dijk GW, van Wensen E, et al. Effect of intracranial pressure monitoring and targeted intensive care on functional outcome after severe head injury. *Crit Care Med. 2005;33(10):2207-2213.*

2**5. Wijdicks FJ, Houwert M, Dijkgraaf M, et al.** Complications after plate fixation and elastic stable intramedullary nailing of dislocated midshaft clavicle fractures: a retrospective comparison. *Int Orthop. 2012;36(10):2139-2145.*

26. Beeres FJP, Hallensleben NDL, Rhemrev SJ, et al. Plate fixation of the proximal humerus: an international multicentre comparative study of postoperative complications. *Arch Orthop Trauma Surg. 2017;137(12):1685-1692.*

27. Hoepelman RJ, Beeres FJP, Beks RB, et al. Non-operative vs. operative treatment for multiple rib fractures after blunt thoracic trauma: a multicenter prospective cohort study. *Eur J Trauma Emerg Surg. 2023;49(1):461-471.*

28. Hoepelman RJ, Minervini F, Beeres FJP, et al. NEXT study group. Quality of life and clinical outcomes of operatively treated patients with flail chest injuries: a multicentre prospective cohort study. *Front Surg. 2023;10:1156489.*

2**9. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J.** Methodological index for non-randomized studies (minors): Stadhouder et al. development and validation of a new instrument. *ANZ J Surg. 2003;73(9):712-716.*

30. Ochen Y, Beks RB, van Heijl M, et al. Operative treatment versus nonoperative treatment of Achilles tendon ruptures: systematic review and meta-analysis. *BMJ. 2019;364:k5120.*

31. Erichsen CJ, Heyde CE, Josten C, et al. Percutaneous versus open posterior stabilization in AOSpine type A3 thoracolumbar fractures. *BMC Muscoskel Disord. 2020;21(1):74.*

32. Myers M, Hall S, Sadek AR, et al. Differences in management of isolated spinal fractures between neurosurgery and orthopaedics: a 6-year retrospective study. *Br J Neurosurg. 2021;35(1):68-72.*

33. O'Neill KR, Smith JG, Abtahi AM, et al. Reduced surgical site infections in patients undergoing posterior spinal stabilization of traumatic injuries using vancomycin powder. *Spine J. 2011; 11(7):641-646.*

34. Blixt S, Mukka S, F¨orsth P, Westin O, Gerdhem P, SunBurst Study Group. Study protocol: the SunBurst trial-a register-based, randomized controlled trial on thoracolumbar burst fractures. *Acta Orthop. 2022;93:256-263.*

35. Fan L, Ou D, Huang X, et al. Surgery vs conservative treatment for type II and III odontoid fractures in a geriatric population: a meta-analysis. *Medicine (Baltim). 2019; 98(44):e10281.*

36. Hosman AJF, Barbagallo G, Popescu EC, et al. The SCI-POEM Study Group. Neurological recovery after early versus delayed surgical decompression for acute traumatic spinal cord injury. *Bone Joint Lett J. 2023;105-B(4):400-411.*

37. Lonjon G, Grelat M, Dhenin A, et al. Survey of French spine surgeons reveals significant variability in spine trauma practices in 2013. *Orthop Traumatol Surg Res. 2015;101(1):5-10.*

38. Öner C, Rajasekaran S, Chapman JR, et al. Spine trauma-what are the current controversies? *J Orthop Trauma. 2017;31(Suppl 4):S1-S6.*

39. Balas M, Pr¨ommel P, Nguyen L, et al. Reality of accomplishing surgery within 24 hours for complete cervical spinal cord injury: clinical practices and safety. *J Neurotrauma. 2021;38(21): 3011-3019.*

40. Stadhouder A, Faraj S, Öner C. Management of spinal trauma patients: a national survey in The Netherlands. *Acta Orthop Belg. 2019;85(1):40-46.*

41. Driessen MLS, Sturms LM, Bloemers FW, et al. The Dutch nationwide trauma registry: the value of capturing all acute trauma admissions. *Injury. 2020;51(11):2553-2559.*

42. Bots SH, Groenwold RHH, Dekkers OM. Using electronic health record data for clinical research: a quick guide. *Eur J Endocrinol. 2022;186(4):E1-E6.*

43. Uslu A, Stausberg J. Value of the electronic medical record for hospital care: update from the literature. *J Med Internet Res. 2021;23(12):e26323.*

44. Kluzek S, Dean B, Wartolowska KA. Patient-reported outcome measures (PROMs) as proof of treatment efficacy. *BMJ Evid Based Med. 2022;27(3):153-155.*

45. Agabegi SS, Stern PJ. Bias in research. *Am J Orthop (Belle Mead NJ). 2008;37(5): 242-248.*

46. Agabegi SS, Antekeier DP, Crawford AH, Crone KR. Postlaminectomy kyphosis in an achondroplastic adolescent treated for spinal stenosis. *Orthopedics. 2008;31(2):168.*

47. Axelrod D, Trask K, Buckley RE, Johal H. The Canadian orthopaedic trauma society. *Bone Joint Lett J. 2021;103-B(5): 898-901.*

4**8. Leatherdale ST.** Natural experiment methodology for research: a review of how different methods can support real-world research. *Int J Soc Res Methodol. 2019;22(1):19-35.*

49. Liu Y, Lee MG, Kim JS. Spine surgery assisted by augmented reality: where have we been? *Yonsei Med J. 2022;63(4): 305-316.*

50. *Available from: https://www.next-studygroup.org/.*

51. Hoepelman RJ, Ochen Y, Beeres FJP, et al. Let's agree to disagree on operative versus nonoperative (LADON) treatment for proximal humerus fractures: study protocol for an international multicenter prospective cohort study. PLoS One. 2022; 17(2):e0264477.

52. Beks RB, de Jong MB, Sweet A, et al. Multicentre prospective cohort study of non-operative versus operative treatment for flail chest and multiple rib fractures after blunt thoracic trauma: study protocol. *BMJ Open. 2019;9(8):e023660.*

53. Marasco SF, Balogh ZJ, Wullschleger ME, et al. Rib fixation in non-ventilator-dependent chest wall injuries: a prospective randomized trial. *J Trauma Acute Care Surg. 2022;92(6): 1047-1053.*

54. Phan K, Rao PJ, Mobbs RJ. Percutaneous versus open pedicle screw fixation for treatment of thoracolumbar fractures: systematic review and meta-analysis of comparative studies. *Clin Neurol Neurosurg. 2015;135:85-92.*

55. Sathish M, Eswar R. Systematic reviews and meta-analysis inspine surgery-how good are they in methodological quality? a systematic review. *Global Spine J. 2021;11(3):378-399.*

56. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol. 2007;7:10.*

57. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med. 2009;6(7): e1000097.*

58. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. *JAMA. 2000;283(15):2008-2012*

Supplementary materials table A.
Original search string (01-07-2021) and final update (30-03-2022)

Database	#	Syntax	Results
MEDLINE (n = 1821)	1	((spinal or spine) adj (fracture* or injur*) adj3 (trauma* or burst)).ti,ab,kw.	827
	2	*Spinal Fractures/pp, su, th and (burst or trauma*).ti,ab,kw.	1832
	3	*Thoracic Vertebrae/in	2482
	4	1 or 2 or 3	4493
	5	exp Osteoporosis/ or osteoporos*.ti,ab,kw.	91707
	6	4 not 5	4135
	7	(exp Pediatrics/ or adolescent/ or exp child/ or exp infant/ or (child* or pediatr* or paediatr* or adolescen* or youth*).ti,ab,kw.) not exp Adult/	2355226
	8	6 not 7	3856
	9	exp Neoplasm Metastasis/ or metasta*.ti,ab,kw.	621026
	10	8 not 9	3818
	11	comment/ or editorial/ or letter/ or case reports/ or (letter or comment* or editorial or case report).ti.	4071872
	12	10 not 11	2855
	13	((exp Animals/ or exp Animal Experimentation/ or exp models, animal/ or (animal* or rat or rats or mice or mouse or dog or dogs or pig or pigs or swine or swines or cow or cows or monkey or monkeys or goat or goats or horse or horses).ti,ab,kw.) not (Humans/ or human*.ti,ab,kw.)) or animal*.ti.	4897417
	14	12 not 13	2780
	15	limit 14 to yr = "2004 -Current	1774
	16	limit 15 to (dutch or english or german)	1572
Last update 30-3-2023	1	orginal search repeated	1821

EMBASE (n= 1287)	1	((spinal or spine) adj (fracture* or injur*) adj3 (trauma* or burst)).ti,ab,kw.	1120
	2	*spine fracture/su, th and (burst or trauma*).ti,ab,kw.	1519
	3	1 or 2	2545
	4	exp osteoporosis/ or osteoporos*.ti,ab,kw.	170682
	5	3 not 4	2390
	6	exp metastasis/ or metasta*.ti,ab,kw.	1000273
	7	exp metastasis/ or metasta*.ti,ab,kw.	2374
	8	(exp child/ or juvenile/ or exp adolescent/ or exp infant/ or exp pediatrics/ or (child* or pediatr* or paediatr* or adolescen* or youth*).ti,ab,kw.) not (adult/ or exp aged/ or middle aged/)	3108037
	9	7 not 8	2207
	10	letter/ or editorial/ or note/ or case report/ or conference paper/ or (letter or comment* or editorial or case report).ti.	5885096
	11	9 not 10	1786
	12	(exp animal/ or exp animal experiment/ or exp animal model/ or (rat or rats or mice or mouse or dog or dogs or pig or pigs or cow or cows or monkey or monkeys or goat or goats or horse or horses or ape or apes or gorilla or gorillas).ti,ab,kw.) not (human/ or human*.ti,ab,kw.)	5738063
	13	11 not 12	1733
	14	limit 13 to yr="2004 -Current"	1359
	15	limit 14 to (dutch or english or german)	1202
	16	limit 14 to (dutch or english or german)	175
	17	15 not 16	1027
Last update 30-3-2023	1	orginal search repeated	1287
CENTRAL (n= 570)	1	((spinal or spine) near/3 (fracture* or injur*) near/3 (trauma* or burst)):ti,ab,kw	357
	2	MeSH descriptor: [Spinal Fractures] explode all trees and with qualifier(s): [physiopathology - PP, surgery - SU, therapy - TH]	295
	3	#1 or #2	640
	4	(osteoporos*):ti,ab,kw	11031
	5	#3 not #4	508
	6	(metasta*):ti,ab,kw	44239
	7	#5 not #6	504
	8	((child* or pediatr* or paediatr* or adolescen* or youth*) not adult*):ti,ab,kw	136919
	9	#7 not #8 in Trials	490
Last update 30-3-2023	1	orginal search repeated	570

Supplementary table B.
Quality assessment in compliance with the MINORS criteria in a systematic review of natural experiments in spinal trauma surgery

Criteria	Reported and adequate (2)	Reported but in adequate (1)	Not reported (0)
A clearly stated aim	Aim stated clearly	Aim unclear but reported	Not reported
Inclusion of consecutive patients	Inclusion criteria and consecutive inlcusion reported	Inclusion criteria/consecutive inclusion unclear	Not reported
Prospective collection of data*	Data collected prospectively and clearly reported	Prospective design, unclearly reported	Not reported/not applicable (retrospective design)
Endpoints appropriate to the aim of the study	Endpoints appropriate and reported clearly	Endpoints inappropriate or reported unclearly	Not reported
Unbiased assessment of the study endpoint	Blinded assessment of outcomes	Reason for not blinding reported	Not reported
Follow-up period appropriate to the aim of the study	Follow-up appropriate and reported clearly	Follow-up inappropriate or reported unclearly	Not reported
Loss to follow-up less than 5%*	≤ 5% and reported	≥ 5% and reported	Not reported/not applicable (retrospective design)
Prospective calculation of the study size*	Sample size calculation and power analysis performed	Only sample size or power analysis performed	Not reported/not applicable (retrospective design)
Additional criteria for comparative studies			
An adequate control group	Natural experiment design	Not applicable (exclusion)	Not applicable (exclusion)
Contemporary groups	Groups managed in the same time period	Not applicable (exclusion)	Not applicable (exclusion)
Baseline equivalence of groups	Comparable baseline characteristics	Incomparable baseline characteristics	Not reported
Adequate statistical analyses	Statistical analysis reported and repeatable	Inadequately reported statistical analysis	Not reported

All items are scored 0 (not reported/not applicable), 1 (reported but inadequate) or 2 (reported and adequate). Overall scores of the MINORS-tool range from 0-24 for comparative studies.
*Overall scores for retrospective studies range from 0-18 as prospective collection of data, loss to follow-up and prospective calculation of the study size are not applicable

9

Summary, discussion and future perspectives

This thesis focused on clinical research in spinal trauma patients concentrating on a novel study methodology alternative to the Randomized Control Trial (RCT) design. In clinical research, the gold standard of performing research is still considered an RCT. Nevertheless, the theoretical and practical difficulties of this methodology in acute surgical problems demand attention.[1,2]

The present thesis starts with an introductory overview discussing controversies in management of spinal trauma patients, start with a historical overview of conservative treatment complemented by operative treatment since the 1980s.

In Chapter 2 we present the results of a national survey amongst surgeons responsible for spinal trauma patients in different trauma centers in the Netherlands. In 1999, the Dutch government, after an alarm raised by Dutch trauma surgeons, installed eleven trauma regions, each with a Level 1 trauma center where polytrauma patients are taken care of with all the relevant facilities and specialties 24/7 present. The eleven trauma centers cover the whole of the country. The goal was to improve the care for trauma patients by enhancing teaching and training, reorganization and regionalization of all healthcare partners involved in trauma care.[3] In 2012, this system was evaluated for a single level trauma center and showed that the odds ratio for in-hospital mortality decreased to 0.74., from 0.84 in this specific trauma region before centralization.[4] Also, the Injury Severity Score of patients increased in the level 1 center and more patients with chest and spinal injuries presented at the ER.[4] This resulted in a national trend. In another level 1 trauma center, the number of spinal injuries presented increased in the years from 2007 until 2016, with also increasing ISS scores.[5] Following centralization of trauma victims, spinal trauma care was also more concentrated in trauma centers. Considering concerns about the variable quality of spine trauma cases, we conducted a survey in 2014 among all level 1 trauma centers (L1TCs) to understand the practical organization and management of spinal trauma patients in these centers. Overall, there were differences between trauma regions in the composition of spine trauma care units, different use of classification systems of spinal fractures, no consensus on timing of surgery and lack of specific protocols.[6] In succession of our paper, Fransen et al performed a survey among 23 Emergency Medical Services (EMSs) and all 11 Level 1 trauma centers in the Netherlands assessing the organization of pre-hospital and acute management of traumatic Spinal Cord Injury patients from 2012-2014.[7] The aim of this study was to investigate the consistency of pre- and in-hospital acute phase care for patients with (suspected) tSCI among EMSs and level-1 trauma centers in the Netherlands. Results, similar to our survey, showed a large variation in pre- and in-hospital care of tSCI patients between the EMSs and L1TCs. The authors conclude that this survey shows the need for standardization of assessments and the development of guidelines recognized by all pre-hospital and in-hospital healthcare providers who are involved in the acute phase treatment of tSCI patients.[7] To follow up on these results the past years several steps were taken in the Netherlands. Hietbrink et al showed that in the past 20 years mortality levels of severely

injured trauma patients diminished with 50% and when adjusted for age even 75%.[8] In addition, in 2007 Rutges performed a systematic review on timing of thoracolumbar spinal injuries which showed that in these often severely injured patients early fracture fixation is associated with less complications, shorter hospital length of stay (HLOS) and ICU length of stay (ILOS).[9] A more recent retrospective study of a French level 1 trauma center showed that the main part of the eighty-three polytrauma patients with spinal injuries were operated within 48h, 62% even within 24 hours. They advise to consider early spinal surgery in polytrauma patients, preferable with less invasive procedures, with an emphasis on thoracic fractures as they are more prone to neurological deterioration.[10] A recent publication by Dijkink showed that 70% of severely injured trauma patients are presented to a Trauma Center in the Netherlands, but as a result it also implies that 30% are not. There seems to be still room for improvement.[11] For the future, research should focus on enhancing pre-hospital triage. Another controversial next step is the discussion if more centralization is desired or required for (spinal) trauma patients. Does more centralization lead to better care for the severely injured patients in the Netherlands or do we have an optimal distribution of patients and facilities as it is now? Future research will need to answer this question. Another aspect of improvement mentioned by caregivers in our survey was the development of a general classification system for spinal fractures. Thanks to the efforts of the AO Knowledge Forum a unified classification system comprising all four anatomical regions (upper cervical, subaxial cervical, thoracolumbar and sacral) is developed for spinal fractures in the last 10 years.[12-15] This systematic classification system facilitates communication between surgeons, provide a tool for setting up best treatment protocols with recommendations for optimal treatment and thus improving spinal trauma care. This AO spinal trauma classification system is already widely adapted by spine surgeons around the world. One more controversial subject in our survey was the timing of treatment of patients with neurological impairment, with wide variations among L1TCs. The best treatment of traumatic spinal cord injury (tSCI) has been an ongoing and much debated dilemma in spinal research, but without doubt a truly relevant topic. Since our survey more than 24000 papers have been published on tSCI in the broadest meaning (PubMed) showing the importance of the matter. Focusing on timing of surgical intervention in patients with traumatic spinal cord injury, a narrative review in 2019 mentioned that there appear to be different patterns for spontaneous recovery in cervical, thoracic, and thoracolumbar tSCI. Neurological recovery by surgical decompression of the spinal cord within 24 hours seems particularly beneficial in patients with complete cervical tSCI. In thoracic or lumbar tSCI this is less clear.[16] A recent European multicenter prospective observational study divided patients with tSCI in early (12h<) and late (12h>) decompression of neural structures. Seventeen centers participated but unfortunately the paper did not elaborate on the specific treatment considerations in the different hospitals. Treatment was based on the judgement of the treating spinal

surgeon. 159 patients were included in the early group and 135 patients had a later decompression. Patients in the early decompression group had significantly more severe neurological impairment. The results of the study did not show a significant difference in lower extremity motor score improvement between early and late decompression after propensity score analysis.[17] This commendable study shows the immense difficulties in performing good clinical spinal trauma research, i.e. the heterogeneity of the patient population spine surgeons are dealing with. Nowadays, we can still not clearly define patients that will benefit from early decompressive surgery in tSCI, which is one of the most important research questions still to answer in spinal trauma treatment.

In Chapter 3, the optimal conservative treatment of thoracolumbar spine fractures is explored. Although inclusion of patients was between 1991 and 1997 and the results seem outdated, the randomized design of this study makes it unique in spinal trauma research. Randomization created comparable groups except for the mean age of female patients with a burst fracture that was significantly higher than male patients (mean difference 21.7 years, CI 1.8-41.5). However, it took 6 years to include 133 patients in two large city hospitals in Amsterdam. Another RCT on conservative treatment of burst fractures in Canada in three spine centers included 96 patients in 7 years.[18] A study performed by the Canadian Orthopedic Trauma Society showed that the average time of presentation of concept to presentation of an RCT took almost 10 years.(19) Although the exact reasons of the (longer) inclusion period is not mentioned in this study, it demonstrates the difficulties in performing RCTs in trauma patients and pleads for alternative study methodologies as explained in Chapter 7 and 8 of this thesis. The results of the study of Bailey show that wearing a Thoraco-Lumbar Contact Orthosis (TLSO) or having a functional treatment after a burst fracture are similar after 2 years.[19] The inclusion criteria of our study were patients younger than 80 years of age, fractures with less than 50% loss of anterior height, with less than 30% reduction of the spinal canal, without signs of posterior element involvement. The study of Bailey included isolated A3 burst fractures according to Magerl classification between T10 and L3 with kyphotic deformity less than 35°, neurologically intact, 16 to 60 years of age, and who were recruited within 3 days of injury. Both studies try to identify in their inclusion criteria so called 'stable' burst fractures that can be treated conservatively. Stable meaning that when mobilizing no substantial further deformity takes place, patients have an acceptable pain level and can go back to their pre-trauma activities. In our study one patient (4%) required surgery because of progressive deformity and pain in the burst group, in the study of Bailey 6 patients (6%) required surgery because of radicular pain (n = 2), severe mechanical back pain on ambulation (n = 3), or severe kyphotic deformity during follow up (n = 1). Identifying stable burst fractures or rather analyzing which burst fractures might fail conservative treatment was the research subject of a recent systematic review of Tan et al in 2022.[20] "Failure" of conservative management was defined as the need for surgical management within 6 months of injury

after initial conservative management. The review included 11 studies (3 RCT's, 8 cohort studies) and pooled analysis of all 601 patients from the 11 included studies showed that the rate of failure in the conservative management of thoracolumbar burst fractures is 9.2% (95% CI: 4.5%-13.9%).[20] Older age, admission kyphotic angle, admission residual canal area and interpedicular distance were factors the authors considered risk factors for failure of conservative treatment and they advise to further investigate in prospective studies these factors to identify the subset of patients prone to failure of conservative management. They also recommend that surgical management should be carefully considered in patients with the above risk factors. There is, however, a significant heterogeneity between studies and the reasons underlying the conversion to operative treatment is less well defined from the included studies.[20] Thus, conservative treatment of patients with a spinal compression or burst fracture gives satisfying results but a substantial percentage of patients do require surgery and identifying these patients is a challenge for spine surgeons. The results of our study also show some disturbing features. According to the VAS, 20 (18%) of the 108 patients with compression fractures suffered from moderate or severe back pain at long-term follow-up; 12 patients had an Oswestry Disability Score (ODI) greater than 40 indicating moderate disability. Of the 25 patients with burst fracture, 3 (12%) had chronic moderate pain and one patient was operated on because of severe persistent pain.[1,21] In this study we used the VAS and ODI as clinical outcome measurements, but these measurements are used in patients with degenerative spinal conditions and one can question the relevance of these outcome scores in spinal trauma patients.

In Chapter 4 we explore the existing outcome measurements of spinal trauma patients at that time. As discussed in chapter 3, issues specific to spinal trauma patients may not be adequately measured by generic outcome measures, or by "spine-specific" outcome measures that were designed for common chronic spinal conditions.[22] We used the WHO's International Classification of Functioning, Disability and Health (ICF) as an expansive theoretical underpinning for newly developed measures targeting, among others, trauma patients.[22] The ICF is a comprehensive and universally accepted framework to describe and classify individuals' functioning, disability, and health. The classification is organized into the components body functions (b), body structures (s), activities and participation (d), and environmental factors (e). As a classification system, the ICF provides alphanumeric codes for each of the ICF categories or functioning domains, arranged in a hierarchical fashion in different levels.[23] The WHO introduced a subset of ICF domains and constructs to generate condition-specific "ICF core sets". In the review we identified commonly used outcome measures that are primarily used in trauma populations: the Functional Independence Measure (FIM), the GOS, the SF-36, and the EQ-5D (standardized instrument of the EuroQol Group), the Musculoskeletal Function Assessment (MFA) and the Health Utilities Index Mark 3 (HUI 3). The only injury-specific outcome measure we identified is

the Functional Capacity Index (FCI). We found that the tSCI patient population is treated as a separate and distinct population in the literature. Similar to the trauma outcomes field, we found that there was no consensus about which outcome assessment tools are to be used in tSCI patients. Outcome measures that are frequently used in tSCI populations included: the Walking Index for Spinal Cord Injury (WISCI), the Spinal Cord Independence Measure (SCIM), the FIM, the SF-36, and the GOS. Psychological outcome and well-being measurement is also an important aspect of polytrauma and spinal trauma outcome. The psychological outcome assessments commonly used in both trauma and tSCI populations are the Hospital Anxiety and Depression Scale (HADS), the Beck Depression Inventory (BDI), and the Hamilton Depression Rating Scale (HDRS). We proposed fourteen domains of relevance in our opinion: mental health, sensory and pain, genitourinary and reproductive, neuromusculoskeletal, nervous system, movement related structures, mobility, self-care, domestic life, interpersonal relationships, major life areas, community-social-civic life, support and relationships, services systems and policies. Following these arguments, the AOSpine Knowledge Forum Trauma in 2016 decided to initiate a project to develop and validate disease-specific outcome instruments for spine trauma patients and health professionals.[23] They performed a systematic search and concluded that there is a great diversity in the use and content of outcome measures to evaluate the function and health of spine trauma patients, with seventeen different outcome measures linked to 57 unique ICF categories. These results support the hypothesis that there is no agreement on outcome assessment in spine trauma research, and that there is no outcome instrument designed or validated for this specific patient population.[23] In addition to this, a consensus meeting was organized with an international panel of 11 spine experts that elected 25 ICF categories as core categories for patient-reported outcome measurement. The group also agreed to use the Numeric Rating Scale 0–100 as response scale in the future universal outcome instrument.[24] Successively the AO Spine Patient Reported Spine Trauma (AOSpine PROST) instrument was developed and published in 2017.[25] A reliability and validation study of the Dutch version of the AOSpine PROST showed very satisfactory results among 163 patients. A unique approach in AOSpine PROST is asking patients to recall their pre-injury level of health, more specifically to compare their current function (0) with their pre-trauma level of function (100), this is very different from other patient outcome measurement scores.[26] A long-term reliability and validity study in 175 patients with a mean follow up of 94 months also showed good reliability and validity. Currently the AO Spine PROST has been, or is being, translated into 17 languages: Arabic, Dutch, English, Filipino, French, German, Hindi, Mandarin Chinese, Nepali, Norwegian, Portuguese, Romanian, Slovak, Spanish, Swahili, Thai, and Turkish.[27] Further developments are the Clinician Reported Outcome Spine Trauma (CROST)[28] and the applicability of PROST in patients with complete traumatic spinal cord injury.[29] To better understand the results of treatments and answer pending research questions

in spinal trauma, the spinal community should incorporate the AOSpine PROST questionnaire in their daily practice. This is especially important in the present time where health systems are suffering from lack of funding and policy makers and health insurance companies demand treatments to be evidence based with clinical and patient related outcome measures.

In Chapter 5 and 6 the concepts of equipoise and clinical equipoise are introduced. There is equipoise if we encounter a state of genuine uncertainty on the part of the clinical investigator regarding the comparative therapeutic evidence of each arm in a trial. It is generally accepted that equipoise is an ethically necessary condition in all cases of clinical research. In trials with several arms, equipoise must exist between all arms of the trial, otherwise the trial design should be modified to exclude the inferior treatment. But equipoise is also fragile, when one treatment is slightly preferred above the other by a researcher when equipoise is disturbed, and the trial is not `ethical` anymore.[30]

Clinical equipoise, as introduced by Freedman in 1987, exists when there is a genuine uncertainty within the expert medical community about the optimal treatment of a certain disease.[30]

A state of clinical equipoise is consistent with a decided treatment preference on the part of the investigators. With this concept the investigators are not the leading subject in research, they must purely recognize that their less-favored treatment is preferred by colleagues whom they consider responsible and knowledgeable.[30] This concept was the starting point of a cohort study where patients are retrospectively identified and prospectively followed where clinical equipoise exists about the best treatment for patients with a spinal fracture, operative or non/operative.[21] Because of the retrospective nature of this study, historical data could be gathered upfront to determine the feasibility concerning outcome and patient numbers needed for relevant outcome. At that time there was the unique opportunity that two university hospitals had different preferred treatment schemes in spinal trauma care in the Netherlands. These two hospitals were similar in their patient population and thus depending on where a patient was admitted, geography determined their treatment. Therefor treatment allocation was determined by factors outside the control of the investigators and confounding is reduced.[31] This clinical equipoise concept was put into practice and 636 patients were identified in two hospitals, where in 190 patients there was a discordant treatment preference by two spine surgeons as representatives of the two hospitals.[21] The two treatment groups both consisted of 95 patients and were essentially comparable. The retrospective nature of the study was challenging in gaining long-term functional results but with a follow up percentage of 79% sufficient.[1,21] Besides in hospital parameters the most important research question was how patient were doing on the long term. For this we used general questionnaires and the ODI, VAS and the Denis work scale. As discussed before, these questionnaires are not disease specific and might not give a complete and adequate picture.[26] However, overall outcome of non-operative and operative treatment

in middle-long-term follow up was comparable, although there seems to be a difference in neurologic recovery patterns in favor of operative treatment.

As we mentioned in our conclusion, future research should focus on cost-effectiveness and short-term outcome parameters to analyze what the best treatment options are for spinal trauma patients, especially in the patients where clinical equipoise exists. Since the 2009 clinical equipoise paper multiple RCT`s (or RCT protocols), systematic reviews and meta-analysis have been performed.[32-37] Results vary from no difference, to better outcome in conservative patients or either operative treatment performing better.[32-37] A recent paper of Camino-Willhuber showed that in AO A3/A4 fractures there was disagreement on treatment between an expert group of global spine surgeons and the treating surgeons. The expert panel recommended surgery for 30% of A3 injuries and 68% of A4 injuries. However, 61% of patients with both A3 and A4 fractures received surgery in the real world.[38] Up to date, it is still not clear how to treat these patients in the best way. So, what is the next step to try to eventually answer this pending research question?

In Chapter 7 and 8 the Natural Experiment concept is introduced as inclusion criterion in prospective and retrospective studies based on clinical equipoise. Natural Experiments (NEs) are implemented in public health studies already for a long time, dating back to the cholera study of John Snow in the 1850`s who postulated that water from certain contaminated water pumps was the cause of the spread of the disease.[39] One definition of a natural experiment is from the UK Medical Research Council defining NEs broadly "to include any event not under the control of a researcher that divides a population in exposed and unexposed groups".[40] Craig et al added two features to this, namely the implementation of the intervention is not dependent on whether there is a plan to evaluate the intervention and random allocation of the intervention is not feasible for ethical or political reasons. These items more likely address public health issues where the natural experiment design is more common.[40-42] The first feature can be debated; Craig stated that natural experiments should only be evaluated when data are available to use an experimental design for the evaluation and non-experimental designs should be avoided[40], but one could also argue that even weaker quality evidence derived from a non-experimental evaluation of a natural experiment may be better than no evidence at all.[41] Leatherdale published a review in 2017 on how different methods can support real-world research in Natural Experiments.[41] He advises on key strategies researchers should strive for when evaluating natural experiments. First is collecting outcome measurements before and after the natural experiment is conducted, the second is to compare an intervention group to a control group and the third is to strive to use best design possible based on the data that are available or will be collected. This was illustrated in the flow diagram of this study.[41] Bias in research can influence internal and external validity.[39] RCT's in general effectively limit internal bias by limiting selection bias. In natural experiments bias due to confounding (a situa-

tion in which the effect (or association) of an intervention on an outcome is distorted by the presence of another variable) can be limited by using experimental research designs with pre- and post-test measures as regression-based analytical modelling and using an adequate control group.[41] If the control group does not seem representative propensity scores, regression discontinuity designs or difference-in-difference models can be used to correct for any potential differences in baseline characteristics (known confounders). For future natural experiment studies, in public health but also in comparative medical studies, registries of currently available longitudinal (and ideally hierarchical) data systems could add value for enabling natural experimental studies. In spinal trauma, research methodologies encompass various approaches, each with its own set of strengths and limitations. Two prominent methodologies include RCTs and the already mentioned NEs. In spine, RCT`s are still the gold standard in clinical research but recently NEs and other alternatives to RCTs have attracted interest because they are useful in evaluating large-scale population health interventions that are not amenable to experimental manipulation.[42] In contrast, randomized interventions or treatment strategies may pose ethical concerns, particularly if there is existing evidence in spinal trauma treatment favoring one treatment over another.[2,31] RCTs often require substantial time, funding, and resources to execute properly, making them impractical for certain spinal research questions. Strict inclusion and exclusion criteria in RCTs may limit the generalizability of findings to real-world clinical settings, and RCTs may necessitate long-term follow-up to assess the sustainability of treatment effects, posing logistical challenges and potential loss to follow-up. This is also the case in the treatment of spinal fractures. Worldwide it is difficult to find individual surgeons who are both skilled and comfortable in operative or non-operative management and do not have a strong preference for one treatment method over the other for specific cases.[38] This and issues of standardization, lack of blinding, limited inclusion criteria and patient and provider preference make a surgical randomized trial in spinal fractures extremely difficult and practically impossible.[43] In this environment of a high degree of collective equipoise and concomitant high degree of individual provider certainty, a large-scale prospective observational study can benefit from equipoise concept and create a study design superior to the RCT in terms of its generalizability and assessment of effectiveness in the real world.[1,21,43] This being the natural experiment design we describe in chapter 7.[31] In 2016, an international multicenter prospective thoracolumbar burst fracture study comparing surgical versus non/surgical treatment was initiated by the AO Spine Knowledge Forum trauma.[44] 208 patients were enrolled with a thoracolumbar burst fracture from T10/L2 and treated surgically or non/surgically as decided by the treating spine surgeon. The original trauma data of this patient group was presented to twenty-two members of the AO Spine Knowledge Forum Trauma (AOSKFT) that is composed of spine trauma opinion leaders from around the world. These twenty-two spine trauma experts formed a panel of experts who would be able

to classify, analyze, and make treatment recommendations on the images from radiographic records from 183 patients included in the Spine TL A3/4 study.[44] In another paper of Dandurand the threshold of equipoise is discussed, medical ethics researchers suggest that a clinical trial is not ethical when there is agreement above 70 or 80%.[46] In our paper in chapter 7 we also discussed the clinical equipoise disagreement. A survey of Institutional Review Board members (IRB) during a conference on bioethics in Florida on collective equipoise, showed that enrolling humans is not ethical anymore when the equipoise level is higher than 80% (80:20 distribution of uncertainty). In children and elderly, IRB members require a higher level of equipoise to be comfortable to approve trials involving life-threatening situations.[47] Dandurand in their study set the equipoise percentage level at 77%, which means that there was disagreement when 17 (or less) out of twenty-two spine surgeons disagreed on treatment. Ghogawala in 2021 performed an RCT on operative treatment of cervical spondylotic myelopathy. They introduced an expert panel of fifteen surgeons who before informed consent of the patient looked at the case and advised on yes/ no to randomization and advised on ventral or posterior surgery. Clinical equipoise was defined as not met when either (1) 80% or more of panel members chose either ventral or dorsal surgery or (2) a simple majority voted against randomization. The researchers found an increased rate of consent to randomization with this study methodology.[48] This is an interesting added step in clinical research where the use of clinical equipoise of experts encourages patients to participate in an RCT. As mentioned before, to be able to reduce confounding in a natural experiment study we advise to constitute an expert panel representative of the two treatment schools. The panel is asked to decide independently on the preferred treatment for the eligible patient as if patients were presented to them in clinical practice at that moment. Patients are included if there is disagreement on treatment choice between the "schools"; they are excluded in case of agreement because there is no clinical equipoise. This ensures that the included study population consists of patients that would have received treatment A in "school A" but were in fact seen and treated according to the conviction of "school B", and vice versa (exchangeability). Our recent review of the NE methodology in spinal trauma only identified 4 articles that used the NE design in the past 19 years.[49] One article used an expert panel, something we would recommend in using the natural experiment methodology in clinical research. Although 4 articles is a disappointing number, the MINORS criteria of the 4 papers showed that the quality of these reports was high, especially compared to other retrospective comparative study designs.[49]

FINAL CONCLUSIONS AND FUTURE PERSPECTIVES

In this thesis, spinal fracture treatment and spinal fracture research methodology are addressed. The last years considerable progress has been made in classification of these fractures with the development of the AOSpine spinal injury classification. With better understanding of the trauma mechanism, patient treatment can be adjusted and improved. Furthermore, the use of outcome scores to analyze treatment outcome with the AO Patient Reported Outcome Measurement Score is a big step forward in analyzing and improving patient care. It should be the duty of the spinal community to implement these outcome scores in their daily practice. However, there still remain questions and controversies in spinal trauma care that are yet not answered or solved and we need to acknowledge that performing good clinical research in spinal trauma is a challenge. The Natural Experiment concept is introduced as a methodology based on clinical equipoise and can be an added value in trauma research. This methodology can be used with an expert panel to further reduce bias and confounding in observational research. NEs are a very promising next step in clinical research but up to now not widely known. To improve visibility of the methodology the Natural Experiments Study Group (NEXT Study Group) is founded. It is an international non-profit collaboration of clinical researchers in the field of emergency and (orthopedic) trauma surgery. The ambition of the NEXT Study Group is to contribute to the improvement of emergency and (orthopedic) trauma surgery patient care. With this NEXT Study Group initiative and future publications in peer-reviewed journals the Natural Experiment concept can be developed further, gain attention in the medical world and improve spinal trauma care. In addition, prospective registries in trauma patients are more widely used and enable another contribution to use natural experiments in (spinal) trauma care.

The Natural Experiments Study Group
(NEXT Study Group)

REFERENCES

1. **Stadhouder A, Buskens E, de Klerk LW, Verhaar JA, Dhert WA, Verbout AJ, et al**. Traumatic thoracic and lumbar spinal fractures: operative or nonoperative treatment: comparison of two treatment strategies by means of surgeon equipoise. Spine (Phila Pa 1976). *2008;33(9):1006-17*.
2. **Canseco JA, Paziuk T, Schroeder GD, Dvorak MF, Öner CF, Benneker LM, et al**. Interobserver Reliability in the Classification of Thoracolumbar Fractures Using the AO Spine TL Injury Classification System Among 22 Clinical Experts in Spine Trauma Care. Global Spine Journal. *2024;14(1_suppl):17S-24S*.
3. **ten Duis HJ, van der Werken C**. Trauma care systems in The Netherlands. Injury. *2003;34(9):722-7*.
4. **Lansink KWW, Leenen LPH**. History, development and future of trauma care for multiple injured patients in the Netherlands. European Journal of Trauma and Emergency Surgery. *2013;39(1):3-7*.
5. **den Ouden LP, Smits AJ, Stadhouder A, Feller R, Deunk J, Bloemers FW**. Epidemiology of Spinal Fractures in a Level One Trauma Center in the Netherlands: A 10 Years Review. Spine. *2019;44(10):732-9*.
6. **Stadhouder A, Faraj S, Öner C**. Management of spinal trauma patients: a national survey in The Netherlands. Acta Orthop Belg. *2019;85(1):40-6*.
7. **Fransen BL, Hosman AJ, van Middendorp JJ, Edwards M, van Grunsven PM, van de Meent H**. Pre-hospital and acute management of traumatic spinal cord injury in the Netherlands: survey results urge the need for standardisation. Spinal Cord. *2016;54(1):34-8*.
8. **Hietbrink F, Houwert RM, van Wessem KJP, Simmermacher RKJ, Govaert GAM, de Jong MB, et al**. The evolution of trauma care in the Netherlands over 20 years. European Journal of Trauma and Emergency Surgery. *2020;46(2):329-35*.
9. **Rutges JPHJ, Öner FC, Leenen LPH**. Timing of thoracic and lumbar fracture fixation in spinal injuries: a systematic review of neurological and clinical outcome. European Spine Journal. *2007;16(5):579-87*.
10. **Joubert C, Cungi PJ, Esnault P, Sellier A, de Lesquen H, Avaro JP, et al**. Surgical management of spine injuries in severe polytrauma patients: a retrospective study. British Journal of Neurosurgery. *2020;34(4):370-80*.
11. **Dijkink S, van Zwet EW, Krijnen P, Leenen LPH, Bloemers FW, Edwards MJR, et al**. The impact of regionalized trauma care on the distribution of severely injured patients in the Netherlands. European Journal of Trauma and Emergency Surgery. *2022;48(2):1035-43*.
12. **Vaccaro AR, Karamian BA, Levy HA, Canseco JA, Rajasekaran S, Benneker LM, et al**. Update on Upper Cervical Injury Classifications: The New AO Upper Cervical Spine Classification System. Clinical Spine Surgery. *2022;35(6):249-55*.
13. **Vaccaro AR, Koerner JD, Radcliff KE, Öner FC, Reinhold M, Schnake KJ, et al**. AOSpine subaxial cervical spine injury classification system. Eur Spine J. *2016;25(7):2173-84*.

14. **Vaccaro AR, Öner C, Kepler CK, Dvorak M, Schnake K, Bellabarba C, et al**. AOSpine Thoracolumbar Spine Injury Classification System: Fracture Description, Neurological Status, and Key Modifiers. Spine. *2013;38(23):2028-37*.
15. **Vaccaro AR, Schroeder GD, Divi SN, Kepler CK, Kleweno CP, Krieg JC, et al**. Description and Reliability of the AOSpine Sacral Classification System. JBJS. *2020;102(16):1454-63*.
16. **Wengel PVt, De Haan Y, Feller RE, Öner FC, Vandertop WP**. Complete Traumatic Spinal Cord Injury: Current Insights Regarding Timing of Surgery and Level of Injury. Global Spine Journal. *2019;10(3):324-31*.
17. **Hosman AJF, Barbagallo G, Middendorp JJv**. Neurological recovery after early versus delayed surgical decompression for acute traumatic spinal cord injury. The Bone & Joint Journal. *2023;105-B(4):400-11*.
18. **Bailey CS, Urquhart JC, Dvorak MF, Nadeau M, Boyd MC, Thomas KC, et al**. Orthosis versus no orthosis for the treatment of thoracolumbar burst fractures without neurologic injury: a multicenter prospective randomized equivalence trial. The Spine Journal. *2014;14(11):2557-64*.
19. **Axelrod D, Trask K, Buckley RE, Johal H**. The Canadian Orthopaedic Trauma Society. The Bone & Joint Journal. *2021;103-B(5):898-901*.
20. **Tan T, Huang MS, Rutges J, Marion TE, Fitzgerald M, Hunn MK, Tee J**. Rate and Predictors of Failure in the Conservative Management of Stable Thoracolumbar Burst Fractures: A Systematic Review and Meta-Analysis. Global Spine Journal. *2021;12(6):1254-66*.
21. **Stadhouder A, Öner FC, Wilson KW, Vaccaro AR, Williamson OD, Verbout AJ, et al**. Surgeon equipoise as an inclusion criterion for the evaluation of nonoperative versus operative treatment of thoracolumbar spinal injuries. Spine J. *2008;8(6):975-81*.
22. **Stadhouder A, Buckens CF, Holtslag HR, Öner FC**. Are existing outcome instruments suitable for assessment of spinal trauma patients? J Neurosurg Spine. 2010;13(5): 638-47.
23. **Öner FC, Jacobs WCH, Lehr AM, Sadiqi S, Post MW, Aarabi B, et al**. Toward the Development of a Universal Outcome Instrument for Spine Trauma: A Systematic Review and Content Comparison of Outcome Measures Used in Spine Trauma Research Using the ICF as Reference. Spine. *2016;41(4):358-67*.
24. **Sadiqi S, Lehr AM, Post MW, Jacobs WCH, Aarabi B, Chapman JR, et al**. The selection of core International Classification of Functioning, Disability, and Health (ICF) categories for patient-reported outcome measurement in spine trauma patients–results of an international consensus process. The Spine Journal. *2016;16(8):962-70*.
25. **Sadiqi S, Lehr AM, Post MW, Dvorak MF, Kandziora F, Rajasekaran S, et al**. Development of the AOSpine Patient Reported Outcome Spine Trauma (AOSpine PROST): a universal disease-specific outcome instrument for individuals with traumatic spinal column injury. European Spine Journal. *2017;26(5):1550-7*.
26. **Sadiqi S, Post MW, Hosman AJ, Dvorak MF, Chapman JR, Benneker LM, et al**. Reliability, validity and responsiveness of the Dutch version of the AOSpine PROST (Patient Reported Outcome Spine Trauma). European Spine Journal. *2021;30(9):2631-44*.

27. Sadiqi S, Öner FC. A disease-specific patient reported outcome instrument for spine trauma is developed, validated and available! Re: Andrzejowski et al. Measuring functional outcomes in major trauma: can we do better? European Journal of Trauma and Emergency Surgery. *2023;49(3):1605-6*.

28. Sadiqi S, de Gendt EEA, Muijs SPJ, Post MWM, Benneker LM, Holas M, et al. Validation of the AO Spine CROST (Clinician Reported Outcome Spine Trauma) in the clinical setting. European Spine Journal. *2024;33(4):1607-16*.

29. Hakbijl AJ, van Diemen T, Dompeling S, Öner FC, Post MWM, Sadiqi S. Applicability of the AO Spine Patient Reported Outcome Spine Trauma (PROST) in people with motor-complete spinal cord injury: a cognitive interview study. Spinal Cord. *2022;60(10):911-6*.

30. Freedman B. Equipoise and the ethics of clinical research. N Engl J Med. *1987;317(3):141-5*.

31. van de Wall BJM, Stadhouder A, Houwert RM, Öner FC, Beeres FJP, Groenwold RHH, Group NS. Natural experiments for orthopaedic trauma research: An introduction. Injury. *2023;54(2):429-34*.

32. Chou T-Y, Tsuang F-Y, Hsu Y-L, Chai CL. Surgical Versus Non-Surgical Treatment for Thoracolumbar Burst Fractures Without Neurological Deficit: A Systematic Review and Meta-Analysis. Global Spine Journal. *2023;14(2):740-9*.

33. Ghobrial GM, Maulucci CM, Maltenfort M, Dalyai RT, Vaccaro AR, Fehlings MG, et al. Operative and nonoperative adverse events in the management of traumatic fractures of the thoracolumbar spine: a systematic review. Neurosurgical Focus FOC. *2014;37(1):E8*.

34. Gnanenthiran SR, Adie S, Harris IA. Nonoperative versus Operative Treatment for Thoracolumbar Burst Fractures Without Neurologic Deficit: A Meta-analysis. Clinical Orthopaedics and Related Research®. *2012;470(2):567-77*.

35. Häckel S, Stienen MN, Martens B, Neuhaus V, Albers CE. Study Protocol for a Randomized Controlled Clinical Trial on the Outcome of Surgical Versus Primary Nonsurgical Treatment of Traumatic Thoracolumbar Spine Burst Fractures in Patients Without Neurological Symptoms—A34RCT. Neurosurgery Practice. *2024;5(2)*.

36. Lanter L, Rutsch N, Kreuzer S, Albers CE, Obid P, Henssler J, et al. Impact of different surgical and non-surgical interventions on health-related quality of life after thoracolumbar burst fractures without neurological deficit: protocol for a comprehensive systematic review with network meta-analysis. BMJ Open. *2023;13(12):e078972*.

37. Wood KB, Buttermann GR, Phukan R, Harrod CC, Mehbod A, Shannon B, et al. Operative Compared with Nonoperative Treatment of a Thoracolumbar Burst Fracture without Neurological Deficit: A Prospective Randomized Study with Follow-up at Sixteen to Twenty-Two Years: *. JBJS. *2015;97(1):3-9*.

38. Camino-Willhuber G, Bigdon S, Dandurand C, Dvorak MF, Öner CF, Schnake K, et al. Expert Opinion, Real-World Classification, and Decision-Making in Thoracolumbar Burst Fractures Without Neurologic Deficits? Global Spine Journal. 2024;14 (1_suppl):49S-55S.

39. Craig P, Katikireddi SV, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annual Review of Public Health. *2017;38(Volume 38, 2017):39-56*.

40. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health. *2012;66(12):1182-6*.

41. Leatherdale ST. Natural experiment methodology for research: a review of how different methods can support real-world research. International Journal of Social Research Methodology. *2019;22(1):19-35*.

42. Craig P, Katikireddi SV, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annu Rev Public Health. *2017;38:39-56*.

43. Dvorak MF, Öner CF, Schnake K, Dandurand C, Muijs S. From Radiographic Evaluation to Treatment Decisions in Neurologically Intact Patients With Thoraco-lumbar Burst Fractures. Global Spine Journal. *2023;14(1_suppl):4S-7S*.

44. Innovation AC. Thoracolumbar burst fractures (AOSpine A3, A4) in neurologically intact patients: an observational, multicenter cohort study comparing surgical versus non-surgical treatment. 2021. *clinicaltrials. gov. 2022***.**

45. Dandurand C, Öner CF, Hazenbiller O, Bransford RJ, Schnake K, Vaccaro AR, et al. Understanding Decision Making as It Influences Treatment in Thoracolumbar Burst Fractures Without Neurological Deficit: Conceptual Framework and Methodology. Global Spine J. *2024;14(1_suppl):8S-16S*.

46. Dandurand C, Dvorak MF, Hazenbiller O, Bransford RJ, Schnake KJ, Vaccaro AR, et al. Using Equipoise to Determine the Radiographic Characteristics Leading to Agreement on Best Treatment for Thoracolumbar Burst Fractures Without Neurologic Deficits. Global Spine Journal. *2024;14(1_suppl):25S-31S*.

47. Mhaskar R, B BB, Djulbegovic B. At what level of collective equipoise does a randomized clinical trial become ethical for the members of institutional review board/ethical committees? Acta Inform Med. 2013;21(3):156-9.

48. Ghogawala Z, Terrin N, Dunbar MR, Breeze JL, Freund KM, Kanter AS, et al. Effect of Ventral vs Dorsal Spinal Surgery on Patient-Reported Physical Functioning in Patients With Cervical Spondylotic Myelopathy: A Randomized Clinical Trial. JAMA. *2021;325(10):942-51*.

49. Stadhouder A, van Rossenberg LX, Kik C, Muijs SPJ, Öner FC, Houwert RM. Natural Experiments as a Study Method in Spinal Trauma Surgery: A Systematic Review. Global Spine J. *2024;14(5):1640-9*.

10

Nederlandse samenvatting

Het ontwikkelen en uitvoeren van kwalitatief goed klinisch onderzoek bij traumapatenten is een uitdaging. In dit proefschrift wordt een kort overzicht gegeven van de historische ontwikkeling van klinisch onderzoek en worden de potentiële uitdagingen benoemd bij chirurgische- en trauma patiënten in Hoofdstuk 1.

GESCHIEDENIS KLINISCH ONDERZOEK

De geschiedenis van het verrichten van klinisch wetenschappelijk onderzoek vangt aan met een beschrijving in 562 BC in het boek van Daniel in de bijbel. Koning Nebuchadnezzar van Babylon verplichtte zijn manschappen om alleen wijn te drinken en vlees te eten. Een aantal edelen kwamen in opstand en de koning stond hen toe om groenten te eten en water te drinken gedurende 10 dagen. Na deze 10 dagen leken de edelen gezonder dan de andere manschappen en derhalve stond de koning hen toe om hun dieet te vervolgen. (1) Nadien volgde Ibn-Sina (Avicenna) in 1025 waar hij bepaalde regels opstelde voor klinisch onderzoek. James Lind in 1747 deed een patiënt gecontroleerd onderzoek naar de relatie tussen de symptomen van scheurbuik en dieet bij 12 patiënten; de 2 patiënten die citroenen en sinaasappelen aten reageerden het beste. (2) Austin Flint in 1863 publiceerde het eerste placebogecontroleerde onderzoek (3, 4) en in 1943 volgde het eerste dubbelblind gecontroleerde onderzoek in opdracht van de Medical Research Council in Groot Brittannië. Hier opvolgend startte in 1946 het tijdperk van de gerandomiseerde klinische onderzoeken. Statisticus dr Hill superviseerde een gerandomiseerde studie naar streptomycine voor tuberculose behandeling en publiceerde dit in 1948 in de British Medical Journal. Dit was een mijlpaal in de methodologie van klinisch onderzoek en de start van een nieuw klinisch onderzoeks-tijdperk. (3, 5)

METHODOLOGIE

Klinisch onderzoek kan worden ingedeeld in observationeel onderzoek of experimenteel onderzoek. Observationeel onderzoek kan beschrijvend of analytisch zijn. In experimenteel onderzoek wordt een hypothese getest waarbij een interventie wordt verricht. (6) Deze studies kunnen wel of niet gecontroleerd zijn en weer worden onderverdeeld in onder andere klinische trials. Deze kunnen wel of niet gerandomiseerd zijn, ook bestaan er crossover trials en factoriele trials. Randomisatie is geïntroduceerd om bias te voorkomen: een systemische vertekening van de relatie tussen een behandeling, risicofactor of blootstelling enerzijds en klinische uitkomsten anderzijds. (7) Er zijn 3 vormen van bias: selectie bias, informatie bias en confounding. (8) Selectie bias is aanwezig wanneer er systematisch fouten worden gemaakt bij de patiënten inclusie. Geïncludeerde patiënten zijn dus geen goede afspiegeling van de beoogde studie populatie. Informatie bias is aanwezig wanneer er op een systematische basis fouten optreden in het vergaren van patiënten data, zowel door patiënten als door de onderzoeker. Confounding is

een bias die kan voorkomen in klinisch onderzoek naar een causaal verband tussen determinant en uitkomst. Er is sprake van confounding wanneer de factor gerelateerd is aan zowel de determinant als de uitkomst en die het causale verband tussen die twee verstoort. (7) Om confounding te voorkomen kan randomisatie of matching toegepast worden in de studie opzet. Vaak kan er voor confounding worden gecorrigeerd door verschillende statistische methoden als multivariabele regressieanalyse of een propensity score methode toe te passen. (9)

CHIRURGIE

Binnen chirurgische vakgebieden zijn een aantal uitdagingen aanwezig om goed klinisch onderzoek te verrichten. Ten eerste is na de ontwikkeling van antiseptica en anesthesie het chirurgisch vak in korte tijd doorontwikkeld en zijn veel chirurgische operaties gestart zonder dat daar voorafgaand klinische onderzoek naar is verricht. Om daar achteraf nog een Randomized Controled Trial (RCT) voor te verrichten is lastig en onethisch. (10) Tevens zijn er factoren als de leercurve door operateurs, de invloed van commerciële bedrijven indien er bepaalde implantaten gebruikt worden, de onmogelijkheid om patiënten en onderzoekers te blinderen voor een bepaalde ingreep, moeite met inclusie van patiënten met zeldzame aandoeningen of wanneer patiënten zich in een spoedsituatie op de spoedeisende hulp presenteren die het verrichten van een RCT bemoeilijken. (10, 11) Deze factoren maken dat de kwaliteit van chirurgisch klinisch onderzoek matig was in het verleden, echter er is vooruitgang: in 2013 analyseerde Ahmed Ali et al de gepubliceerde chirurgische RCT's tussen 1999 en 2009. In dit uitgebreide overzicht steeg het aantal chirurgische artikelen waarin een RCT-methodologie werd toegepast van 300 in 1999 tot 450 in 2009 (50%). De stijging was met name in andere continenten dan Noord-Amerika was, waar er juist een daling van 23% was. De kwaliteit van deze RCT's nam ook toe in deze 10 jaar. De studie liet ook zien dat de kwaliteit van Europese studies aanzienlijk verbeterde, Noord-Amerikaanse studies verbeterden ook, onderzoeken uit Azië/Oceanië lieten gedurende deze 10 jaar geen verbetering zien. Per land uitgesplitst had Nederland het hoogste percentage onderzoeken met een laag risico op bias (50%). (12) In 2023 werd een update van deze review gepresenteerd met dezelfde methodologie voor chirurgische onderzoeks-papers tot het jaar 2019. In dit onderzoek werden 438 papers opgenomen, wat een stabiel aantal is ten opzichte van 2009. Gastro-intestinale/oncologische chirurgie was de meest voorkomende subspecialiteit (50,1%), terwijl traumachirurgie slechts 3,6% van het totaal aantal gepubliceerde chirurgische RCT's bevatte. De kwaliteit van de studies verbeterde vooral in Azië, waar het percentage studies met een laag risico op bias steeg tot 18,1% (RR 3,50, 1,70 tot 7,32; Blz < 0,001). RCT's uit Afrika/Zuid-Amerika bleven nog steeds zeer laag in risico op bias (<10%) en Europa en Noord-Amerika verbeterden niet significant. De auteurs bespreken dat het conforme aantal RCT's aantoont dat er mogelijk een plateaufase is bereikt, wat niet per se een negatieve ontwikkeling is. (13)

TRAUMAPATIËNTEN EN SPINALE TRAUMAPATIËNTEN

Klinisch onderzoek bij traumapatiënten is vaak nog uitdagender in vergelijking met chirurgische patiënten: Traumachirurgen hebben vaak een voorkeur voor bepaalde operatieve of conservatieve behandelingen vanwege hun opleiding, ervaringen uit het verleden en technische vaardigheden. De lokale ziekenhuiscultuur speelt een belangrijke rol in de manier waarop patiënten worden behandeld en de infrastructuur, protocollen en dagelijkse routines in het ziekenhuis voor traumapatiënten zijn moeilijk te veranderen. Dan verschillen behandelingen aanzienlijk tussen traumacentra en ziekenhuizen en zijn chirurgen terughoudend om bepaalde traumapatiënten te includeren in een RCT waarbij chirurgen een behandeling moeten uitvoeren waar ze minder toe geneigd zijn. Bij traumapatiënten kan dringende klinische besluitvorming nodig zijn, wat inclusie en informed consent extra problematisch maakt. (14, 15) In onderzoek naar spinale trauma patiënten is al het bovenstaande van toepassing, maar er zijn extra factoren die bijdragen aan het geringe aantal RCT's. Een van de factoren is de variatie in classificatiesystemen voor wervelfracturen die in het verleden werden gebruikt. Met de introductie in 2013 en 2016 van de AOSpine thoracolumbale en cervicale subaxiale classificatie is dit echter verbeterd. (16, 17) Een andere complicerende factor was het ontbreken van een universeel ziekte specifiek uitkomstinstrument voor patiënten met spinaal trauma. De AOSpine PROST werd hiervoor ontwikkeld, gevalideerd en geïmplementeerd. (18-21) Naast deze voorwaarden blijft er in de spinale wereld een grote variatie in behandelmogelijkheden en voorkeur van spinaal chirurgen bestaan voor patiënten met traumatisch wervelletsel. Derhalve resteert de vraag wat de optimale behandeling is voor patiënten met een spinaal trauma? (22)

CLINICAL EQUIPOISE EN NATUURLIJKE EXPERIMENTEN

Clinical equipoise, zoals beschreven door Freedman, bestaat wanneer er binnen de deskundige medische gemeenschap onzekerheid bestaat over de optimale behandeling van een bepaalde aandoening. (23) In 2008 werd dit concept geïntroduceerd in een retrospectieve vergelijkende studie over wervelfracturen. (14, 24) Sindsdien is er een toenemend aantal klinische studies die equipoise gebruiken als uitgangspunt voor klinisch onderzoek. In 2023 waren er 212 publicaties in PubMed met equipoise als trefwoord. In de aanwezigheid van equipoise in een klinische setting is het gebruikelijk om 'scholen' te vormen op basis van overtuigingen ten aanzien van de beste behandeling onder behandelende artsen of chirurgen. (14, 25) In plaats van chirurgen te dwingen tot een prospectieve gerandomiseerde klinische studie, kan de 'schoolvorming` worden gebruikt in een observationele onderzoeksopzet, waarbij de expertise in niet-operatieve en operatieve behandeling wordt gebruikt als een voordeel bij de optimale behandeling van patiënten. (14) Observationele studies lijken steeds meer betrouwbaar bewijs

te leveren, zelfs vergelijkbaar met RCT's, mits de observationele studies van voldoende kwaliteit zijn. (15, 26) Observationele studies zijn meer representatief voor de gangbare dagelijkse praktijk waar artsen de behandeling kunnen uitvoeren die ze verkiezen. Deze studies genereren veel sneller resultaten en zijn minder duur. (15) Echter, in observationele studies zonder randomisatie kan bias of confounding optreden. Om dit te ondervangen is een alternatieve studie opzet ontwikkeld: natuurlijke experimenten (NE's). NE's hebben een lange geschiedenis in observationeel algemeen bevolkingsonderzoek. Onlangs zijn NE's als alternatief voor RCT's geïntroduceerd, niet alleen in algemeen bevolkingsonderzoek, maar steeds meer in klinisch onderzoek. (15, 27-29) Van de Wall beschrijft NE`s als: 'observationele studies waarin patiënten worden blootgesteld aan de experimentele of de standaardbehandeling, waarbij de toewijzing van de behandeling wordt bepaald door factoren buiten de controle van de onderzoekers.' Het proces dat de toewijzing van behandelingen regelt, lijkt aantoonbaar op de willekeurige toewijzing in een experimentele setting, vandaar de naam natuurlijk experiment. (15)

In hoofdstuk 2 worden de resultaten gepresenteerd van een enquête onder Nederlandse chirurgen verantwoordelijk voor de zorg voor patiënten met een wervelfractuur in de verschillende traumacentra. Doel was om in kaart te brengen hoe de traumazorg voor deze patiënten was georganiseerd, wat de knelpunten waren en wat er verbeterd kon worden.

SPINALE TRAUMAZORG IN NEDERLAND

De enquête werd in 2013 afgenomen. Resultaten lieten zien dat in een relatief klein land als Nederland de trauma zorg toentertijd in de 11 traumacentra verschillend was georganiseerd. Elk ziekenhuis had een gespecialiseerd team van behandelaars van spinaal trauma patiënten met een wisselende samenstelling van orthopedisch chirurgen, traumachirurgen, revalidatieartsen, algemeen chirurgen en neurochirurgen. In 8 van de 11 ziekenhuizen waren specifieke traumaprotocollen voor spinaal trauma patiënten. Patiënten met neurologisch letsel werden bij voorkeur naar een gespecialiseerd ziekenhuis getransporteerd. Er werden toentertijd verschillende classificatiesystemen gebruikt en er werd aangegeven dat er een noodzaak was in gebruik van een meer universeel classificatie systeem. Er werd wisselend gedacht over hoe snel patiënten met neurologisch letsel geopereerd dienden te worden, dit varieerde van binnen 4 uur tot binnen 48 uur. De chirurgen gaven de behandeling voor wervelfracturen een gemiddeld cijfer van 7.7 (5-10) op een schaal van 0 tot 10 in hun eigen ziekenhuis. 64% van de ondervraagden gaf aan dat het noodzakelijk is de spinale traumazorg in Nederland meer te concentreren.

VERBETERINGEN

In de opvolgende jaren zijn al veel stappen genomen om de spinale zorg te verbeteren. Er is een universeel classificatiesysteem gekomen (16, 17), we kunnen monitoren met Patient Related Outcome Measurements hoe patiënten na het trauma herstellen (19-21) en de Advance Trauma Life Support opvang werd geïmplementeerd. (30). Resterende vragen als wat de beste timing is om patiënten met neurologisch letsel te opereren of welke fracturen beter operatief of conservatief behandeld kunnen worden blijven echter bestaan.

De conservatieve behandeling van wervelfracturen varieert per behandelend chirurg en per ziekenhuis. Om duidelijkheid in deze behandeling te scheppen werden een aantal gangbare conservatieve behandelmethodes geïdentificeerd. Vervolgens werd een gerandomiseerde studie opgezet als beschreven in hoofdstuk 3.

CONSERVATIEVE BEHANDELING WERVELFRACTUREN

In twee Amsterdamse stadsziekenhuizen werden in totaal 133 patiënten geïncludeerd en aansluitend gerandomiseerd tussen 1991 en 1997. Er waren 108 patiënten met een compressie fractuur en 25 patiënten met een burst fractuur. Patiënten werden gerandomiseerd tussen behandelmethodes fysiotherapie en houdingsinstructies, een brace voor 6 weken, of een gipscorset voor 6 of 12 weken voor de compressie fracturen. De burst fracturen werden behandeld met 12 weken een brace of een gipscorset. Demografische gegevens van patiënten voor compressie fracturen waren vergelijkbaar behalve voor geslacht, leeftijd, type trauma en tijd tussen trauma en opname. In burst fracturen was de gemiddelde leeftijd van vrouwen (57 jaar) hoger dan voor mannen (36 jaar) met een Confidence Interval van 1.8–41.5. Van 75.4% van de patiënten waren follow up gegevens beschikbaar met een gemiddelde follow up duur van 7.1 jaar (1-12 jaar, SD 3.0) Resultaten lieten geen significante verschillen zien ten aanzien van gemeten kyphose hoeken tijdens en na behandeling. In de compressie groep scoorde een brace significant een lagere pijnscore voor persisterende pijnklachten dan een gips voor 12 weken (mean difference 19.0, CI 1.87–36.2, calculated power 0.60). Ook de Oswestry Disability Index liet een significant verschil zien ten opzichte van 12 weken gips in het voordeel van een brace (mean difference 10.1, CI 0.25–20.0, calculated power 0.57). Dit geldt ook ten opzichte van fysiotherapie (mean difference 14.9, CI 2.7–27.1, calculated power 0.70). Fysiotherapie vonden patiënten een betere behandeling dan een gips voor 6 of 12 weken (mean difference 33.9, CI of 16.6–51.3, calculated power 0.97; mean difference 21.6, CI 3.4–39.8, calculated power 0.81). In de burst fracturen groepen werden geen significante verschillen aangetoond. 18% van de patiënten in de compressie groep gaven een Visual Analogue Score van meer dan 50 (matige pijn) en 9% een VAS score van 70 (ernstige pijn). In de burst groep is vanwege ernstige pijn en progressieve deformiteit 1 patiënt geopereerd. Geen van de patiënten had een VAS score boven de 70, 3 patiënten

(12%) hadden een VAS score boven de 50. Multivariaat analyses in beide groepen konden geen prognostische factoren vaststellen om resterende pijnklachten te voorspellen. Concluderend toonde deze studie dat een brace behandeling met ondersteunende fysiotherapie de beste behandeling voor patiënten met een compressie fractuur was. Wel had 20% van de patiënten na behandeling na langere follow up nog matig tot ernstige rugklachten. (31)

Gezien bovenstaande resultaten rees de vraag wat voor patiënten en behandelaars belangrijk is als uitkomst maat bij wervelfracturen. De VAS en ODI zijn veel gebruikte vragenlijsten in patiënten met degeneratieve rugklachten en derhalve in bovenstaande studie gebruikt. Maar zijn deze vragenlijsten bruikbaar in spinale trauma patiënten of dienen we naar andere factoren te kijken? Derhalve wordt in hoofdstuk 4 een search verricht welke uitkomst scores toentertijd aanwezig waren in de literatuur.

UITKOMST SCORES IN SPINALE TRAUMA PATIËNTEN

Aangezien de zorg voor polytrauma patiënten sterk is verbeterd en deze patiënten langer leven is het belangrijk om de geleverde zorg ook te evalueren. Veel van deze patiënten hebben namelijk ook één of meerdere wervelfracturen. (32) De World Health Organization heeft een International Classification of Functioning, Disability and health (ICF) document opgesteld. Dit beschrijft in detail verschillende gezondheidsdomeinen van kwaliteit van leven. Deze kunnen gebruikt worden bij het vergelijken van instrumenten voor uitkomsten en voor het beoordelen van de validiteit om een zinvollere analyse mogelijk te maken. Er werd een literatuuronderzoek in naar uitkomst scores voor patiënten verricht in 2009 waarbij 6090 papers werden geïdentificeerd. Deze werden door twee lezers beoordeeld, inclusief referenties, net als verwante artikelen in PubMed. Wat opviel was dat er geen specifiek scoringssysteem is voor spinale trauma patiënten. Er zijn algemene gezondheids instrumenten als de SF-36. Dit is een veelgebruikte generieke maatstaf met 36 items verdeeld over acht gezondheidsdomeinen. Hoewel het op veel patiëntpopulaties is gevalideerd, is het niet specifiek ontworpen voor spinale trauma en heeft het beperkingen, met name voor de spinal cord injury patiënten. De EQ-5D is een instrument dat vijf dimensies van gezondheid (mobiliteit, zelfzorg, activiteiten, pijn, angst/depressie) meet en omvat een algemene gezondheids-VAS. Het is ontwikkeld door de EuroQol Group en wordt steeds vaker als op zichzelf staand instrument gebruikt in wervelkolomonderzoek. Het kan ook nuttig zijn voor economische evaluaties. De Health Utilities Index (HUI) is een generieke maat voor gezondheidstoestand en Quality Of Life, met een focus op functioneel potentieel in plaats van prestaties. Dit instrument maakt ook economische evaluaties mogelijk. De Sickness Impact Profile (SIP) is een 136-items meetinstrument en evalueert 12 activiteitengebieden bij patiënten en wordt gebruikt om veranderingen in gezondheidstoestand over de tijd en tussen groepen te detecteren. Het is nuttig voor evaluatie en beleidsvorming, maar wordt minder vaak gebruikt in spinale trauma.

De gebruikte mentale/ psychologische scorings instrumenten zijn de Hamilton Depression Rating Scale (HDRS) en Beck Depression Inventory (BDI). Beide instrumenten worden veel gebruikt in depressieonderzoek, maar zijn niet gevalideerd voor spinale trauma. De HDRS meet aspecten zoals depressieve stemming en werk interesses, terwijl de BDI emotionele en somatische symptomen evalueert. Dan is er ook de Hospital Anxiety and Depression Scale (HADS): Dit zelfbeoordelingsinstrument is ontworpen om angst- en depressie-stoornissen te identificeren bij niet-psychiatrische patiënten en presteert goed bij het screenen van deze dimensies. Hoewel veel gebruikt bij SCI-patiënten, is de validiteit voor spinale trauma niet goed onderzocht. Verder zijn er meerdere wervelkolom specifieke uitkomst instrumenten, vooral veel gebruikt bij patiënten met lage rugpijn. De Oswestry Disability Index (ODI) en Roland-Morris Disability Questionnaire (RMDQ) zijn de meest gebruikte hierin. De ODI lijkt iets gevoeliger voor veranderingen bij patiënten met ernstige pijnklachten, terwijl de RMDQ goed gevalideerd en betrouwbaar is voor patiënten met minder ernstige lage rugpijn. Andere uitkomst instrumenten zijn de Million Visual Analog Scale (MVAS), Low-Back Outcome Score (LBOS), en de Quebec Back Pain Disability Scale (QBPDS): allen gebruikt bij lage rugpijn klachten maar hebben beperkte validatie bij spinale trauma patiënten. Ook bestaat de Resumption of Activities of Daily Living Scale (RADLS): Dit instrument kijkt naar de hoe patiënten functioneerden voor aanvang van ziekte/trauma qua dagelijkse activiteiten in vergelijking met hoe het daarna gaat. Functionele instrumenten focussen op een specifiek item qua uitkomst. Return to Work (RTW) en Health Service Use (HSU) kunnen gebruikt worden voor economische analyses. Ze zijn eenvoudig te meten maar moeten voorzichtig worden geïnterpreteerd vanwege de variabele invloed van patiënt gerelateerde factoren. De Functional Capacity Index (FCI) is een instrument dat is gevalideerd in grote trauma-populaties maar niet specifiek voor spinale trauma. Het meet functionele capaciteit over tien dimensies. De Functional Independence Measure (FIM) en Spinal Cord Independence Measure (SCIM) worden veel gebruikt bij SCI-patiënten en zijn betrouwbaar en gevalideerd. De SCIM zou kunnen worden gebruikt bij patiënten met wervelfracturen en neurologische uitval. (33) Het ICF-raamwerk, dat een biopsychosociaal model van gezondheid en functioneren weerspiegelt, biedt een generieke aanpak voor het selecteren van relevante domeinen. Voor musculoskeletaal trauma werden op een consensus meeting 47 categorieën geselecteerd voor een 'core set'. SCI patiënten werden in een andere 'core' set geplaatst. In het artikel selecteerden we 12 domeinen die wij relevant achtten voor spinale trauma patiënten waarbij we aangeven hoe de diverse uitkomst instrumenten deze 12 domeinen ook adresseren. De domeinen zijn mentaal functioneren, gevoel en pijn, urogenitaal en voortplanting, zenuwstelsel, beweging, mobiliteit, zelfverzorging, algemene dagelijkse activiteiten in huis, relaties, grote levensgebeurtenissen, sociaal leven, steun en relaties en gebruik voorzieningen en beleid. (33) Concluderend zijn de beschreven uitkomstinstrumenten onvoldoende afgestemd op de specifieke behoeften van spinale trauma patiënten. De meeste beschikbare instrumenten zijn niet specifiek ontworpen voor deze populatie, wat leidt tot beperkingen en gebrek aan nauwkeurigheid

bij het meten van uitkomsten. Er is een dringende behoefte aan een speciaal ontwikkeld instrument dat de unieke dynamiek van spinale trauma weerspiegelt. Dit zou kunnen bestaan uit een combinatie van bestaande vragenlijsten en zou de nadruk moeten leggen op de terugkeer van activiteiten en functioneren ten opzichte van de situatie vóór het trauma. Computeradaptive tests of item response theory kunnen hierbij helpen door relevante domeinen en activiteiten te combineren. Met deze review werd een begin gemaakt met een evidence -based lijst van domeinen voor verder onderzoek naar de beste uitkomst maten voor herstel van patiënten met traumatisch wervelletsel. Dit heeft in de jaren hierna een verder vervolg gekregen door aanvullende studies uiteindelijk resulterend in de AO Patient Reported Outcome Measurement Score ontwikkeling. (20)

EQUIPOISE

In hoofdstuk 5 en 6 worden de concepten equipoise en klinische equipoise geïntroduceerd. Er is equipoise als er sprake is van oprechte onzekerheid bij de klinische onderzoeker over de vergelijkende therapeutische bewijsvoering van elke behandeling in een onderzoek. Het wordt algemeen aanvaard dat equipoise een ethisch voorwaarde is bij het verrichten van klinisch onderzoek. In onderzoeken met meerdere behandelingen moet er equipoise bestaan van alle behandelingen. Indien dit niet zo is moet het onderzoeksprotocol aangepast worden om behandelingen met elkaar te vergelijken volgens ethische principes. Equipoise is belangrijk echter maakt het verrichten van goed onderzoek ook moeilijk; wanneer een onderzoeker de ene behandeling boven de andere verkiest, wordt de equipoise verstoord en is het onderzoek niet langer ethisch. (23) Klinische equipoise, zoals geïntroduceerd door Freedman in 1987, bestaat wanneer er oprechte onzekerheid bestaat binnen de medische expertgemeenschap over de optimale behandeling van een bepaalde ziekte of aandoening. (23) Klinische equipoise is aanwezig wanneer er een duidelijke behandelvoorkeur is van de onderzoekers. Bij dit concept zijn de onderzoekers niet het leidende onderwerp in het onderzoek; ze moeten simpelweg erkennen dat een door hen minder geprefereerde behandeling de voorkeur geniet van collega's die zij als verantwoordelijk en deskundig beschouwen. (23) Dit concept was het uitgangspunt van een cohortonderzoek bij patiënten met een wervelfractuur uitgevoerd in 2008. Patiënten werden retrospectief geïdentificeerd en prospectief gevolgd waarbij er klinische equipoise bestaat over de beste behandeling voor patiënten met een wervelfractuur, operatief of niet-operatief in dit geval. (14) Vanwege het retrospectieve karakter van deze studie konden historische gegevens vooraf worden verzameld om de haalbaarheid te bepalen met betrekking tot uitkomst ten aanzien van de benodigde patiënten aantallen. Destijds was er een unieke gelegenheid waarin twee universitaire ziekenhuizen verschillende voorkeurs behandelingsschema's hadden bij werveltrauma zorg in Nederland. Deze twee ziekenhuizen waren vergelijkbaar in

hun patiëntenpopulatie. Door patiënten met een wervelfractuur die werd opgenomen en behandeld te identificeren, werd, afhankelijk van welk ziekenhuis een patiënt werd opgenomen, de behandeling bepaald door waar zij terechtkwamen. Hierdoor werd de behandelings-toewijzing bepaald door factoren buiten de controle van de onderzoekers en werd confounding verminderd. (15) Middels dit klinische equipoise-concept werden 636 patiënten retrospectief geïdentificeerd in twee ziekenhuizen. De traumadata van deze patiënten werden geblindeerd aangeboden aan een expert panel werkzaam in een van beide ziekenhuizen welke een behandeling voorstelde. Bij 190 patiënten was er sprake van een verschillend behandel voorstel van twee wervelkolomchirurgen ieder representatief als vertegenwoordigers van de twee ziekenhuizen. (24) De twee patiënten groepen operatief en conservatief bestonden elk uit 95 patiënten en waren vergelijkbaar qua demografische gegevens. In verband met het retrospectieve karakter van de studie was het een uitdaging om lange termijn functionele resultaten te verkrijgen van patiënten, maar met een follow-up percentage van 79% na gemiddeld 6,2 jaar, voldoende. (24) Naast de in het ziekenhuis gemeten parameters, was de belangrijkste onderzoeksvraag wat de lange termijn uitkomsten waren van patiënten. Hiervoor gebruikten we algemene gezondheids vragenlijsten en de ODI, VAS en de Denis-werkschaal. Zoals eerder besproken, zijn deze vragenlijsten niet ziektespecifiek en geven mogelijk geen volledig en adequaat beeld. (20) Over het algemeen waren de resultaten van niet-operatieve en operatieve behandelingen op middellange termijn vergelijkbaar, hoewel er een verschil leek te zijn in neurologische herstel ten gunste van patiënten die een operatieve behandeling hadden ondergaan.

We zijn van mening dat toekomstig onderzoek zich moeten richten op kosten/effectiviteit en direct postoperatief en korte termijn uitkomst parameters om te analyseren wat de beste behandelingsopties zijn voor patiënten met spinale trauma's, vooral bij patiënten waar klinische equipoise bestaat. Sinds het bovengenoemde klinische equipoise-artikel uit 2009 zijn meerdere RCT's (of RCT-protocollen), systematische reviews en meta-analyses uitgevoerd. (34-38) De resultaten variëren van geen verschil tot betere uitkomst bij conservatieve patiënten of juist beter herstel bij operatieve behandeling. Een recent artikel van Camino-Willhuber in 2024 toonde aan dat bij AO A3/A4-fracturen er discussie was over de voorkeursbehandeling tussen een expertgroep van wereldwijde wervelkolomchirurgen en de chirurgen welke de patiënten daadwerkelijk behandelden. Het expertpanel adviseerde een operatie voor 30% van de A3-letsels en 68% van de A4-letsels. Echter, 61% van de patiënten met zowel A3- als A4-fracturen ondergingen in de praktijk een operatie. (39) Tot op heden is het nog steeds niet duidelijk wat de beste manier is om deze patiënten met dit type fractuur te behandelen. Dus, wat is de volgende stap om deze onderzoeksvraag te beantwoorden?

NATUURLIJK EXPERIMENT

In hoofdstuk 6 en 7 wordt het concept van het Natuurlijke Experiment geïntroduceerd als methode in prospectieve en retrospectieve studies gebaseerd op klinische equipoise. Natuurlijke Experimenten (NE's) worden al lange tijd toegepast in algemene bevolkingsonderzoeken. Dit gaat terug tot 1850 naar het cholera-onderzoek van John Snow, die stelde dat water van bepaalde besmette waterpompen de oorzaak was van de verspreiding van de ziekte in Londen.(27) Een definitie van een Natuurlijk Experiment komt van het Britse Medical Research Council, die NE's breed definieert als "elk gebeurtenis dat niet onder de controle van een onderzoeker valt en dat een populatie verdeelt in blootgestelde en niet-blootgestelde groepen". (40) Craig et al voegden hier twee voorwaarden aan toe: namelijk dat de implementatie van de interventie niet afhankelijk is van het al dan niet bestaan van een plan om de interventie te evalueren en dat willekeurige toewijzing van de interventie niet haalbaar is vanwege ethische of politieke redenen. Deze items richten zich waarschijnlijk meer op algemeen bevolkingsonderzoek waar het natuurlijke experimentontwerp vaker wordt toegepast. (27, 40, 41) Er is discussie over de eerste voorwaarde: Craig stelde dat natuurlijke experimenten alleen moeten worden geëvalueerd wanneer er gegevens beschikbaar zijn om een experimenteel ontwerp te gebruiken voor de evaluatie en niet-experimentele ontwerpen moeten worden vermeden. (40) Maar men zou ook kunnen beweren dat zelfs zwakker bewijs verkregen uit een niet-experimentele evaluatie van een natuurlijk experiment beter is dan helemaal geen bewijs. (41) Leatherdale publiceerde in 2017 een review over hoe verschillende onderzoeks voorwaarden de uitkomsten van natuurlijke experimenten kunnen ondersteunen. (41) Hij adviseert over bepaalde strategieën waar onderzoekers naar moeten streven bij het evalueren van natuurlijke experimenten. Ten eerste is het belangrijk om uitkomstmetingen vóór en na het natuurlijke experiment te verzamelen, ten tweede is het belangrijk om een interventiegroep met een controlegroep te vergelijken, en ten derde te streven naar het best mogelijke ontwerp op basis van de beschikbare of te verzamelen gegevens. Dit alles om bias te voorkomen aangezien dit de interne en externe validiteit van de uitkomsten kan beïnvloeden. (40) RCT's beperken over het algemeen effectief interne bias door selectiebias te beperken. In natuurlijke experimenten kan bias door confounding (een situatie waarin het effect (of de associatie) van een interventie op een uitkomst wordt vertekend door de aanwezigheid van een andere variabele) worden beperkt door gebruik te maken van statistische analyses met pre- en post-test metingen, regressie-modellen en het gebruik van een adequate controlegroep. (41) Als de controlegroep niet representatief lijkt, kunnen propensity scores, regressieanalyses of difference-in-difference modellen worden gebruikt om te corrigeren voor mogelijke verschillen in baseline karakteristieken (bekende confounders). Voor toekomstige natuurlijke experiment studies, zowel in bevolkingsonderzoek maar ook in vergelijkende klinisch medische studies, zouden patiënten registers van waarde kunnen zijn bij het gebruik van natuurlijke experimentele studies. Deze zijn ook voor trauma patiënten steeds meer in ge-

bruik. Het huidige onderzoek in spinale trauma patiënten is diverse in onderzoeksmethodologie: RCT's zijn nog steeds de gouden standaard in klinisch onderzoek, maar recentelijk hebben NE's en andere alternatieven voor RCT's belangstelling gekregen omdat ze toepasbaar zijn bij het evalueren van grootschalige volksgezondheidsinterventies die moeilijk te onderzoeken zijn in een experimentele omgeving. (40) Een van de nadelen van een RCT is dat bepaalde interventies of behandelstrategieën ethische bezwaren oproepen bij behandelaars, vooral als er al bestaand bewijs is in de behandeling van spinaal trauma dat de ene behandeling betere uitkomsten heeft dan de andere. (15, 42) Het verrichten van RCT's kost verder veel tijd, is duur en vraagt veel middelen om correct uit te voeren. Vooral in traumatisch wervelkolom onderzoek is dit vaak niet praktisch. Strikte inclusie- en exclusiecriteria in RCT's kunnen de generaliseerbaarheid van bevindingen naar echte klinische omstandigheden beperken. De lange termijn follow-up in RCT's geeft vaak logistieke uitdagingen en mogelijk verlies aan follow-up. Deze argumenten zijn zeker van toepassing op de behandeling van spinale fracturen. Wereldwijd is het moeilijk om individuele chirurgen te vinden die zowel bekwaam als comfortabel zijn in operatieve of niet-operatieve behandeling en geen sterke voorkeur hebben voor de ene behandelmethode boven de andere voor specifieke patiënten. (39) Daarnaast is het moeilijk om patiënten met een werveltrauma te standaardiseren, blindering van behandeling is niet mogelijk, de inclusiecriteria zijn beperkt in een RCT en er is dus vaak sprake van een voorkeur voor behandeling van patiënt en behandelaar. Dit maakt een chirurgische gerandomiseerde trial bij spinale fracturen buitengewoon moeilijk en praktisch onmogelijk.(43) Aldus kan bij deze collectieve equipoise en een hoge mate van voorkeur bij de zorgverlener een prospectieve observationele studie gebruik maken van het equipoise-concept en studie opzet creëren dat superieur is aan de RCT in termen van generaliseerbaarheid en beoordeling van effectiviteit in de 'echte' wereld. (14, 24, 43) Dit is het natuurlijke experimentontwerp zoals beschreven in hoofdstuk 7. In 2016 werd een internationale multicenter prospectieve studie naar thoracolumbale burstfracturen opgezet die chirurgische versus niet-chirurgische behandeling vergeleek, geïnitieerd door het AO Spine Knowledge Forum trauma. (44) 208 patiënten met een thoracolumbale burstfractuur van T10 tot L2 werden chirurgisch of conservatief behandeld waarbij het beleid werd bepaald door de behandelende wervelkolomchirurg op dat moment. De oorspronkelijke traumagegevens van deze patiëntengroep werden gepresenteerd aan 22 leden van het AO Spine Knowledge Forum Trauma (AOSKFT), dat bestaat uit opinieleiders op het gebied van werveltrauma van over de hele wereld. Deze 22 werveltrauma-experts vormden een panel van experts die in staat zouden zijn om de beelden van röntgenologische gegevens van 183 patiënten die waren behandeld in de Spine TL A3/4-studie te classificeren, analyseren en behandeladviezen te geven. (45) Van de 183 patiënten was er in 8% van de patiënten volledige overeenstemming over classificatie, mate van Posterior Ligament Injury, mate van comminutie en behandeling. In een ander artikel van Dandurand wordt de drempel van equipoise besproken;

onderzoekers op het gebied van medische ethiek suggereren dat een klinische trial niet ethisch is wanneer er overeenstemming is van meer dan 70- 80%. (46) In ons artikel in hoofdstuk 7 bespraken we ook de klinische equipoise drempel. Een enquête onder leden van Medisch Ethische Commissies tijdens een conferentie over bio-ethiek in Florida over collectieve equipoise toonde aan dat het niet langer ethisch is om patiënten te includeren wanneer het equipoise niveau hoger is dan 80% (80:20 verdeling van onzekerheid). Bij kinderen en ouderen en in levensbedreigende situaties gaven deze leden aan een hoger niveau van equipoise te wensen voor de goedkeuring voor deze studies. (47) Dandurand, in de eerdergenoemde studie, stelde het equipoise-percentage op 77%, wat betekent dat er geen consensus was wanneer 17 (of minder) van de 22 wervelkolomchirurgen het oneens waren over de behandeling. Ghogawala voerde in 2021 een RCT uit over operatieve behandeling van cervicale spondylotische myelopathie. Ze introduceerden een expertpanel van 15 chirurgen die vóór de geïnformeerde toestemming van de patiënt de casus bekeken en advies gaven over wel of geen kandidaat voor randomisatie en advies gaven over ventrale of dorsale chirurgie. Klinische equipoise werd gedefinieerd als niet voldaan wanneer (1) 80% of meer van de panelleden koos voor ventrale of dorsale chirurgie of (2) een eenvoudige meerderheid tegen randomisatie stemde. De onderzoekers vonden een verhoogde mate van instemming van de patiënt met de randomisatie met deze onderzoeks opzet. (48) Dit is een interessante extra stap in klinisch onderzoek waarbij het gebruik van klinische equipoise door experts, patiënten aanmoedigt om deel te nemen aan een RCT. Zoals eerder vermeld, om confounding te verminderen in een natuurlijke experimentstudie, adviseren we een expertpanel samen te stellen dat representatief is voor de twee behandelscholen. Het panel wordt gevraagd om onafhankelijk te beslissen over de geprefereerde behandeling voor de patiënt, alsof patiënten op dat moment aan hen werden gepresenteerd in de klinische praktijk. Patiënten worden geïncludeerd als er geen consensus is over de behandelingskeuze tussen de 'scholen'; ze worden uitgesloten bij overeenstemming omdat er geen klinische equipoise is. Dit zorgt ervoor dat de geïncludeerde onderzoekspopulatie bestaat uit patiënten die behandeling A zouden hebben ontvangen in "school A", maar die feitelijk werden gezien en behandeld volgens de overtuiging van "school B", en vice versa (uitwisselbaarheid). Onze recente review van de NE-methodologie in spinaal trauma identificeerde slechts 4 artikelen die het NE-ontwerp in de afgelopen 19 jaar hebben gebruikt. (hoofdstuk 8) (49) Eén artikel gebruikte een expertpanel, iets wat we zouden aanbevelen bij het gebruik van de natuurlijke experimentmethodologie in klinisch onderzoek. Hoewel 4 artikelen een teleurstellend aantal is, toonden de MINORS-criteria van de 4 artikelen aan dat de kwaliteit van deze studies hoog was, vooral in vergelijking met ander retrospectieve vergelijkend onderzoek. (49)

CONCLUSIES EN TOEKOMST

In dit proefschrift worden de behandeling van wervelfracturen en de onderzoeksmethodologie Natuurlijk Experiment bij patiënten met een spinaal trauma bediscussieerd. De afgelopen jaren is er aanzienlijke vooruitgang geboekt in de classificatie van deze fracturen met de ontwikkeling van de AOSpine classificatie. Met een beter begrip van het traumamechanisme kan de behandeling van patiënten worden geoptimaliseerd. Bovendien is het gebruik van klinische uitkomst scores met de AO Patient Reported Outcome Measurement Score een grote stap vooruit in het analyseren en verbeteren van de patiëntenzorg. Het zou de plicht van de wervelkolomgemeenschap moeten zijn om deze uitkomstscores in hun dagelijkse praktijk te implementeren. Er blijven echter resterende vragen en onduidelijkheden bestaan in de werveltraumazorg die nog niet zijn beantwoord of opgelost, en we moeten erkennen dat het uitvoeren van goed klinisch onderzoek bij wervelfracturen een uitdaging is. Het concept van het Natuurlijke Experiment wordt geïntroduceerd als een methodologie gebaseerd op klinische equipoise en kan een toegevoegde waarde zijn in traumaonderzoek. Deze methodologie kan worden gebruikt met een expertpanel om bias en confounding verder te verminderen bij het toepassen van observationeel onderzoek. NE's zijn een veelbelovende volgende stap in klinisch onderzoek, maar zijn tot nu toe niet algemeen bekend. Om de zichtbaarheid van de methodologie te verbeteren, is de Natural Experiments Study Group (NEXT Study Group) opgericht. Het is een internationale non-profit samenwerking van klinische onderzoekers op het gebied van spoed- en (orthopedische) traumachirurgie. De ambitie van de NEXT Study Group is bij te dragen aan de verbetering van de zorg voor patiënten in de spoed- en (orthopedische) traumachirurgie. Met dit NEXT Study Group-initiatief en toekomstige publicaties in peer-reviewed tijdschriften kan het concept van het Natuurlijke Experiment verder worden ontwikkeld, aandacht krijgen in de medische wereld en de zorg voor spinale trauma patiënten verbeteren. Bovendien kunnen prospectieve registers bij traumapatiënten een verdere bijdrage leveren aan het gebruik van Natuurlijke Experimenten in (spinaal) traumaonderzoek.

The Natural Experiments Study Group
(NEXT Study Group)

REFERENCES

1. **Collier R**. Legumes, lemons and streptomycin: a short history of the clinical trial. CMAJ. *2009;180(1):23-4.*
2. **Hughes RE**. James Lind and the cure of scurvy: an experimental approach. Med Hist. *1975;19(4):342-51.*
3. **Bhatt A**. Evolution of clinical research: a history before and beyond james lind. Perspect Clin Res. *2010;1(1):6-10.*
4. **Flint A**. ART. I.--*A contribution toward the Natural History of Articular Rhcumatism; consisting of a report of thirteen cases treated solely with palliative measures. The American Journal of the Medical Sciences (1827-1924). 1863;46(91):17.*
5. **Hill CA, Riley C, Gifford JH**. The Use of Intrathecal Heparin in Conjunction with Streptomycin in the Treatment of Tuberculous Meningitis: Preliminary Report. J Clin Pathol. *1948;1(3):157-61.*
6. **Chidambaram AG, Josephson M**. Clinical research study designs: The essentials. Pediatr Investig. *2019;3(4):245-52.*
7. **Groenwold RH**. [Three types of bias: distortion of research results and how that can be prevented]. Ned Tijdschr Geneeskd. *2013;157(40):A6497.*
8. **KLEINBAUM DG, MORGENSTERN H, KUPPER LL**. SELECTION BIAS IN EPIDEMIOLOGIC STUDIES12. American Journal of Epidemiology. *1981;113(4):452-63.*
9. **Vetter TR, Mascha EJ**. Bias, Confounding, and Interaction: Lions and Tigers, and Bears, Oh My! Anesth Analg. 2017;125(3):1042-8.
10. **McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D**. Randomised trials in surgery: problems and possible solutions. BMJ. *2002;324(7351):1448-51.*
11. **Hirst A, Agha RA, Rosin D, McCulloch P**. How can we improve surgical research and innovation?: the IDEAL framework for action. Int J Surg. *2013;11(10):1038-42.*
12. **Ahmed Ali U, van der Sluis PC, Issa Y, Habaga IA, Gooszen HG, Flum DR, et al**. Trends in worldwide volume and methodological quality of surgical randomized controlled trials. Ann Surg. *2013;258(2):199-207.*
13. **Pronk AJM, Roelofs A, Flum DR, Bonjer HJ, Abu Hilal M, Dijkgraaf MGW, et al**. Two decades of surgical randomized controlled trials: worldwide trends in volume and methodological quality. Br J Surg. *2023;110(10):1300-8.*
14. **Stadhouder A, Öner FC, Wilson KW, Vaccaro AR, Williamson OD, Verbout AJ, et al**. Surgeon equipoise as an inclusion criterion for the evaluation of nonoperative versus operative treatment of thoracolumbar spinal injuries. Spine J. *2008;8(6):975-81.*
15. **van de Wall BJM, Stadhouder A, Houwert RM, Öner FC, Beeres FJP, Groenwold RHH, Group NS**. Natural experiments for orthopaedic trauma research: An introduction. Injury. *2023;54(2):429-34.*
16. **Vaccaro AR, Koerner JD, Radcliff KE, Öner FC, Reinhold M, Schnake KJ, et al**. AOSpine subaxial cervical spine injury classification system. Eur Spine J. *2016;25(7):2173-84*

17. Vaccaro AR, Öner C, Kepler CK, Dvorak M, Schnake K, Bellabarba C, et al. AOSpine thoracolumbar spine injury classification system: fracture description, neurological status, and key modifiers. Spine (Phila Pa 1976). *2013;38(23):2028-37*.

18. Buijs GS, de Gendt EEA, Sadiqi S, Post MW, Muijs SPJ, Öner FC. Long-Term Reliability and Validity of the AO Spine PROST (Patient-Reported Outcome Spine Trauma). Spine (Phila Pa 1976). *2022;47(17):E562-E9*.

19. Sadiqi S, de Gendt EEA, Muijs SPJ, Post MWM, Benneker LM, Holas M, et al. Validation of the AO Spine CROST (Clinician Reported Outcome Spine Trauma) in the clinical setting. Eur Spine J. *2024;33(4):1607-16*.

20. Sadiqi S, Dvorak MF, Vaccaro AR, Schroeder GD, Post MW, Benneker LM, et al. Reliability and Validity of the English Version of the AOSpine PROST (Patient Reported Outcome Spine Trauma). Spine (Phila Pa 1976). *2020;45(17):E1111-E8*.

21. Sadiqi S, Lehr AM, Post MW, Dvorak MF, Kandziora F, Rajasekaran S, et al. Development of the AOSpine Patient Reported Outcome Spine Trauma (AOSpine PROST): a universal disease-specific outcome instrument for individuals with traumatic spinal column injury. Eur Spine J. *2017;26(5):1550-7*.

22. Kweh BTS, Tee JW, Dandurand C, Vaccaro AR, Lorin BM, Schnake K, et al. The AO Spine Thoracolumbar Injury Classification System and Treatment Algorithm in Decision Making for Thoracolumbar Burst Fractures Without Neurologic Deficit. Global Spine J. *2024;14(1_suppl):32S-40S*.

23. Freedman B. Equipoise and the ethics of clinical research. N Engl J Med. *1987;317(3):141-5*.

24. Stadhouder A, Buskens E, de Klerk LW, Verhaar JA, Dhert WA, Verbout AJ, et al. Traumatic thoracic and lumbar spinal fractures: operative or nonoperative treatment: comparison of two treatment strategies by means of surgeon equipoise. Spine (Phila Pa 1976). *2008;33(9):1006-17*.

25. Dandurand C, Dvorak MF, Hazenbiller O, Bransford RJ, Schnake KJ, Vaccaro AR, et al. Using Equipoise to Determine the Radiographic Characteristics Leading to Agreement on Best Treatment for Thoracolumbar Burst Fractures Without Neurologic Deficits. Global Spine J. *2024;14(1_suppl):25S-31S*.

26. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. *2000;342(25):1887-92*.

27. Craig P, Katikireddi SV, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annu Rev Public Health. *2017;38:39-56*.

28. Hoepelman RJ, Beeres FJP, Beks RB, Sweet AAR, Ijpma FF, Lansink KWW, et al. Non-operative vs. operative treatment for multiple rib fractures after blunt thoracic trauma: a multicenter prospective cohort study. *Eur J Trauma Emerg Surg*. *2023;49(1):461-71*.

29. Hoepelman RJ, Minervini F, Beeres FJP, van Wageningen B, FF IJ, van Veelen NM, et al. Quality of life and clinical outcomes of operatively treated patients with flail chest injuries: A multicentre prospective cohort study. Front Surg. *2023;10:1156489*.

30. Styner JK. The Birth of Advanced Trauma Life Support. Journal of Trauma Nursing | JTN. *2006;13(2):41-4*.

31. Stadhouder A, Buskens E, Vergroesen DA, Fidler MW, de Nies F, Öner FC. Non-operative treatment of thoracic and lumbar spine fractures: a prospective randomized study of different treatment options. J Orthop Trauma. *2009;23(8):588-94*.

32. Holtslag HR, Post MW, Lindeman E, Van der Werken C. Long-term functional health status of severely injured patients. Injury. *2007;38(3):280-9*.

33. Stadhouder A, Buckens CF, Holtslag HR, Öner FC. Are existing outcome instruments suitable for assessment of spinal trauma patients? J Neurosurg Spine. 2010;13(5):638-47.

34. Chou T-Y, Tsuang F-Y, Hsu Y-L, Chai CL. Surgical Versus Non-Surgical Treatment for Thoracolumbar Burst Fractures Without Neurological Deficit: A Systematic Review and Meta-Analysis. Global Spine Journal. *2023;14(2):740-9*.

35. Ghobrial GM, Maulucci CM, Maltenfort M, Dalyai RT, Vaccaro AR, Fehlings MG, et al. Operative and nonoperative adverse events in the management of traumatic fractures of the thoracolumbar spine: a systematic review. Neurosurgical Focus FOC. *2014;37(1):E8*.

36. Gnanenthiran SR, Adie S, Harris IA. Nonoperative versus Operative Treatment for Thoracolumbar Burst Fractures Without Neurologic Deficit: A Meta-analysis. Clinical Orthopaedics and Related Research®. *2012;470(2):567-77*.

37. Häckel S, Stienen MN, Martens B, Neuhaus V, Albers CE. Study Protocol for a Randomized Controlled Clinical Trial on the Outcome of Surgical Versus Primary Nonsurgical Treatment of Traumatic Thoracolumbar Spine Burst Fractures in Patients Without Neurological Symptoms–A34RCT. Neurosurgery Practice. *2024;5(2)*.

38. Wood KB, Buttermann GR, Phukan R, Harrod CC, Mehbod A, Shannon B, et al. Operative Compared with Nonoperative Treatment of a Thoracolumbar Burst Fracture without Neurological Deficit: A Prospective Randomized Study with Follow-up at Sixteen to Twenty-Two Years: *. JBJS. *2015;97(1):3-9*.

39. Camino-Willhuber G, Bigdon S, Dandurand C, Dvorak MF, Öner CF, Schnake K, et al. Expert Opinion, Real-World Classification, and Decision-Making in Thoracolumbar Burst Fractures Without Neurologic Deficits? Global Spine Journal. 2024;14 (1_suppl):49S-55S.

40. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health. *2012;66(12):1182-6*.

41. Leatherdale ST. Natural experiment methodology for research: a review of how different methods can support real-world research. International Journal of Social Research Methodology. *2019;22(1):19-35*.

42. Canseco JA, Paziuk T, Schroeder GD, Dvorak MF, Öner CF, Benneker LM, et al. Interobserver Reliability in the Classification of Thoracolumbar Fractures Using the AO Spine TL Injury Classification System Among 22 Clinical Experts in Spine Trauma Care. Global Spine Journal. *2024;14(1_suppl):17S-24S*.

43. Dvorak MF, Öner CF, Schnake K, Dandurand C, Muijs S. From Radiographic Evaluation to Treatment Decisions in Neurologically Intact Patients With Thoraco-lumbar Burst Fractures. Global Spine Journal. *2023;14(1_suppl):4S-7S*.

44. Innovation AC. Thoracolumbar burst fractures (AOSpine A3, A4) in neurologically intact patients: an observational, multicenter cohort study comparing surgical versus non-surgical treatment. 2021. *clinicaltrials. gov. 2022***.**

45. Dandurand C, Öner CF, Hazenbiller O, Bransford RJ, Schnake K, Vaccaro AR, et al. Understanding Decision Making as It Influences Treatment in Thoracolumbar Burst Fractures Without Neurological Deficit: Conceptual Framework and Methodology. Global Spine J. *2024;14(1_suppl):8S-16S*.

46. Dandurand C, Dvorak MF, Hazenbiller O, Bransford RJ, Schnake KJ, Vaccaro AR, et al. Using Equipoise to Determine the Radiographic Characteristics Leading to Agreement on Best Treatment for Thoracolumbar Burst Fractures Without Neurologic Deficits. Global Spine Journal. *2024;14(1_suppl):25S-31S*.

47. Mhaskar R, B BB, Djulbegovic B. At what level of collective equipoise does a randomized clinical trial become ethical for the members of institutional review board/ethical committees? Acta Inform Med. 2013;21(3):156-9.

48. Ghogawala Z, Terrin N, Dunbar MR, Breeze JL, Freund KM, Kanter AS, et al. Effect of Ventral vs Dorsal Spinal Surgery on Patient-Reported Physical Functioning in Patients With Cervical Spondylotic Myelopathy: A Randomized Clinical Trial. JAMA. *2021;325(10):942-51*.

49. Stadhouder A, van Rossenberg LX, Kik C, Muijs SPJ, Öner FC, Houwert RM. Natural Experiments as a Study Method in Spinal Trauma Surgery: A Systematic Review. Global Spine J. *2024;14(5):1640-9*.

11

Dankwoord

Het afronden van dit proefschrift heeft wat voeten in de aarde gehad. Dit was niet gelukt zonder de hulp van veel mensen. Ik wil iedereen heel hartelijk bedanken: mijn collega's in de ziekenhuizen waar ik heb gewerkt en werk, mijn vrienden die bleven informeren wat de stand van zaken was en natuurlijk mijn familie. Tim Jiya en Ricardo Feller wil ik bedanken dat ze mijn paranymfen willen zijn en daarnaast geweldige collega's. Tevens mijn co-promotoren Marijn Houwert en Sander Muijs welken een finale stimulans gaven in de afronding. Ook wil ik de promotie commissie bedanken voor het lezen en beoordelen van dit proefschrift. Onderstaande mede auteurs ben ik ook veel dank verschuldigd:

Stan Buckens
Erik Buskens
Sayf Faraj
Malcolm Fidler
Herman Holtslag
Charlotte Kik
Luuk de Klerk
Mattie van Rijen
Luke van Rossenberg
Ab Verbout
Diederik Vergroesen
Alex Vaccaro
Jan Verhaar
Bryan van de Wall
Owen Williamson

Dan uiteraard mijn promotor Cumhur Öner voor zijn geduld.

Als laatste Francesca voor de onvoorwaardelijke steun.

12

The author

About the Author

Agnita Stadhouder was born on the 14th of october 1972 in Enkhuizen, the Netherlands. After finishing secondary school, VWO RSG Enkhuizen, she started her study Medicine at the Vrije Universiteit in Amsterdam. During her studies she went to India for 9 months and decided that orthopedic surgery was the next step in her career. She started as a resident not in training at the Onze Lieve Vrouwe Gasthuis in Amsterdam and later the Reinier de Graaf Gasthuis in Delft. In 2007 she started her orthopedic residency training with general surgery in Delft (head dr de Graaf) and continued orthopedic surgery at the Utrecht Medical University Center in Utrecht (head prof Verbout and prof Castelein). She did part of the resident training also at the Onze Lieve Vrouwe Gasthuis (head dr Willems). This training also included one year of fulltime research under the supervision of professor Öner. After finishing as orthopedic surgeon she did a spine fellowship in the UMCU (prof FC Öner and prof Castelein) and later was appointed staff member in the same hospital combined with a position as orthopedic surgeon in the Sint Maartenskliniek in Woerden (dr Pavlov). In 2010 she changed jobs and started working as a spine surgeon at VU University Medical Center in Amsterdam (head prof van Royen). Prof Marinus de Kleuver also joined the spine section of de VU University Medical Center as an associate professor. In 2012 she visited prof Mazda in Paris, did a paediatric spine fellowship in Toronto (dr Zeller en dr Lewis) in 2015 and visited dr Miladi in Paris in 2019. In 2018 she combined her academic activities in the VU University Hospital with work in the Dijklander hospital in Hoorn focusing on idiopathic scoliosis and general spine surgery together with Eric Kraaneveld. From 2011 until 2020 she was a board member and President of the Dutch Spine Society. In 2018 the two University Hospitals in Amsterdam merged and her work address changed to the Meibergdreef in Amsterdam with professor Kerkhoffs as head of the department. She lives together with Francesca and has a son Renzo and daughter Senna.

13

Publications

Eur Spine J 2024 Risk factors for neurophysiological events related to intraoperative halo-femoral traction in spinal deformity surgery. Hanneke I Berends, Agnita Stadhouder, Barend J van Royen, Henricus L Journée , Alida A Gouw

Global Spine J 2023 Natural Experiments as a Study Method in Spinal Trauma Surgery: A Systematic Review. Agnita Stadhouder , Luke Xander van Rossenberg, Charlotte Kik , S P J Muijs , F C Öner , R Marijn Houwert

Injury 2023 Natural experiments for orthopaedic trauma research: An introduction Bryan J M van de Wall , Agnita Stadhouder , R Marijn Houwert , F Cumhur Öner, Frank J P Beeres , Rolf H H Groenwold ; NEXT Study Group

Journal of Thoracic Oncology Clinical and research reports 2023 Superior Sulcus Tumors Invading the Spine: Multimodal Treatment Outcomes From the Preimmunotherapy Era. S Unal, R Feller, A Stadhouder, DJ Heineman, TU Jiya, M van Dorp, I Bahce, J Braun, S Senan, M Dahele, C Dickhoff

European Spine Journal 2023 Estimated cumulative radiation exposure in patients treated for adolescent idiopathic scoliosis. J Cool, GJ Streekstra, J van Schuppen, A Stadhouder, J C van den Noort, B J van Royen

Spine deformity 2022 The risk of delayed spinal cord injury in paediatric spinal deformity surgery. JPH.J Rutges, JJM Renkens, DHR. Kempen, C Faber, A Stadhouder, MC Kruyt, A Mostert, PP Horsting, LWL de Klerk, M de Kleuver, RM Castelein & TPC Schlösser

Spine deformity 2023 Current models to understand the onset and progression of scoliotic defromities in adolescent idiopathic scoliosis: a systematic review. Meiring A, de Kater EP, Stadhouder A, van Royen BJ, Breedveld P, Smit TH

Injury 2022 Natural experiments for orthopaedic trauma research: an introduction. Bryan JM van de Wall, Agnita Stadhouder, R Marijn Houwert, F Cumhur Öner, Frank JP Beeres, Rolf HH Groenwold.

Stud Health Technol Inform 2021 The mechanobiology of differential growth in Adolescent Idiopathic Scoliosis. Smit TH, Stadhouder A.

The Journal of Bone and Joint Surgery 2021 High Failure Rates of a Unilateral Posterior Peri Apical Distraction Device (ApiFix) for Fusionless Treatment of Adolescent Idiopathic Scoliosis. Stadhouder, Agnita MD,a; Holewijn, Roderick M. MD, PhD; Haanstra, Tsjitske M. PhD; van Royen, Barend J. MD, PhD; Kruyt, Moyo C. MD PhD; de Kleuver, Marinus MD

Neurosurg Focus 2019 The biomechanical effect of single-level laminectomy and posterior instrumentation on spinal stability in degenerative lumbar scoliosis: a human cadaveric study Christine M E Rustenburg, Sayf S A Faraj, Roderick M Holewijn, Idsart Kingma,, Barend J van Royen, Agnita Stadhouder, Kaj S Emanuel

Acta Orthopedica Belgica. 2019 Management of spinal trauma patients: a national survey in the Netherlands. Stadhouder A, Faraj S, Öner FC

Journal of Orthopaedic Case reports 2018 Parvimonas micra Spondylodiscitis: A Case Report and Systematic review of the Literature. D.C. van Duijvenbode, J.W. Kuiper, R.W. Holewijn, A. Stadhouder

Spine Journal 2019 Prediction of mechanical complications in adult spinal deformity surgery- the GAP score versus the Schwab classification. Jacobs E, van Royen BJ, van Kuijk SMJ, Merk JMR, Stadhouder A, van Rhijn LW, Willems PC.

Spine 2019 Epidemiology of Spinal Fractures in a Level One Trauma Center in the Netherlands: A 10 Years Review Lars P den Ouden, Arjen J Smits, Agnita Stadhouder, Ricardo Feller, Jaap Deunk, Frank W Bloemers

European Spine Journal 2018 Timing of surgery in traumatic spinal cord injury: a national, multidisciplinary survey. Ter Wengel PV, Feller RE, Stadhouder A, Verbaan D, Öner FC, Goslings JC, Vandertop P.

BMJ open 2018 Is post-operative bracing after pedicle screw fixation of spine fractures necessary? Study protocol of the ORNOT study; a randomized controlled multicenter trial *A. Smits, J. Deunk, A. Stadhouder, M.C. Altena, D. H. R. Kempen, F. W. Bloemers*

Neuropediatrics 2018 Risk factors for dystonia after selective dorsal rhizotomy in nonwalking children and adolescents with bilateral spasticity. L. van de Pol, J vermeulen, C van `t Westende, P. van Schie, E. Bolster, P. van Ouwerkerk, R. Strijers, J. Becher, A. Stadhouder, P. de Graaf, A. Buizer.

Global Spine J. 2017 A novel spinal implant for fusionless scoliosis correction: a biomechanical analysis of a posterior peri-apical concave distraction device. R.M. Holewijn, M. De Kleuver, A.J. van der Veen, K.S. Emanuel, A. Bisschop, A. Stadhouder, B.J. van Royen, I. Kingma

Eur J Paediatr Neurol. 2017 Effect of selective dorsal rhizotomy on daily care and comfort in non-walking children and adolescents with severe spasticity. Buizer AI, van Schie PE, Bolster EA, van Ouwerkerk WJ, Strijers RL, van de Pol LA, Stadhouder A, Becher JG, Vermeulen RJ.

Spine 2017 Restrained differential growth: the initiating event of adolescent idiopathic scoliosis.
Crijns TJ, Stadhouder A, Smit TH.

J Emerg Med. 2016 The Clay Shoveler's fracture: a case report and review of the literature. Posthuma de Boer J, van Wulfften Palthe AF, Stadhouder A, Bloemers FW.

Ann Surg Oncol. 2016 Complications after percutaneous pedicle screw fixation for the treatment of unstable spinal metastases. A.L. Versteeg, J.J. Verlaan, P. de Baat, T.U. Jiya, A. Stadhouder, C.H. Diekerhof, G. B. van Solinge, F.C. Öner

Spine Deform 2015 How Does Spinal Release and Ponte Osteotomy Improve Spinal Flexibility? The Law of Diminishing Returns. Roderick M Holewijn, Tom P C Schlösser, Arno Bisschop, Albert J van der Veen, Agnita Stadhouder, Barend J van Royen, René M Castelein, Marinus de Kleuver

Global Spine J 2015 The effects of single-level instrumented lumbar laminectomy on adjacent spinal biomechanics Arno Bisschop, Roderick M Holewijn, Idsart Kingma, Agnita Stadhouder, Pieter-Paul A Vergroesen, Albert J van der Veen, Jaap H van Dieën, Barend J van Royen

Clin Biomech 2014 Single level lumbar laminectomy alters segmental biomechanical behavior without affecting adjacent segments. Arno Bisschop, Susanne J P M van Engelen, Idsart Kingma, Roderick M Holewijn, Agnita Stadhouder, Albert J van der Veen, Jaap H van Dieën, Barend J van Royen

Eur Spine J 2014 The outcome and survival of palliative surgery in thoraco-lumbar spinal metastases: contemporary retrospective cohort study
R M Nemelc, A Stadhouder, B J van Royen, T U Jiya

Spine J 2014 Reliability and validity of the adapted Dutch version of the revised Scoliosis Research Society 22-item questionnaire. Tom P C Schlösser, Agnita Stadhouder, Janneke J P Schimmel, A Mechteld Lehr, Geert J M G van der Heijden, René M Castelein

Scoliosis 2011 The flying buttress construct for posterior spinopelvic fixation: a technical note. Barend J van Royen, Martijn van Dijk, Dirk P van Oostveen, Bas van Ooij, Agnita Stadhouder

Journal of Neurosurgery Spine. 2010 Are existing outcome measurements suitable for assesment of spinal trauma patients? A Stadhouder, CFM Buckens, HR Holtslag, FC Öner

Journal of Orthopedic Trauma. 2009 Non-operative treatment of thoracic and lumbar spine fractures: a prospective randomized study of different treatment options.
A Stadhouder, E Buskens, DA Vergroesen, MW Fidler, F de Nies, FC Öner

Spine. 2008 Traumatic thoracic and lumbar spine fractures: operative or nonoperative treatment: comparison of two treatment strategies by means of surgeon equipoise
A Stadhouder, E Buskens, LW de Klerk, JA Verhaar, WA Dhert, AJ Verbout, AR Vaccaro, FC Öner

Spine Journal. 2008 Surgeon equipoise as an inclusion criterium for the evaluation of nonoperative versus operative treatment of thoracolumbar spinal fractures
A Stadhouder, FC Öner, KW Wilson, AR Vaccaro, OD Williamson, AJ Verbout, JA Verhaar, LW de Klerk, E Buskens

BOOK-CHAPTERS

Textbook 'The Spine' Editor A. Vaccaro. Chapter: Nonoperative Treatment of Fractures of the Thoracolumbar Spine. 2019
F.C Öner, A. Stadhouder