

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

Global Spine Journal  
2024, Vol. 14(5) 1640–1649  
© The Author(s) 2023  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/21925682231220889  
[journals.sagepub.com/home/gsj](https://journals.sagepub.com/home/gsj)



Agnita Stadhouders, MD<sup>1</sup> , Luke Xander van Rosenberg, MD<sup>2,3</sup> ,  
Charlotte Kik, MD<sup>4</sup>, S. P. J. Muijs, MD, PhD<sup>5</sup>, F. C. Öner, MD, PhD<sup>5</sup>, and  
R. Marijn Houwert, MD, PhD<sup>6</sup>

## 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.

## Keywords

trauma, methodology, natural experiment, systematic review, spinal fracture

## 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.<sup>1–3</sup> However, randomized controlled trials in surgical fields may encounter certain difficulties which reduce reliability of results and allow introduction of bias.<sup>4–6</sup> 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.<sup>7–11</sup> Methodical difficulties include

<sup>1</sup> Department of Orthopaedics and Sports Medicine, Amsterdam University Medical Centers, Amsterdam, Netherlands

<sup>2</sup> Faculty of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland

<sup>3</sup> Department of Trauma Surgery, Diaconessenhuis, Utrecht, Netherlands

<sup>4</sup> Department of Neurosurgery, Erasmus MC, Rotterdam, Netherlands

<sup>5</sup> Department of Orthopaedics, University Medical Center Utrecht, Utrecht, Netherlands

<sup>6</sup> Department of Trauma Surgery, University Medical Center Utrecht, Utrecht, Netherlands

### Corresponding Author:

Agnita Stadhouders, MD, Department of Orthopaedics and Sports Medicine, Amsterdam University Medical Centers, Meibergdreef 9, Amsterdam 1105 AZ, Netherlands.

Email: [a.stadhouders@amsterdamumc.nl](mailto:a.stadhouders@amsterdamumc.nl)



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further

permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

challenges in acquiring informed consent, blinding of patients and randomization of patients.<sup>9,11,12</sup> In acute surgery fields, where urgent lifesaving treatment is often involved, randomized controlled trials are difficult to conduct properly due to these challenges. This is also the case for spinal trauma care.<sup>10,13</sup> Other study designs might pose a more viable solution.<sup>9</sup>

Observational studies have historically been used to demonstrate credible results in situations where a randomized controlled trial is either unethical or unfeasible.<sup>14</sup> However, observational studies are more prone to bias and confounding.<sup>14–16</sup> To minimize confounding, observational studies must be carefully and rigorously designed.<sup>17</sup> In therapeutic studies a randomized design has greater value and credibility of results compared to observational studies and Vandembroucke<sup>17</sup> 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 Vandembroucke<sup>17</sup> in the *Lancet* in 2004.<sup>17</sup> 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.<sup>17,18</sup> 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.”<sup>18</sup> 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.<sup>19</sup> Clinical equipoise is “a state of genuine uncertainty on the part of the clinical investigator regarding the comparative therapeutic merits of different treatment options”<sup>20</sup> 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.<sup>21</sup> A natural experiment becomes feasible when clinical equipoise is present and allocation of treatment is dependent on external factors.<sup>21</sup> This is especially true for trauma patients. Generally, trauma patients will receive acute

care at the nearest hospital able to facilitate adequate treatment.<sup>22,23</sup> 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.<sup>21</sup> Multiple natural experiments in trauma surgery have been conducted reporting results matching the credibility of randomized controlled trials.<sup>24–28</sup> 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 30<sup>st</sup> 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

**Table 1.** Summary of Inclusion and Exclusion Criteria.

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

NE, natural experiment.

<sup>a</sup>(RCT, case series, case control, case report, observational cohort studies without NE\* design).

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’).<sup>29</sup> 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. 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.<sup>30</sup> Further information on the assessment of methodological quality is provided in [Appendix B](#).

## 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.<sup>13,31–33</sup>

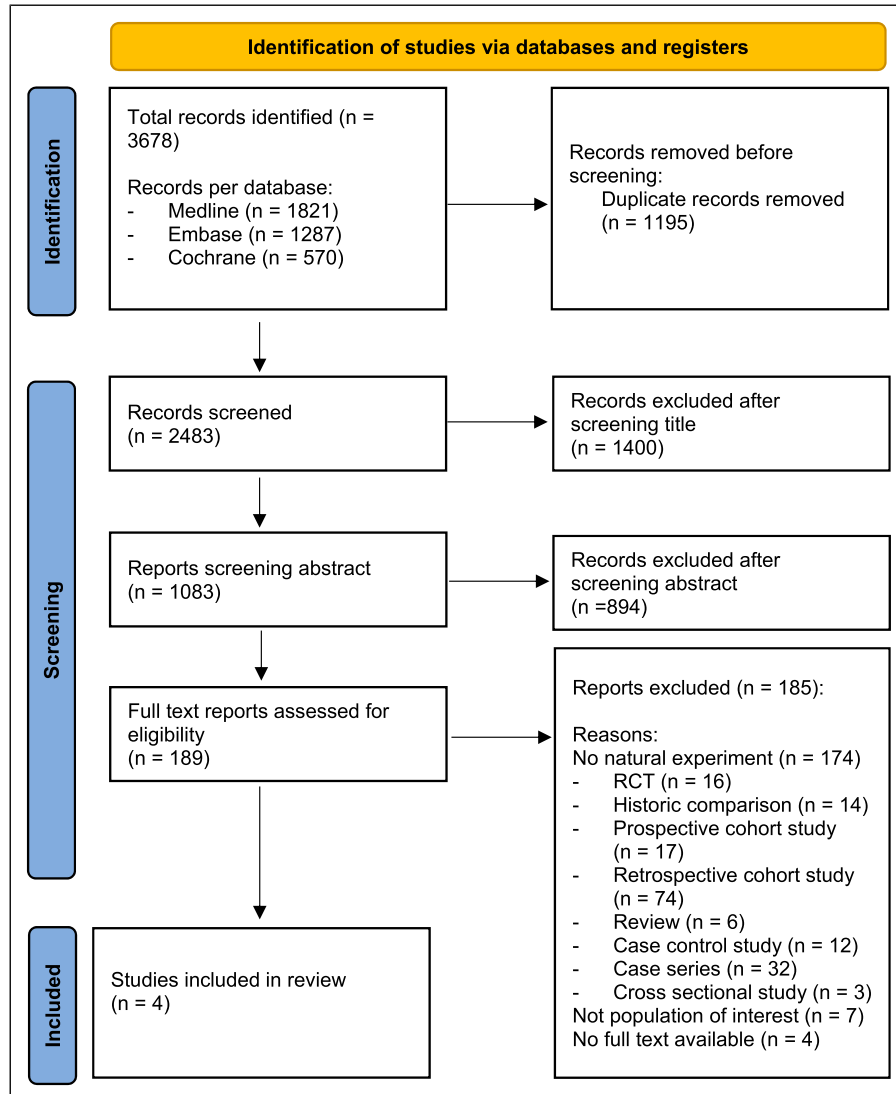
All four studies were designed retrospectively,<sup>13,31–33</sup> of which one study retrieved participant retrospectively but collected patient reported outcomes actively of included participants.<sup>13</sup> Three studies were performed in Western Europe,<sup>13,31,32</sup> while one was conducted in the United States of America.<sup>33</sup> Two studies included a multi-center setting.<sup>13,31</sup> All participating hospitals in the studies were level-1 trauma centers.<sup>13,31–33</sup> Three studies included a comparison between different surgical treatment modalities for acute spinal fractures,<sup>13,31,32</sup> while one study focused on the rate of infectious complications by the use of vancomycin powder in posterior spinal stabilization of traumatic injuries.<sup>33</sup> Of the 852 participants in the included studies 481 (56%) were male, and the mean age ranged from 37 to 69 years.<sup>13,31–33</sup> Mean follow-up ranged from 6 months to 74 months.<sup>13,31–33</sup> Two studies reported the injury severity score (ISS) of which all participants had mean ISS scores ≤ 16.<sup>31,32</sup> 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).<sup>13,31,32</sup> 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%).<sup>13,31–33</sup> Three studies reported presence of neurological impairment, of all 765 patients 157 (20.5%) were partially or completely neurologically impaired.<sup>13,31,32</sup> One study excluded patients with cervical fractures and/or neurological impairment.<sup>33</sup>

### 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,<sup>13</sup> followed by O’Neill et al and Erichsen et al both with score of 16.<sup>31–33</sup> Myers et al scored lowest with a score of 12.<sup>32</sup> 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).<sup>13,31–33</sup> 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



**Figure 1.** Flowchart of study selection. From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71.

retrieval began in 2004, and the study was published in 2011.<sup>33</sup> The study demonstrated the possibility to use a natural experiment design in pharmaceutical studies.

Similarly, Stadhouders 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.<sup>13</sup> 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.<sup>31,32</sup> 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

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

Study and Year	Study Period	Design	Country	Natural Experiment Comparison <sup>a</sup>	Number of Patients		Age in Years, Mean (SD)/ Range		Gender Male, n (%)	Follow Up in Months, Mean (SD)/Range	Mean (SD) ISS <sup>b</sup>	Trauma Mechanism, n (%)				Fracture Location, n (%)		Neurological Impairment, n (%)
					Total	Case/Control	Case/Control	Case/Control				Case/Control	Case/Control	Fall	Traffic Accident	Other	Cervical	
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	na	8.7 (4.6)/8.9 (6.9)	446 (95.9%) <sup>c</sup>	15 (3.2%) <sup>c</sup>	4 (0.8%) <sup>c</sup>	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) <sup>c</sup>	ISS ≤ 16 (no SD) <sup>c</sup>	21 (47.7%)/30 (69.8%)	9 (20.4%)/11 (25.5%)	14 (31.8%)/2 (4.6%)	Excluded	87 (100%) <sup>c</sup>	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%)		
Stadhouder, 2008	1991-2002	Retrospective	The Netherlands	Schools, MS	190	95/95	38.5 (18-84)/37.1 (18-79)	50 (52%)/64 (67%)	74.4 (no SD) <sup>c</sup>	na	46 (48%)/42 (44%)	27 (28%)/26 (27%)	22 (2.4%)/27 (29%)	61 (32.1%) <sup>c</sup>	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.

<sup>a</sup>What level of comparison (high to low = schools/expertise centers, departments, surgeons), SS single center, MS, multicenter.

<sup>b</sup>Injury severity score.

<sup>c</sup>Only overall value available, no distinction between case and control groups.



**Table 3.** MINORS Score.

MINORS Quality Assessment of Included Studies in a Systematic Review of Natural Experiments in Spinal Trauma Surgery				
Criteria	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 <sup>a</sup>	0	0	0	0
Endpoints appropriate to the aim of the study	2	2	2	2
Unbiased assessment of the study endpoint	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% <sup>a</sup>	0	0	0	0
Prospective calculation of the study size <sup>a</sup>	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).

<sup>a</sup>All included studies are retrospectively designed, scores range from 0-24 for comparative studies and 0-18 for retrospective comparative studies.

period using the Oswestry 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.<sup>31</sup> 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.<sup>32</sup> Both research groups had similar baseline characteristics, admittance practice strategies and exclusion rates.<sup>31,32</sup> 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.<sup>31,32</sup>

## 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.<sup>13,31-33</sup> 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.<sup>13,31-33</sup> 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,<sup>34</sup> treatment strategy of C2 fractures in the elderly<sup>35</sup> and timing

of intervention in patients with spinal cord injury.<sup>36</sup> For this matter natural experiments can be of value since within spinal trauma treatment, the different schools and treatments are common practice already.<sup>37-40</sup> 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.<sup>41</sup> As are the common practice of Electronic Medical Records (EMR) in hospitals,<sup>42,43</sup> and Patient Reported Outcome Measurements (PROMS) prospectively gathered in specific patient groups.<sup>44</sup> In the included retrospective natural experiment study of Stadhouder et al<sup>13</sup> 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.<sup>45</sup> 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.<sup>45</sup> Results therefore should be carefully interpreted when there is a high and difference between groups number of drop outs.<sup>45,46</sup>

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.<sup>47</sup> 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’.<sup>48</sup> 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.<sup>18</sup> 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).<sup>13,31–33</sup> A shorter duration of study time can be a contributing factor to conducting research in quickly developing specialties as orthopedic trauma and spine surgery.<sup>18,49</sup>

Two included papers were published more than 10 years ago,<sup>13,33</sup> two papers more recently.<sup>31,32</sup> 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 non-profit collaboration of clinical researchers in the field of emergency and (orthopedic) trauma surgery.<sup>50</sup> They so far published four relevant papers with a natural experiment as methodology and more studies are being conducted and soon to be published.<sup>27,28,51,52</sup> One study showed that with a natural experiment design on rib fixation there was no difference in outcome between nonoperative and operative treated patients.<sup>27</sup> The inclusion was finished one year earlier than predicted and took three years.<sup>27</sup> Before this publication, a RCT was conducted in Australia which took four years and where almost half of the eligible patients refused to participate in this study.<sup>53</sup> 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,<sup>53</sup> comparable with the natural experiment paper.<sup>27</sup> 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.<sup>29</sup> It comprises twelve items with a maximum score of 24, that applies to meticulously designed RCT’s.<sup>29</sup> The studies included in our paper had a score of respectively 12, 16, 16 and 17 points.<sup>13,31–33</sup> 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, we consider 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<sup>54</sup> 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.”<sup>54</sup> Another more recent systematic review by Sathish et al evaluated 96 systematic reviews published in spine surgery.<sup>55</sup> Reviews were scored by the AMSTAR score (A measurement Tool to Assess systematic Reviews),<sup>56</sup> PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)<sup>57</sup> and MOOSE (Meta-analyses Of Observational Studies in Epidemiology).<sup>58</sup> 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.<sup>48</sup> As stated in a previous published paper on natural experiments we suggest to collect data on key prognostic patient factors, either prospective or retrospective.<sup>18</sup> Further, it is important to correct for confounding by stratification, regression adjustment or matching.<sup>18</sup> Another solution is to use clinical equipoise as an inclusion criterion.<sup>10,18–20,45,51</sup> 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.<sup>10,18,51</sup> In this review one paper used an expert panel,<sup>13</sup> the other three studies did not.<sup>31–33</sup>

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.<sup>45</sup> 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.<sup>45</sup> 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.



## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iDs

Agnita Stadhouders  <https://orcid.org/0000-0001-8882-3842>

Luke Xander van Rossenberg  <https://orcid.org/0000-0002-6649-5910>

## Supplemental Material

Supplemental material for this article is available online.

## References

- 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.
- Kaptchuk TJ. The double-blind, randomized, placebo-controlled trial: gold standard or golden calf? *J Clin Epidemiol.* 2001;54(6):541-549.
- Groenwold RH. [Three types of bias: distortion of research results and how that can be prevented]. *Ned Tijdschr Geneesk.* 2013;157(40):A6497.
- Meakins JL. Innovation in surgery: the rules of evidence. *Am J Surg.* 2002;183(4):399-405.
- Pollock AV. Surgical evaluation at the crossroads. *Br J Surg.* 1993;80(8):964-966.
- Stirrat GM, Farrow SC, Fardon J, Dwyer N. The challenge of evaluating surgical procedures. *Ann R Coll Surg Engl.* 1992;74(2):80-84.
- 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.
- Chalmers TC. Randomization of the first patient. *Med Clin.* 1975;59(4):1035-1038.
- McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ.* 2002;324(7351):1448-1451.
- Stadhouders A, Oner 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.
- 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.
- 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.
- Stadhouders 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.
- Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med.* 2000;342(25):1878-1886.
- 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.
- 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.
- Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet.* 2004;363(9422):1728-1731.
- van de Wall BJM, Stadhouders A, Houwert RM, Oner FC, Beeres FJP, Groenwold RHH, NEXT Study Group. Natural experiments for orthopaedic trauma research: an introduction. *Injury.* 2023;54(2):429-434.
- Khullar D, Jena AB. "Natural experiments" in health care research. *JAMA Heal For.* 2021;2(6):e210290.
- Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med.* 1987;317(3):141-145.
- 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.
- 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.
- 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.
- 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.
- 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.
- Beeres FJP, Hallensleben ND, 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.
- 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.
- 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.
- Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors):

- 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 Musculoskel 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, Försth 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. Oner 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, Prömmel 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, Oner 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. Post-laminectomy 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.
  48. 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 nonoperative 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 in spine 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.