





## ARTICLE

# Sample sizes and statistical methods in interventional studies on individuals with spinal cord injury: A systematic review

Georg Zimmermann<sup>1,2,3</sup>  | Lisa-Maria Bolter<sup>3</sup> | Ronny Sluka<sup>4</sup> | Yvonne Höller<sup>1</sup>  |  
 Arne C. Bathke<sup>3</sup>  | Aljoscha Thomschewski<sup>1,2,4</sup> | Stefan Leis<sup>1</sup> | Simona Lattanzi<sup>5</sup> |  
 Francesco Brigo<sup>6,7</sup>  | Eugen Trinko<sup>1,8</sup>

<sup>1</sup>Department of Neurology, Christian Doppler Medical Centre, Paracelsus Medical University, Salzburg, Austria

<sup>2</sup>Spinal Cord Injury and Tissue Regeneration Centre Salzburg, Paracelsus Medical University, Salzburg, Austria

<sup>3</sup>Department of Mathematics, Paris Lodron University, Salzburg, Austria

<sup>4</sup>Department of Psychology, Paris Lodron University, Salzburg, Austria

<sup>5</sup>Neurological Clinic, Department of Experimental and Clinical Medicine, Marche Polytechnic University, Ancona, Italy

<sup>6</sup>Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona, Italy

<sup>7</sup>Division of Neurology, Franz Tappeiner Hospital, Merano, Italy

<sup>8</sup>Department of Public Health, Health Services Research and Health Technology Assessment, UMIT, Hall i. T., Austria

## Correspondence

Georg Zimmermann, Spinal Cord Injury and Tissue Regeneration Centre Salzburg, Paracelsus Medical University, Strubergasse 22, A 5020 Salzburg, Austria.  
 Email: georg.zimmermann@pmu.ac.at

## Abstract

**Aim:** Prevalence and incidence of spinal cord injury (SCI) are low. However, sample sizes have not been systematically examined yet, although this might represent useful information for study planning and power considerations. Therefore, our objective was to determine the median sample size in clinical trials on SCI individuals. Moreover, within small-sample size studies, statistical methods and awareness of potential problems regarding small samples were examined.

**Methods:** We systematically reviewed all studies on human SCI individuals published between 2014 and 2015, where the effect of an intervention on one or more health-related outcomes was assessed by means of a hypothesis test. If at least one group had a size <20, the study was classified as a small sample size study. PubMed was searched for eligible studies; subsequently, data on sample sizes and statistical methods were extracted and summarized descriptively.

**Results:** Out of 8897 studies 207 were included. Median total sample size was 18 (range 4-582). Small sample sizes were found in 167/207 (81%) studies, resulting limitations and implications for statistical analyses were mentioned in 109/167 (65%) studies.

**Conclusions:** Although most recent SCI trials have been conducted with small samples, the consequences on statistical analysis methods and the validity of the results are rarely acknowledged.

## KEYWORDS

effect size, power, rare disease, spinal cord trauma, statistical data analysis

## 1 | INTRODUCTION

Spinal cord injury (SCI) individuals have to face various impairments, ranging from deficits in motor function to further complications, which may be even more serious and life-threatening, such as urinary tract infections and bladder dysfunction,<sup>1,2</sup> or cardiovascular and respiratory diseases.<sup>3,4</sup> Recent studies indicate that the prevalence of SCI is about 25/100 000,<sup>5</sup> although estimates vary from 11 to

112/100 000 due to methodological and regional differences.<sup>6,7</sup> Annual incidences range from about 8 to 80/1 000 000.<sup>7</sup> In order to improve outcomes in SCI and the persons' quality of life, numerous therapeutic approaches have been investigated. The scope ranges from pharmacological therapies<sup>8</sup> to interventions targeting brain reorganization, neuroplasticity induced by stimulation techniques (eg, transcranial magnetic stimulation<sup>9</sup>) as well as rehabilitative and training programs for SCI individuals, either with or without the use of

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors. *Journal of Evidence-Based Medicine* published by Chinese Cochrane Center, West China Hospital of Sichuan University and John Wiley & Sons Australia, Ltd

assistive devices.<sup>10,11</sup> These efforts have been complemented by systematic reviews of the scientific status quo,<sup>12</sup> with the ultimate goal of setting up evidence-based guidelines for treatment and rehabilitation of people with SCI.<sup>13,14</sup>

However, from a methodological point of view, the question whether the relatively small prevalence of SCI and well-known difficulties regarding recruitment and compliance<sup>15</sup> translate to small sample sizes warrants a systematic investigation. Ioannidis considered the probability that a research finding is indeed true,<sup>16</sup> showing that this probability is

- positively associated with the power of a study (ie, the probability that the treatment effect can be detected statistically, given that there is one) as well as the ratio of true to false hypotheses in a particular field, but
- negatively associated with the type I error level (eg, falsely claiming that a new therapy is superior to the standard one) and the amount of bias.

It should be noted that the power of a study is in turn influenced by several factors: Needless to say that one-sided tests are more powerful than their two-sided counterparts, that the power increases with growing effect size and larger significance level, and that the spread of the data is negatively associated with power. In particular, the latter fact explains why a paired test has higher power than its unpaired counterpart, given that all other specifications are the same. Furthermore, it is well known in mathematical statistics that some frequently used statistical methods may either over- or underreject the respective hypotheses in small sample situations.<sup>17–19</sup> For example, power differences between the unpaired *t*-test and the Wilcoxon-Mann-Whitney test depend on whether or not the assumptions underlying the *t*-test are met. Moreover, power differences may result from the fact that the respective hypotheses being tested are not the same. Summing up, the choice of the statistical analysis method may lead to a further deflation of the probability that a research finding is true. Apart from that, especially small sample sizes might lead to a decrease in power<sup>20</sup> and, thus, also to a decreased post-study-probability. Moreover, small sample sizes are also related to other methodological aspects: for example, there is evidence that an adequate sample size is required for avoiding covariate imbalance, which in turn could affect the validity of the results.<sup>21</sup> Needless to say, however, that it would be inappropriate to judge the quality of SCI studies based on the examination of sample sizes and statistical methods only. Rather, the considerations from above demonstrate what *could* potentially happen, given that sample sizes are small.

The aim of the present investigation is twofold: first, we would like to determine the median sample size in clinical trials on human individuals with SCI. Although we may speculate that the sample sizes are rather small, the principle of evidence-based, systematic research warrants a thorough investigation.<sup>22</sup> Moreover, as a secondary analysis, we consider the subset of studies with small sample sizes—the definition is stated in the Methods section—and examine some key characteristics of the statistical methods that were used as well as the

question to which extent the authors mentioned potential problems related to small sample sizes.

## 2 | METHODS

### 2.1 | Study selection and data extraction

We chose a sample covering the two consecutive years that were most recent at the point we started the evaluation. In order to minimize the risk of bias due to delayed publication and post hoc changes in electronic databases, we considered all research papers, which were published between 01 January 2014 and 31 December 2015, and extracted the full list of abstracts on 23 January 2017 from PubMed. We selected articles that reported results from interventional studies on individuals with traumatic or nontraumatic SCI (eg, ischemia, infections). We used a rather broad definition of an interventional study, including any study where participants were prospectively assigned to at least one health-related intervention, followed by an evaluation of its effect with respect to a health outcome.<sup>23</sup> We deliberately did not restrict ourselves to randomized clinical trials, although this design is considered as the gold standard of interventional studies, because a substantial number of SCI studies might be nonrandomized due to various practical reasons.<sup>15</sup> Moreover, we focused on studies, which included at least one group of humans with traumatic or nontraumatic SCI, thus excluding postmortem examinations, in-vitro studies, case studies and papers dealing with healthy controls or animals only. Furthermore, we included only publications in English language.

In addition to these criteria, which had been specified in advance, further restrictions were set up in the course of the screening process, adhering to the spirit of the pre-defined criteria as closely as possible. First, we refined the distinction between clinical trials and case series, based on whether hypothesis tests or descriptive evaluations of the intervention effect were carried out. Then, we decided to exclude case series, because they are supposed to have less impact on decision-making than clinical trials. Moreover, with respect to the secondary objective of investigating the statistical methods, there would have been no use in considering studies where only descriptive or narrative analyses had been done. Second, a research paper was excluded if the authors solely assessed the reliability of an instrument in the interventional part of the study (eg, reliability of blood pressure responses to a sit-up test<sup>24</sup>), because the focus in these studies was not on improving any participant-relevant outcome. Third, regarding duplicate publications, we applied a quite narrow definition and excluded studies only if they reported results from exactly one and the same sample. If small sample sizes were present and the statistical methods differed between studies, the information concerning the latter was merged. With respect to our main outcomes, this way of dealing with duplicate publications allowed for reducing bias while not losing substantial information.

Screening and data extraction were conducted in the following way: based on a consensus decision of GZ, YH, AB, AT, and SL, the query “spinal cord injury OR tetraplegia OR paraplegia OR tetraparesis OR

paraparesis OR spinal cord infarction OR spinal cord ischemia” was entered in the advanced search form in PUBMED, with the time frame for the date of publication set to “Jan 1, 2014 to Dec 31, 2015.” Since PubMed automatically applies term mapping, additional MeSH term searches were not done. We deliberately did not apply further filters to the searches, in order to minimize the risk of bias due to algorithm-based exclusion. In order to enhance transparency, we provide the text file, which contains the abstracts of the search results, upon request. Due to the considerable amount of studies, we think that this makes more sense than providing lengthy tables containing all individual study details.

As a first step, all abstracts were screened for eligibility by RS under the supervision of GZ, YH, and AT. This part included a pilot phase covering almost 100 abstracts, with the purpose of getting used to the screening process and refining inclusion and exclusion criteria. In order to avoid bias resulting from the fact that only one single person did the abstract screening, studies were only excluded if they clearly failed to meet at least one of the inclusion criteria mentioned above. In case there were doubts at this stage, the corresponding item was left in the list of included studies, postponing the decision to the next screening stage. As a second step, the full texts of all remaining studies—if available—were screened independently by GZ and LB. Studies were either excluded, following the aforementioned criteria, or classified as eligible for subsequent extraction of basic publication characteristics, sample sizes and statistical methods. If consensus could not be reached, the final decision was left to AB (statistician), YH (neuroscientist) or SL (clinician). Although some items were not applicable due to the specific methodological focus of our review, we tried to adhere to recommendations for conducting and reporting systematic reviews<sup>22,25</sup> as closely as possible (the PRISMA checklist is provided as a supplemental file).

## 2.2 | Outcomes and statistical methods

The primary outcomes were the total sample size and the number of individuals with SCI. Both quantities refer to the sizes of the respective groups that were finally included in the statistical analyses rather than to the numbers of participants which had been initially considered as eligible.

As to the secondary outcomes, within the subgroup of “small-sample size studies,” additional information regarding the statistical test(s) used and the respective group sizes were recorded. A study had to meet the following two criteria, in order to qualify as a “small-sample size study.” First, it was required that several groups (eg, different treatment arms in a clinical trial or repeated measurements taken from one single group before and after an intervention) were compared to each other with respect to a primary or secondary outcome by means of a statistical hypothesis test. Second, at least one of those groups had to comprise less than 20 individuals. Although this cutoff is somewhat arbitrary, the underlying rationale is that some frequently used hypothesis tests are based on certain assumptions, which cannot be meaningfully assessed in such scenarios. For example, based on 20 observations, it is impossible to assess whether the data was generated

from a normal distribution or not (Supplementary Figure S1). Apart from that, the power for sample sizes below that cutoff might be considerably decreased even for comparisons of paired groups and moderate effects (please refer to the R code in the supplementary material). Within the small-sample size studies, we also recorded if potentially related statistical problems were mentioned. In this case, a further distinction between “just mentioned,” “rule-of-thumb-like implication for the choice of statistical methods,” and “statistically sound reasoning for the choice of statistical methods” was made. The detailed definitions of these categories are provided in Table S1 in the supplementary material. All variables collected in the present review had been specified by GZ, YH, and AB in advance. The primary and secondary outcomes defined above were analyzed using basic descriptive summary statistics and graphical displays of the collected data.

## 3 | RESULTS

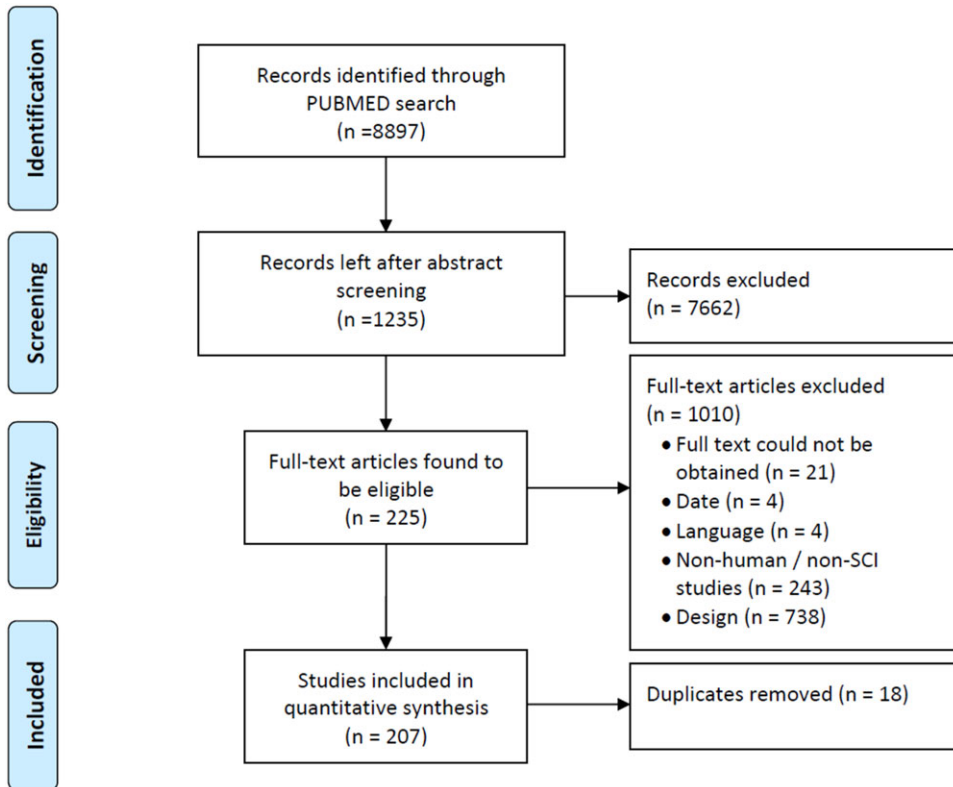
The search on PubMed identified 8897 publications. Applying the exclusion criteria lead to a substantial reduction down to 207 studies, which were eventually considered eligible for qualitative and quantitative syntheses. Details about the selection process can be seen in the flow chart (Figure 1).

### 3.1 | Sample sizes

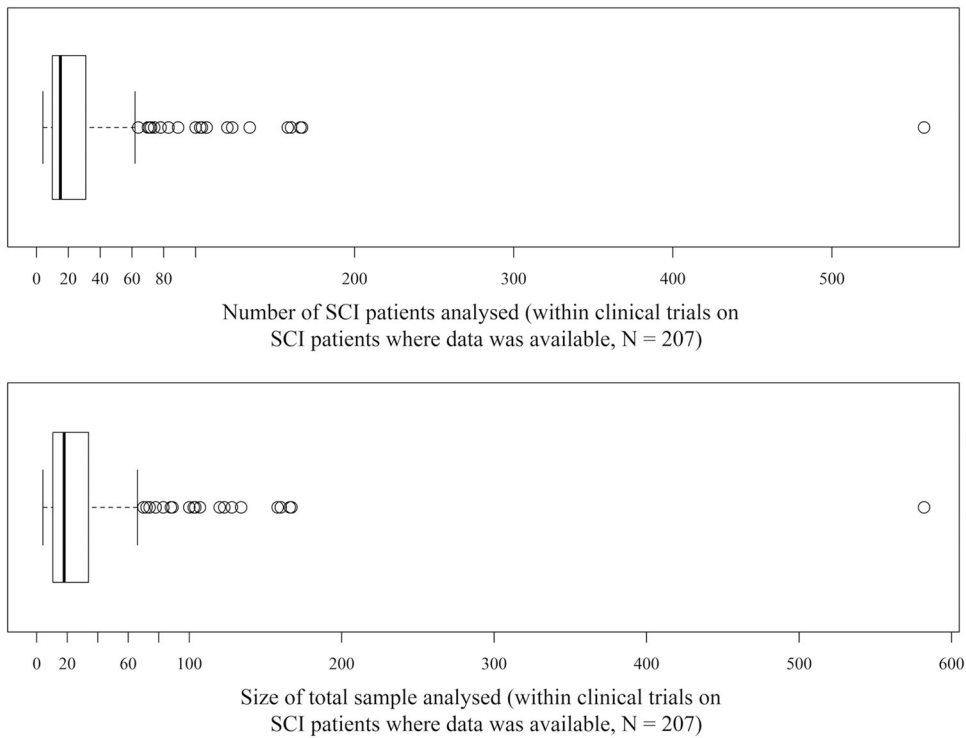
The median number of individuals with SCI was 15 (range: 4-558, interquartile range: 10-32), the median total sample size was 18 (range: 4-582, interquartile range: 10-34), see Figure 2. The subset of studies with a total sample size >100 ( $n = 12$ ) comprised multicenter<sup>26-28</sup> as well as single-center trials.<sup>29-35</sup> We found that 167 out of 207 (81%) eligible studies met our definition of a small-sample size study. Within this subset, the median number of participants with SCI and total sample size were 13 (range: 4-160, interquartile range: 9-19) and 15 (range: 4-160, interquartile range: 10-22), respectively, see Figure 3.

### 3.2 | Statistical methods and mentioning potential problems

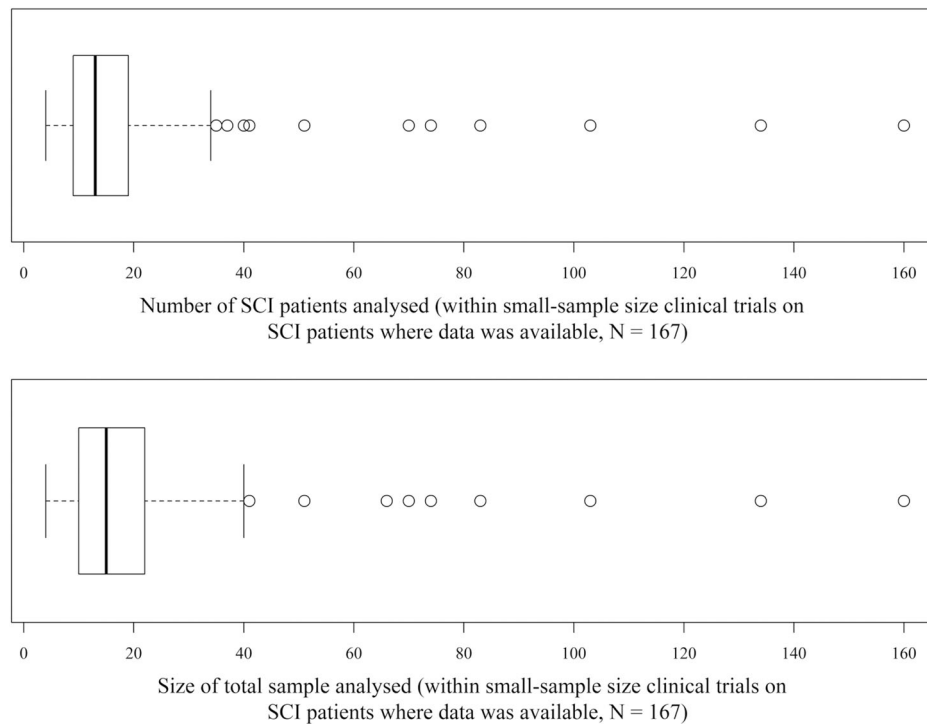
Three of 167 small-sample size studies did not provide sufficient information about the statistical methods and were therefore excluded from subsequent analyses. In the remaining 164 studies, repeated measures analyses were conducted most frequently (parametric/nonparametric comparisons of paired groups:  $n = 143$  studies vs unpaired tests:  $n = 52$  studies). Parametric methods were more often used than nonparametric methods (123 vs 72 studies). The sizes of the groups that were compared showed large variation, due to a broad variety of design specifications (eg, different numbers of repeated measurements). Group sizes were equal in some studies, especially when paired comparisons were made, because balance is being imposed by design in this case (eg, 2 groups of size 14 each<sup>36</sup>). However, in other studies, highly unbalanced scenarios were found (eg, 2 groups of sizes 6 and 10, respectively<sup>37</sup>). Moreover, sometimes,



**FIGURE 1** Flow chart of the screening process, adapted from Ref. (25)



**FIGURE 2** Distribution of the number of participants with spinal cord injury (SCI) and the total sample size in clinical trials on individuals with SCI



**FIGURE 3** Distribution of the number of participants with spinal cord injury (SCI) and the total sample size within the subgroup of small-sample size clinical trials on people with SCI

comparisons of very small groups were reported (eg, pre-post comparison for a group of three individuals<sup>38</sup>; two unpaired groups with three and five individuals, respectively<sup>39</sup>).

The authors did not mention the small sample size issue at all in 58 out of 167 studies (35%). Awareness of potential problems related to small sample size, yet without referring to the choice of a particular statistical method, was present in 92 studies (55%). Different implications thereof were mentioned, ranging from considerable baseline variability, which could not be accounted for,<sup>40,41</sup> to limited generalizability of the results,<sup>42,43</sup> decreased power,<sup>44,45</sup> and the need for replication in larger samples.<sup>46,47</sup> Seventeen studies (10%) stated a rule-of-thumb like justification for the choice of a particular statistical approach, mainly that nonparametric methods were used instead of parametric approaches due to the small sample size.<sup>48-55</sup> However, no study was found where the authors provided a theoretically sound reason (eg, by referring to a simulation study, which indicates a good performance in small sample settings).

## 4 | DISCUSSION

The results indicate that the total sample sizes in SCI trials are often quite small. Only few multicenter clinical trials could be identified, where the number of participants was large. Even in studies with moderate to large sample sizes, comparisons of subgroups were frequently conducted, leading to a proportion of small sample size studies of about 80 percent. Furthermore, we found that within small-sample

size studies, parametric methods were used more often than their nonparametric counterparts, and repeated-measures analyses were conducted most frequently.

Although these results might have some implications for the post-study probability of the research findings being true, as outlined in the Introduction, they are by no means supposed to serve as an assessment of the quality of SCI research in general. It has been pointed out that systematic reviews are highly relevant for evidence-based decision making. If conducted with methodological rigor and based on low-bias studies, they provide a high level of evidence.<sup>22</sup> However, reviews were excluded in the present study, which of course makes sense with respect to the investigation of sample sizes, but nevertheless prevents us from drawing definite conclusions regarding whether or not to trust the results from SCI research in general.

Regarding the statistical methods used, we could come up with some remarkable findings.

First, researchers used parametric methods more often than their nonparametric counterparts. However, parametric tests (eg, *t*-tests) might yield erroneous results in terms of type I error probabilities (ie, the probability of getting a false positive), but also with respect to power, if the underlying assumptions are violated. We can only speculate about the extent to which this problem is present in SCI research, because a thorough examination of the question whether or not the assumptions underlying the respective tests were met would have exceeded the scope of the present review. However, it should be noted that in small datasets, it is virtually impossible to assess the underlying assumptions in a meaningful way. Apart from that, in cases where ordinal outcomes (eg, ASIA scores) are compared between groups by using

a parametric test, the underlying assumptions are clearly violated, regardless whether the sample is small or large. Even though the classical nonparametric alternatives such as the Wilcoxon-Mann-Whitney test are applicable in more general settings, they might show suboptimal performance in small sample sizes, too.<sup>56</sup> Likewise, we observed that chi-squared tests were used quite frequently for the analysis of categorical data, although it is well-known that they are based on an approximation and, hence, should only be applied to samples of reasonable size.<sup>57</sup>

Second, comparisons of paired observations (ie, repeated measures) were more frequently conducted than between-subject comparisons. In general, for fixed group sizes, a repeated measures design yields larger power than the comparison of unpaired groups, because the variance resulting from paired observations is smaller. Hence, even for small samples, the power might still be reasonably large. Nevertheless, consider, for example, a simple pre-post-design, where measurements are taken before and after a particular intervention of interest. Even if the effect size is moderate (ie, Cohen's  $d$ , which is the mean difference divided by the standard deviation, is equal to 0.5), the power is only about 56% for a sample size of 20 (details are provided in the R code in the supplementary material). With smaller sample sizes, the power decreases to 44% ( $n = 15$ ) and 29% ( $n = 10$ ). Nevertheless, at least the values, which have been extracted from the few publications where Cohen's  $d$  was provided indicate that for significant results, the corresponding median Cohen's  $d$  was equal to 1 (for details regarding effect sizes and variances, see Supplementary Material S1). Hence, as long as the group sizes do not drop below  $n = 10$  (power for  $n = 9$ : 74.8%), the power might be acceptable for paired tests. However, if we took the first quartile instead of the median, that is, if we set  $d = 0.8$ , the power would drop to 56.0% for a group size  $n = 9$ . Hence, reporting of means and standard deviations of the group differences definitely needs to be improved, in order to allow for a thorough analysis of effect sizes, power, and sample size calculation methods.

Third, in about two-thirds of the small-sample size studies, the authors were aware of the fact that the small sample size represented a limitation. Whereas in some cases, only one single, short sentence regarding this issue had been put down in the publication, other authors also discussed potential implications to some extent. Nevertheless, the impact of small sample sizes on power and post-study probabilities was not mentioned. Likewise, references to studies examining the appropriateness of certain statistical methods in particular small-sample size settings were not made.

We found clear signs that a considerable number of researchers are aware of the problems regarding small sample sizes, at least to some extent. So, obviously, there is a chance of improving the situation—but how? In our opinion, this goal can be achieved only in a collaborative effort of all scientists and stakeholders involved in SCI research. Clinical researchers could try to establish new connections with other research groups, in order to set the stage for conducting adequately powered multicenter trials.<sup>16</sup> Although such a trial might be demanding with respect to resources and does not come without methodological challenges (eg, standardization issues), there is no other way to

obtain reliable large-scale evidence. Clinical trial networks could serve as a convenient framework for such collaborative efforts. Apart from that, we propose that the editors of SCI journals consider the advice of expert statistics reviewers, and publish methodological papers on a more or less regular basis. A short tutorial-like paper, aimed at explaining and illustrating a certain methodological or statistical topic in a nontechnical style, could substantially facilitate understanding and increase awareness of potential problems related to small-sample settings. However, not only clinical researchers and journal editors could contribute to improving the status quo. Also, statisticians might have to change their attitudes towards publishing more in topic-oriented medical journals. Although a paper focused on theory is, of course, a great achievement, the merit of providing a nontechnical explanation of a certain methodological topic is equally outstanding, because this could lead to an increased reliability of research findings concerning treatment options for severely impaired people. Moreover, sessions that are focused on statistical methods should become an integral part of clinical research conferences, in order to further stimulate the interdisciplinary exchange of new methodological approaches. Apart from discussing advances regarding the statistical techniques, those interactions should also be focused on a shift towards more emphasis on clinical rather than statistical significance. Especially in people with SCI, the treatment groups, which are compared with respect to a certain outcome, might be very heterogeneous with respect to several patient characteristics (eg, incomplete vs complete, considerable variation in the years since injury). Therefore, identifying responders to interventions (ie, individuals that make clinically relevant improvements) on a case-by-case basis might sometimes be more sensible than comparing average responses between very small subgroups by means of a statistical hypothesis test.

The quality of the included studies was not taken into account in the analyses, because quality appraisal is more relevant for systematic reviews focused on health-related outcomes. Although there might be some associations between the choice of statistical methods and other aspects of a study's quality, a closer look at this issue lies beyond the scope of the present publication. Likewise, a risk of bias assessment, which is usually recommended for systematic reviews,<sup>25</sup> was not carried out, because within-study bias (eg, bias due to lack of randomization and unblinded outcome assessment) is irrelevant with respect to sample sizes and statistical methods. By contrast, bias across studies could indeed play a role. Although the relationship might not be perfect, it is generally assumed that small studies are more prone to publication bias.<sup>58</sup> Consequently, our results might still overestimate the "typical" sample size of a SCI trial. Likewise, publication bias and selective outcome reporting are also likely to have a considerable impact on the assessment of the statistical methods that have been used. In this case, however, we can only speculate about the direction of bias.

We have tried to minimize the risk of selection bias by carefully considering several methodological issues in advance. First, when doing the database search, we kept the application of filters to a minimum, in order to avoid bias due to algorithm-based pre-selection of studies. Second, we employed a quite conservative methodology

for the process of study selection, since we excluded studies at the abstract screening stage only if we could be really sure that they did not meet our inclusion criteria. Furthermore, each stage was supervised by clinically and statistically experienced researchers, who shared their opinion with the reviewers when unclear cases were encountered. Moreover, even the unambiguous results of the abstract screening underwent a rough overall check for errors by two reviewers. However, potential bias introduced at that stage cannot be entirely excluded, because it was decided due to the limited available resources that the abstract screening should be done by one single person (RS), only. Third, clinical trial registries were not sought. However, with respect to our research question, the information provided in registries would not have helped much anyway, since obviously, the final sample sizes cannot be obtained from a study protocol. Moreover, with regard to the statistical methods, there could be a considerable difference between what had been prespecified and what was actually carried out.

The present systematic review has some limitations. For feasibility reasons, we have decided to perform the literature search in one single database only. However, presumably, additional searches in databases like EMBASE would mainly increase the number of studies that were published as conference abstracts, etc and, therefore, lead to even smaller median sample sizes. Furthermore, we have not conducted any stratified analyses of the sample sizes (eg, stratified according to different types of interventions), because we wanted to provide a broad overall picture. Finally, the present systematic review is not supposed to serve as an overall assessment of the validity of findings from SCI studies, nor should it be regarded as an evaluation of the statistical quality in general. Rather, our work is focused on two particular goals, namely on determining the median sample size in SCI interventional studies and examining some key characteristics of the statistical methods.

To conclude, for the first time, we provide systematically collected evidence, which indicates that sample sizes in SCI interventional studies are small. Moreover, some awareness of potential problems related to small sample sizes seems to be present, yet without giving a more detailed account of implications for the choice of statistical methods in most cases. Further systematic assessments of selected methodological topics (eg, sample size calculation and power) are warranted.

## ACKNOWLEDGMENT

The authors would like to thank Mrs Roxane Dilcher for providing support in the planning phase of the study.

## DATA ACCESSIBILITY STATEMENT

In order to enhance transparency, we provide the text file, which contains the abstracts of the search results, upon request. Due to the considerable amount of studies, we think that this makes more sense than providing lengthy tables containing all individual study details.

## ORCID

Georg Zimmermann  <https://orcid.org/0000-0002-8282-1034>  
 Yvonne Höller  <https://orcid.org/0000-0002-1727-8557>  
 Arne C. Bathke  <https://orcid.org/0000-0002-6260-3726>  
 Francesco Brigo  <https://orcid.org/0000-0003-0928-1577>

## REFERENCES

1. Craggs MD, Balasubramaniam AV, Chung EAL, Emmanuel AV. Aberrant reflexes and function of the pelvic organs following spinal cord injury in men. *Auton Neurosci Basic Clin*. 2006;126–127:355–370.
2. Cruz CD, Cruz F. Spinal cord injury and bladder dysfunction: new ideas about an old problem. *Sci World J*. 2011;11:214–234.
3. DeVivo MJ, Krause JS, Lammertse DP. Recent trends in mortality and causes of death among persons with spinal cord injury. *Arch Phys Med Rehabil*. 1999;80(11):1411–1419.
4. Sabre L, Reikand T, Asser T, Korv J. Mortality and causes of death after traumatic spinal cord injury in Estonia. *J Spinal Cord Med*. 2013;36(6):687–694.
5. Noonan VK, Fingas M, Farry A, et al. Incidence and prevalence of spinal cord injury in Canada: a national perspective. *Neuroepidemiology*. 2012;38(4):219–226.
6. Blumer CE, Quine S. Prevalence of spinal cord injury: an international comparison. *Neuroepidemiology*. 1995;14(5):258–268.
7. Singh A, Tetreault L, Kalsi-Ryan S, Nouri A, Fehlings MG. Global prevalence and incidence of traumatic spinal cord injury. *Clin Epidemiol*. 2014;6:309–331.
8. Bracken MB. Steroids for acute spinal cord injury. *Cochrane database Syst Rev [Internet]*. 2012;1:CD001046. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22258943>
9. Versace V, Langthaler PB, Höller Y, et al. Abnormal cortical neuroplasticity induced by paired associative stimulation after traumatic spinal cord injury: a preliminary study. *Neurosci Lett*. 2018;664:167–171.
10. Alcobendas-Maestro M, Esclarín-Ruz A, Casado-López RM, et al. Lokomat robotic-assisted versus overground training within 3 to 6 months of incomplete spinal cord lesion: randomized controlled trial. *Neurorehabil Neural Repair*. 2012;26(9):1058–1063.
11. Yang JF, Musselman KE, Livingstone D, et al. Repetitive mass practice or focused precise practice for retraining walking after incomplete spinal cord injury? A pilot randomized clinical trial. *Neurorehabil Neural Repair*. 2014;28(4):314–324.
12. Nardone R, Höller Y, Brigo F, et al. Descending motor pathways and cortical physiology after spinal cord injury assessed by transcranial magnetic stimulation: a systematic review. *Brain Res*. 2015;1619:139–154.
13. Fehlings MG, Cadotte DW, Fehlings LN. A series of systematic reviews on the treatment of acute spinal cord injury: a foundation for best medical practice. *J Neurotrauma*. 2011;28(8):1329–1333.
14. Martin Ginis KA, van der Scheer JW, Latimer-Cheung AE, et al. Evidence-based scientific exercise guidelines for adults with spinal cord injury: an update and a new guideline. *Spinal Cord*. 2017;1–14.
15. Ginis KA, Hicks AL. Exercise research issues in the spinal cord injured population. *Exerc Sport Sci Rev*. 2005;33(1):49–53.
16. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):696–701.
17. Krishnamoorthy K, Lu F, Mathew T. A parametric bootstrap approach for ANOVA with unequal variances: fixed and random models. *Comput Stat Data Anal*. 2007;51(12):5731–5742.

18. Konietschke F, Bathke AC, Harrar SW, Pauly M. Parametric and non-parametric bootstrap methods for general MANOVA. *J Multivar Anal.* 2015;140:291–301.
19. Friedrich S, Brunner E, Pauly M. Permuting longitudinal data in spite of the dependencies. *J Multivar Anal.* 2017;153:255–265.
20. Brigo F, Trinka E. Randomized controlled trials in status epilepticus: size matters. *Epilepsia.* 2017;58(5):915.
21. Hsu LM. Random Sampling, Randomization, and equivalence of contrasted groups in psychotherapy outcome research. *J Consult Clin Psychol.* 1989;57(1):131–137.
22. Biondi-Zoccai G, Lotrionte M, Landoni G, Modena MG. The rough guide to systematic reviews and meta-analyses. *HSR Proc Intensive Care Cardiovasc Anesth.* 2011;3:161–173.
23. World Health Organization. Health topics: clinical trials [Internet]. 2018 [cited 2018 Jan 9]. Available from: [http://www.who.int/topics/clinical\\_trials/en/](http://www.who.int/topics/clinical_trials/en/). [accessed 30 June 2019]
24. Currie KD, Wong SC, Warburton DE, Krassioukov AV. Reliability of the sit-up test in individuals with spinal cord injury. *J Spinal Cord Med.* 2015;38(4):563–566.
25. Moher D, Liberati A, Altman DG, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 2009;6(7):e1000100.
26. Ottomanelli L, Barnett SD, Goetz LL. Effectiveness of supported employment for veterans with spinal cord injury: 2-year results. *Arch Phys Med Rehabil.* 2014;95(4):784–790.
27. Richards JS, Bombardier CH, Wilson CS, et al. Efficacy of venlafaxine XR for the treatment of pain in patients with spinal cord injury and major depression: a randomized, controlled trial. *Arch Phys Med Rehabil.* 2015;96(4):680–689.
28. Alam MS, Phan K, Karim R, et al. Surgery for spinal tuberculosis: a multi-center experience of 582 cases. *J Spine Surg.* 2015;1(1):65–71.
29. Gu X-D, Wang J, Yu P, et al. Effects of electroacupuncture combined with clean intermittent catheterization on urinary retention after spinal cord injury: a single blind randomized controlled clinical trial. *Int J Clin Exp Med.* 2015 Oct 15;8(10):19757–19763.
30. Derakhshanrad N, Vosoughi F, Yekaninejad MS, Moshayedi P, Saberi H. Functional impact of multidisciplinary outpatient program on patients with chronic complete spinal cord injury. *Spinal Cord.* 2015;53(12):901.
31. Froehlich-Grobe K, Lee J, Aaronson L, Nary DE, Washburn RA, Little TD. Exercise for everyone: a randomized controlled trial of project workout on wheels in promoting exercise among wheelchair users. *Arch Phys Med Rehabil.* 2014;95(1):20–28.
32. Wong S, Jamous A, O'Driscoll J, et al. A *Lactobacillus casei* Shirota probiotic drink reduces antibiotic-associated diarrhoea in patients with spinal cord injuries: a randomised controlled trial. *Br J Nutr.* 2014;111(4):672–678.
33. Cui H, Guo J, Yang L, Guo Y, Guo M. Comparison of therapeutic effects of anterior decompression and posterior decompression on thoracolumbar spine fracture complicated with spinal nerve injury. *Pakistan J Med Sci.* 2015;31(2):346–350.
34. Liu X, Meng Q, Yu D, Zhao X, Zhao T. Novel medical bathing with traditional Chinese herb formula alleviates paraplegia spasticity. *Int J Nurs Pract.* 2014;20(3):227–232.
35. Sabour H, Norouzi Javidan A, Latifi S, et al. Omega-3 fatty acids' effect on leptin and adiponectin concentrations in patients with spinal cord injury: a double-blinded randomized clinical trial. *J Spinal Cord Med.* 2015;38(5):599–606.
36. Conte A, Giannantoni A, Gubbiotti M, et al. Intradetrusorial botulinum toxin in patients with multiple sclerosis: a neurophysiological study. *Toxins (Basel).* 2015;7(9):3424–3435.
37. Milosevic M, Masani K, Kuipers MJ, et al. Trunk control impairment is responsible for postural instability during quiet sitting in individuals with cervical spinal cord injury. *Clin Biomech.* 2015;30(5):507–512.
38. Gorgey A, Khalil R. Neuromuscular electrical stimulation training increases intermuscular fascial length but not tendon cross-sectional area after spinal cord injury. *Top Spinal Cord Inj Rehabil.* 2015;21(1):87–92.
39. Alibai EA, Baghban F, Farrokhi MR, Mohebbi N, Ashraf MH. Effects of human erythropoietin on functional outcome of patients with traumatic cervical cord injury; a pilot randomized clinical trial. *Bull Emerg Trauma.* 2015 Jul;3(3):79–85.
40. Hayes HB, Jayaraman A, Herrmann M, Mitchell GS, Rymer WZ, Trumbower RD. Daily intermittent hypoxia enhances walking after chronic spinal cord injury: a randomized trial. *Neurology.* 2014;82(2):104–113.
41. Tamplin J, Baker FA, Buttifant M, Berlowitz DJ. The effect of singing training on voice quality for people with quadriplegia. *J Voice.* 2014;28(1):128.e19–128.e26.
42. Evans N, Hartigan C, Kandilakis C, Pharo E, Clesson I. Acute cardiorespiratory and metabolic responses during exoskeleton-assisted walking overground among persons with chronic spinal cord injury. *Top Spinal Cord Inj Rehabil.* 2015;21(2):122–132.
43. Tsang WWN, Gao KL, Chan KM, et al. Sitting Tai Chi improves the balance control and muscle strength of community-dwelling persons with spinal cord injuries: a pilot study. *Evidence-Based Complement Altern Med.* 2015;2015:1–9.
44. Arazpour M, Mehrpour SR, Bani MA, Hutchins SW, Bahramizadeh M, Rahgozar M. Comparison of gait between healthy participants and persons with spinal cord injury when using a powered gait orthosis—a pilot study. *Spinal Cord.* 2014;52(1):44–48.
45. Inada T, Takahashi H, Yamazaki M, et al. Multicenter prospective non-randomized controlled clinical trial to prove neurotherapeutic effects of granulocyte colony-stimulating factor for acute spinal cord injury: analyses of follow-up cases after at least 1 year. *Spine (Phila Pa 1976).* 2014;39(3):213–219.
46. El-Kheir WA, Gabr H, Awad MR, et al. Autologous bone marrow-derived cell therapy combined with physical therapy induces functional improvement in chronic spinal cord injury patients. *Cell Transplant.* 2014;23(6):729–745.
47. Oo WM. Efficacy of addition of Transcutaneous electrical nerve stimulation to standardized physical therapy in subacute spinal spasticity: a randomized controlled trial. *Arch Phys Med Rehabil.* 2014;95(11):2013–2020.
48. Arazpour M, Bani MA, Hutchins SW, Curran S, Javanshir MA, Mousavi ME. Influence of orthotic gait training with powered hip orthosis on walking in paraplegic patients. *Disabil Rehabil Assist Technol.* 2014;9(3):226–230.
49. Fleerkotte BM, Koopman B, Buurke JH, Van Asseldonk EHF, Van Der Kooij H, Rietman JS. The effect of impedance-controlled robotic gait training on walking ability and quality in individuals with chronic incomplete spinal cord injury: an explorative study. *J Neuroeng Rehabil.* 2014;11(1). <https://doi.org/10.1186/1743-0003-11-26>.
50. Leech KA, Kinnaird CR, Hornby TG. Effects of serotonergic medications on locomotor performance in humans with incomplete spinal cord injury. *J Neurotrauma.* 2014;31(15):1334–1342.
51. Sharif H, Gammage K, Chun S, Ditor D. Effects of FES-ambulation training on locomotor function and health-related quality of life in individuals with spinal cord injury. *Top Spinal Cord Inj Rehabil.* 2014;20(1):58–69.
52. Liu LQ, Ferguson-Pell M. Pressure changes under the ischial tuberosities during gluteal neuromuscular stimulation in spinal cord injury: a comparison of sacral nerve root stimulation with surface functional electrical stimulation. *Arch Phys Med Rehabil.* 2015;96(4):620–626.



53. Marques CG, Santos VC, Levada-Pires AC, et al. Effects of DHA-rich fish oil supplementation on the lipid profile, markers of muscle damage, and neutrophil function in wheelchair basketball athletes before and after acute exercise. *Appl Physiol Nutr Metab*. 2015;604(6): 596.
54. Guillon B, Van-Hecke G, Iddir J, et al. Evaluation of 3 pushrim-activated power-assisted wheelchairs in patients with spinal cord injury. *Arch Phys Med Rehabil*. 2015;96(5):894–904.
55. Van Straaten MG, Cloud BA, Morrow MM, Ludewig PM, Zhao KD. Effectiveness of home exercise on pain, function, and strength of manual wheelchair users with spinal cord injury: a high-dose shoulder program with telerehabilitation. *Arch Phys Med Rehabil*. 2014;95(10):1810–1817.
56. Brunner E, Munzel U. The nonparametric Behrens-Fisher problem: asymptotic theory and a small-sample approximation. *Biometrical J*. 2000;42(1):17–25.
57. Agresti A. *Categorical Data Analysis*. 3rd edition, Vol. 45. *Wiley Series in Probability and Statistics*. New York: Wiley; 2012.
58. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14(5):365–376.

#### SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Zimmermann G, Bolter L-M, Sluka R, et al. Sample sizes and statistical methods in interventional studies on individuals with spinal cord injury: A systematic review. *J Evid Based Med*. 2019;12:200–208. <https://doi.org/10.1111/jebm.12356>