
Ana Antonic1, Emily S. Sena2,3, Jennifer S. Lees3, Taryn E. Wills1,2, Peta Skeers1, Peter E. Batchelor1,2, Malcolm R. Macleod3, David W. Howells2*

1 Department of Medicine, University of Melbourne, Lance Townsend Building, Austin Hospital, Heidelberg, Victoria, Australia, 2 Florey Institute of Neuroscience and Mental Health, Victoria, Australia, 3 Division of Clinical Neurosciences, University of Edinburgh, Western General Hospital, Edinburgh, United Kingdom

Abstract

Spinal cord injury (SCI) is a devastating condition that causes substantial morbidity and mortality and for which no treatments are available. Stem cells offer some promise in the restoration of neurological function. We used systematic review, meta-analysis, and meta-regression to study the impact of stem cell biology and experimental design on motor and sensory outcomes following stem cell treatments in animal models of SCI. One hundred and fifty-six publications using 45 different stem cell preparations met our prespecified inclusion criteria. Only one publication used autologous stem cells. Overall, allogeneic stem cell treatment appears to improve both motor (effect size, 27.2%; 95% Confidence Interval [CI], 25.0%–29.4%; 312 comparisons in 5,628 animals) and sensory (effect size, 26.3%; 95% CI, 7.9%–44.7%; 23 comparisons in 473 animals) outcome. For sensory outcome, most heterogeneity between experiments was accounted for by facets of stem cell biology. Differentiation before implantation and intravenous route of delivery favoured better outcome. Stem cell implantation did not appear to improve sensory outcome in female animals and appeared to be enhanced by isoflurane anaesthesia. Biological plausibility was supported by the presence of a dose–response relationship. For motor outcome, facets of stem cell biology had little detectable effect. Instead most heterogeneity could be explained by the experimental modelling and the outcome measure used. The location of injury, method of injury induction, and presence of immunosuppression all had an impact. Reporting of measures to reduce bias was higher than has been seen in other neuroscience domains but were still suboptimal. Motor outcomes studies that did not report the blinded assessment of outcome gave inflated estimates of efficacy. Extensive recent preclinical literature suggests that stem-cell–based therapies may offer promise, however the impact of compromised internal validity and publication bias mean that efficacy is likely to be somewhat lower than reported here.


Academic Editor: Doug G. Altman, Oxford University, United Kingdom

Received December 4, 2012; Accepted November 1, 2013; Published December 17, 2013

Copyright: © 2013 Antonic et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The Florey Institute of Neuroscience and Mental Health acknowledges the strong support from the Victorian Government and in particular the funding from the Operational Infrastructure Support Grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: AD, Alzheimer’s disease; BBB, Basso Beattie and Bresnahan; BMS, Basso mouse scale; CAMARADES, Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies; CI, Confidence interval; CNS, Central nervous system; EAE, Experimental autoimmune encephalomyelitis; FCI, Focal cerebral ischaemia; SCI, Spinal cord injury; PD, Parkinson’s disease; χ², Chi-squared, heterogeneity.

* E-mail: david.howells@unimelb.edu.au

† These authors contributed equally to this work.

Introduction

Stem cells, from which all tissues can be generated, offer the potential to reconstitute tissues damaged by injury and disease. However, realising this potential will demand a detailed knowledge of the genetic and internal environmental cues that specify a cell’s type, location, and interaction with its neighbours. It will also require a thorough understanding of stem cell behaviour in the context of lesioned or damaged tissues.

Stem cell transplantation was pioneered in the 1950s using hematopoietic stem cells to repopulate the bone marrow in patients with cancers of the blood and bone marrow [1]. Such is the success of this approach that an estimated 50,000 of these transplants are performed each year [2]. As understanding of stem cell biology has increased, so too has the ambition for restoring more complex tissues. In animal models, hepatocytes derived from stem cells can be engrafted into the damaged liver [3], and lineage-specific stem cells can repair damaged cornea [4,5]. Recent studies also demonstrate the generation of artificial tissues with key features of complex solid organs including blood vessels [6], heart [7–9], lung [10], and kidney [11]. Even in the CNS, where the breadth of cell types and the complexity of their interactions are maximal, stem cell implants appear able to integrate into the existing circuitry [12–14]. In patients, lineage-specific stem cells have been reported to show efficacy in the regeneration of craniofacial bones [13] and of damaged cornea [5].

Integration into the host environment and tissue reconstruction are not the only potentially relevant biological effects of stem cells.
Results

Study Characteristics

Electronic searching identified 156 full publications that met our prespecified inclusion criteria (Table S1). Forty-five different stem cell types had been investigated, from which over a third were derived from adult rats. The duration of experiments following the induction of SCI ranged from 7 d to 6 mo.

Risk of Bias

We describe the reporting of study quality checklist items reported for each included publication in Table S2. All studies included in this analysis came from peer-reviewed papers; while we identified a number of potentially relevant abstracts, none of these reported data in sufficient detail to be included. One hundred and eleven of 156 publications (71%) reported compliance with animal welfare regulations, and 25 (16%) reported whether or not a conflict of interest existed.

Allocation concealment was reported in 14 of 156 publications (9%). Random allocation to treatment group (72, 46%) and blinded assessment of outcome (72, 46%) were reported more frequently in these publications than in the modelling of other neurological disorders [39–42], but the reporting of a sample size calculation (less than 1%) was consistent with the proportions observed elsewhere (Table 1). No publication reported all four of these measures to minimise bias.

Despite the reported benefits of hypothermia in SCI [43–45], in other animal models of neurological disease [46] and in humans with ischaemic neurological injury [47,48], only 33 (21%) studies described controlling temperature during the experiments.

There were only sufficient data to assess publication bias in studies using allogeneic stem cells where outcome was measured as a motor score. Small study bias was suggested with symmetry of the funnel plot (Figure 2A) and Egger regression (Figure 2B) but not by Trim and Fill.

Meta-Analysis

As expected, our search identified a diverse range of experiments. There was substantial between-study heterogeneity for studies using allogeneic stem cells both where outcome was measured as a motor score [heterogeneity (\( \chi^2 \)) = 9,735, 311 degrees of freedom (df), \( p \leq 10^{-99} \), effect size, 27.2% improvement in outcome [95% confidence interval, 25.0%–29.4%]; 312 comparisons] and as a sensory outcome (\( \chi^2 = 183, df = 23, p < 10^{-26} \), effect size, 26.3% [7.9%–44.7%]; 24 comparisons).

Motor score in experiments using allogeneic stem cells. In meta-regression, eight study characteristics accounted
Figure 1. Summary of data included in meta-analysis of use of stem cells to treat spinal cord injury with individual comparisons ranked according to their effect on (A) % improvement in motor score and (B) % improvement in sensory score. The shaded grey bar represents the 95% confidence limits of the global estimate. The vertical error bars represent the 95% confidence intervals for the individual estimates.

doi:10.1371/journal.pbio.1001738.g001

Table 1. Reporting of study quality criteria.

<table>
<thead>
<tr>
<th>Item</th>
<th>SCI</th>
<th>FCI [41]</th>
<th>EAE [42]</th>
<th>PD [40]</th>
<th>AD [39]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random allocation to group</td>
<td>46%</td>
<td>36%</td>
<td>9%</td>
<td>16%</td>
<td>15%</td>
</tr>
<tr>
<td>Blinded assessment of outcome</td>
<td>46%</td>
<td>29%</td>
<td>16%</td>
<td>15%</td>
<td>21%</td>
</tr>
<tr>
<td>Sample size calculation</td>
<td>1%</td>
<td>3%</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>0%</td>
</tr>
<tr>
<td>Compliance with animal welfare regulations</td>
<td>71%</td>
<td>57%</td>
<td>32%</td>
<td>40%</td>
<td>54%</td>
</tr>
<tr>
<td>Statement of a potential conflict of interest</td>
<td>16%</td>
<td>23%</td>
<td>6%</td>
<td>2%</td>
<td>11%</td>
</tr>
</tbody>
</table>

Abbreviations: AD, Alzheimer’s disease; EAE, experimental autoimmune encephalomyelitis; FCI, focal cerebral ischaemia; PD, Parkinson's disease; SCI, spinal cord injury.

doi:10.1371/journal.pbio.1001738.t001
for a significant proportion of the between-study heterogeneity in studies reporting a change in motor score (Table 2). More influence was apparent for factors related to the lesion model than those related to stem cell biology. There was no detectable effect of stem cell dose, derivation (adult or embryonic), manipulation in culture (genetic, growth factor, antibiotic), number of passages in culture, method of stem cell selection prior to implantation, route of administration, frequency of administration, the presence or absence of a supporting scaffold, time of assessment, anaesthetic used, or temperature regulation during surgery.

The neurobehavioural test used (Figure 3A) accounted for most of the observed heterogeneity (adjusted $R^2 = 12.2\%$, $p<0.00001$). Seventy percent of the data (228 comparisons, 4,042 animals) was used accounted for a high proportion of the heterogeneity (adjusted $R^2 = 5.8\%$, $p<0.005$). For cyclosporine A [78 comparisons, 1,242 (22% of total) animals], efficacy was 19.6% ($13.7–25.4$) compared with 30.2% ($27.2–33.1$) in 226 comparisons and 4,259 animals where no immunosuppression was used. Efficacy also appeared smaller in a smaller number of experiments [6 comparisons, 80 (1.4%) animals] using FK506 (Figure 3D).

The approach used to induce injury had a smaller but significant effect (adjusted $R^2 = 3.4\%$, $p<0.01$, Figure 3E). The most common approach was contusion injuries [149 comparisons, 2,847 animals; 23.8% improvement, ($20.1–27.5$)] with compressive approaches providing improvements of a similar magnitude [59 comparisons, 1,135 animals; 25.8% ($18.8–32.8$)]. Slightly higher estimates of effect size were obtained when the cord had been transected [63 comparisons, 928 animals; 30.5% ($24.1–37.0$)] or hemisected [38 comparisons, 717 animals; 37.6% ($29.1–46.2$)].

Efficacy was highest with treatment strategies using cell lines (7 comparisons, 131 animals) rather than primary cells, and amongst primary cells those derived from mice were the least effective (Figure 3F, adjusted $R^2 = 4.3\%$, $p<0.005$).

Efficacy was lower in studies reporting the blinded assessment of outcome [147 comparisons, 2,653 animals, 23.6% ($18.5–28.7$)] than in those that did not [165 experiments, 2,975 animals, 30.3% ($23.1–50.1$)]. No effect was seen for reporting of allocation concealment, randomisation, or sample size calculations.

**Motor score subanalyses.** A large proportion of the data (115 comparisons, 2,165 animals) were obtained from rats implanted with allogeneic stem cells, after injury created with an impactor, at the midthoracic level and assessed by the BBB test, where the sex of the animal was explicitly stated. This large and experimentally homogeneous subset of the data was analysed separately to establish whether a clearer picture of the key determinants of stem cell biology and implantation emerged.

Heterogeneity was reduced from 9.735 ($\chi^2$) over 312 individual comparisons to 1.420 over 115 comparisons, confirming the validity of this approach. As in the full analysis, stem cell dose, number of passages during culture, the presence of additional antibiotics or growth factors in the culture medium, selection methodology, the use of adult or embryonic stem cells and the species of origin, route of administration, presence of a supporting scaffold, and prior differentiation or transfection of the stem cells had no significant effect.

In this subpopulation of comparisons (Table 3) the anaesthetic used accounted for a high proportion of the heterogeneity (adjusted $R^2 = 16.3\%$, $p<0.001$). Isoflurane was infrequently used (3 comparisons, 47 animals) and was associated with the largest improvement in outcome. Of the most commonly used anaesthetics, chloral hydrate [21 comparisons, 417 animals, 33.0% ($16.0–50.1$)] was associated with the largest effect size (Figure 4A).
The interval from lesioning to outcome assessment accounted for 11.0% of the heterogeneity such that absolute effect size fell by 1.7% for every additional week of delay to outcome assessment. The presence of immunosuppression also accounted for a large proportion of the heterogeneity in this constrained dataset (adjusted $R^2 = 10.4\%$, $p<0.01$); both cyclosporine A and FK506 substantially reduced the benefit derived from stem cells (Figure 4B). BBB scores were lower in experiments where other tests had also been reported [22 comparisons, 473 animals, 14.0% (4.7–23.3)] than where BBB was reported alone [93 comparisons, 26.7 (23.9–29.4)] (Table 2).

**Table 2. Study characteristics accounting for heterogeneity of motor score.**

<table>
<thead>
<tr>
<th>Motor Score</th>
<th>Effect size % (95% CI)</th>
<th>Number of Animals</th>
<th>Number of Comparisons</th>
<th>Adjusted $R^2$</th>
<th>$p&lt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled estimate</td>
<td>27.2 (25.0–29.4)</td>
<td>5,628</td>
<td>312</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NBS Motor tests</td>
<td>–8.1 (–37.7–21.4)</td>
<td>49</td>
<td>3</td>
<td>12.24%</td>
<td>0.00001</td>
</tr>
<tr>
<td>Staircase</td>
<td>–2.0 (–43.3–39.3)</td>
<td>12</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMS</td>
<td>24.5 (11.2–37.7)</td>
<td>196</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple tests</td>
<td>24.5 (17.8–31.2)</td>
<td>1,053</td>
<td>56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BBB</td>
<td>26.7 (23.9–29.4)</td>
<td>4,042</td>
<td>228</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forelimb placing test</td>
<td>47.9 (18.8–77.1)</td>
<td>76</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tarlov</td>
<td>73.1 (57.5–88.7)</td>
<td>200</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location of injury</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cervical</td>
<td>32.2 (12.2–52.3)</td>
<td>156</td>
<td>13</td>
<td>10.64%</td>
<td>0.00001</td>
</tr>
<tr>
<td>Lower thoracic/lumbar</td>
<td>48.1 (39.7–55.5)</td>
<td>456</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midthoracic</td>
<td>24.9 (22.3–27.5)</td>
<td>5,016</td>
<td>271</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>22.9 (19.6–26.3)</td>
<td>2,906</td>
<td>171</td>
<td>9.69%</td>
<td>0.00001</td>
</tr>
<tr>
<td>Male</td>
<td>27.4 (21.7–33.1)</td>
<td>1,704</td>
<td>87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>35.7 (27.9–43.5)</td>
<td>676</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both</td>
<td>48.7 (37.6–59.7)</td>
<td>341</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immunosuppression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclosporine A/MP</td>
<td>–11.5 (–92.5–69.5)</td>
<td>12</td>
<td>1</td>
<td>5.83%</td>
<td>0.0026</td>
</tr>
<tr>
<td>FK506</td>
<td>11.6 (–8.0–31.2)</td>
<td>80</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclosporine A</td>
<td>19.6 (13.7–25.4)</td>
<td>1,242</td>
<td>78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>30.2 (27.2–33.1)</td>
<td>4,259</td>
<td>226</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclophosphamide</td>
<td>44.4 (–0.8–89.7)</td>
<td>36</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method used to induce SCI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impactor with spacer</td>
<td>11.6 (–11.3–34.5)</td>
<td>79</td>
<td>5</td>
<td>4.40%</td>
<td>0.0115</td>
</tr>
<tr>
<td>Aneurysm clip</td>
<td>18.7 (7.9–29.4)</td>
<td>356</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impactor</td>
<td>24.1 (20.4–27.8)</td>
<td>2,768</td>
<td>144</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>27.8 (19.7–36)</td>
<td>665</td>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balloon compression</td>
<td>28.4 (15.8–41.1)</td>
<td>235</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compression weight</td>
<td>30.1 (20.3–39.9)</td>
<td>544</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blade</td>
<td>33.5 (26.4–40.6)</td>
<td>682</td>
<td>52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scissors</td>
<td>42.5 (30–54.9)</td>
<td>278</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filament</td>
<td>79.2 (4.8–53.6)</td>
<td>20</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source of cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell line</td>
<td>41.1 (25.1–57.1)</td>
<td>131</td>
<td>7</td>
<td>4.34%</td>
<td>0.0034</td>
</tr>
<tr>
<td>Human</td>
<td>28.0 (21.6–34.3)</td>
<td>1,483</td>
<td>77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>18.0 (11.2–24.8)</td>
<td>877</td>
<td>56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>29.2 (25.9–32.6)</td>
<td>3,136</td>
<td>172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of injury</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contusion</td>
<td>23.8 (20.1–27.5)</td>
<td>2,847</td>
<td>149</td>
<td>3.44%</td>
<td>0.0073</td>
</tr>
<tr>
<td>Compression</td>
<td>25.8 (18.8–32.8)</td>
<td>1,135</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transection</td>
<td>30.5 (24.1–37)</td>
<td>928</td>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemisection</td>
<td>37.7 (29.1–46.2)</td>
<td>717</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinded assessment of outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not blinded</td>
<td>30.3 (26.8–33.8)</td>
<td>2,975</td>
<td>165</td>
<td>2.21%</td>
<td>0.01</td>
</tr>
<tr>
<td>Blinded</td>
<td>23.6 (18.5–28.7)</td>
<td>2,653</td>
<td>147</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The interval from lesioning to outcome assessment accounted for 11.0% of the heterogeneity such that absolute effect size fell by 1.7% for every additional week of delay to outcome assessment. The presence of immunosuppression also accounted for a large proportion of the heterogeneity in this constrained dataset (adjusted $R^2 = 10.4\%$, $p<0.01$); both cyclosporine A and FK506 substantially reduced the benefit derived from stem cells (Figure 4B). BBB scores were lower in experiments where other tests had also been reported [22 comparisons, 473 animals, 14.0% (4.7–23.3)] than where BBB was reported alone [93 comparisons,
There was no impact of whether stem cells were given once, at multiple times, or by continuous infusion; the sex of the animals; or the reporting of randomisation, allocation concealment, or blinded assessment of outcome.

A second subanalysis of the motor dataset was performed to examine whether restriction of the analysis to higher quality studies appreciably altered the results. This analysis was hampered by the paucity of truly high-quality data. None of the contributing papers reported each of four key measures of internal validity (randomisation, blinded assessment of outcome, allocation concealment, and sample size calculation), and only 20 individual comparisons came from papers describing three of the four. As a compromise we analysed the 25% of the

Figure 3. Study characteristics which account for heterogeneity of total motor dataset. (A) Behavioural test used, (B) location of injury, (C) sex of animals, (D) immunosuppressant used, (E) type of Injury, (F) stem cell source, and (G) effect of blinding. The shaded grey bar represents the 95% confidence limits of the global estimate. The vertical error bars represent the 95% confidence intervals for the individual estimates.
doi:10.1371/journal.pbio.1001738.g003
motor dataset that reported having both randomisation and blinding.

Restricting the analysis in this way reduced the number of animals assessed from 5,628 to 1,466 and heterogeneity fell from 9,735 to 945 ($\chi^2$). Despite this, the key features of both the full and the subanalysis are the same. The characteristics of the animal model still have more impact than the type of cells implanted (Tables 2 and 4).

Immunosuppression no longer has an effect on heterogeneity and the effect size in animals immunosuppressed with cyclosporine-A [mean, 24.3; 95% CI, 13.2–35.3] is the same as in animals where immune suppression is not used [mean, 24.9; 95% CI, 18.3–31.6]. Allocation concealment emerges as significant, though not in the expected direction. Also the type of cell culture medium and type of cell manipulation prior to implantation also begin to have an impact, but it should be noted that in both cases it is the experiments where the precise conditions are “unknown” that report the greatest effect. In the subanalysis, the mean number of cells implanted is substantially lower than in the full analysis (6.3 × 10^3 versus 7.4 × 10^6), and a dose–response relationship is evident.

**Sensory score in experiments using allogeneic stem cells.** While motor behaviour was relatively unaffected by most factors specific to stem cell biology, the reverse was true for studies reporting a change in sensory outcome (Table 5).

Of the five study characteristics accounting for a significant proportion of the between-study heterogeneity, the type of manipulation in culture had the largest effect (adjusted $R^2 = 61.3\%$, $p<0.005$). Prior differentiation was associated with larger effect sizes, while transfection was associated with smaller effects (Figure 5A). The number of cells administered had a clear dose–response effect (adjusted $R^2 = 31.7\%$, $p<0.02$; Figure 5B). Studies that delivered cells intravenously were associated with significantly larger effects than studies transplanting the cells directly into the lesion area of the spinal cord (adjusted $R^2 = 19.2\%$, $p<0.05$) (Figure 5C).

As with the motor score subanalysis, the anaesthetic agent had a large effect (adjusted $R^2 = 42.8\%$, $p<0.05$). The use of isoflurane to induce anaesthesia in three individual comparisons was associated with substantial additional benefit compared to other methods of anaesthesia (Figure 5D). All studies assessed sensory outcome in either all male or all female cohorts, with studies using female animals appearing to offer no benefit (Figure 5E; adjusted $R^2 = 21.5\%$, $p<0.05$).

**Discussion**

Systematic review and meta-analysis have helped identify biases within clinical trials [49], providing an impetus to improve standards [50]. This approach offers similar benefits for animal studies [28,41,51] by describing the impact of biological and experimental factors on reported efficacy in a systematic and transparent summary of all available data. This allows judgement of the extent to which conclusions are at risk of bias [52]. In this study we apply these techniques to provide a detailed systematic analysis of the animal literature describing stem-cell–based therapies in spinal cord injury.

Overall, treatment with allogeneic stem cells improves both motor and sensory outcome after spinal cord injury by around 25%, but with important differences between the two datasets. Because of the amount of data, conclusions relating to motor outcome (5,628 animals) are probably more robust than those relating to sensory outcomes (473 animals). For both outcomes there was a broad range of experimental approaches, reflected in the high levels of heterogeneity seen. This is typical for systematic reviews in animal studies and validates our choice of a random effects model, and our summary estimates should be considered as the average efficacy rather than the best estimate of a single “true” efficacy. Interestingly, improvement in sensory outcome seems to be sensitive to differences in factors relating to treatment (i.e., stem cell biology), while motor outcome appears to be more sensitive to factors relating to the lesion and the outcome measure used, and to be less dependent on the biological features of the stem cells used.

Evidence supporting a dose–response relationship for sensory outcome suggests the presence of a biologically plausible effect. We

TABLE 3. Study characteristics accounting for heterogeneity of motor score subanalysis.

<table>
<thead>
<tr>
<th>Motor Score</th>
<th>Effect Size % (95% CI)</th>
<th>Number of Animals</th>
<th>Number of Comparisons</th>
<th>Adjusted $R^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled estimate</td>
<td>24.1 (20.1–28.1)</td>
<td>2,165</td>
<td>115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaesthetic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halothane</td>
<td>15.5 (0.3–30.8)</td>
<td>147</td>
<td>6</td>
<td>16.3%</td>
<td>0.0007</td>
</tr>
<tr>
<td>Ketamine</td>
<td>17.4 (0.6–34.2)</td>
<td>508</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pentobarb</td>
<td>17.5 (1.1–33.9)</td>
<td>740</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>28.8 (10.8–46.8)</td>
<td>265</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chloral Hydrate</td>
<td>33.0 (16.0–50.1)</td>
<td>417</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isoflurane</td>
<td>59.2 (31.8–86.7)</td>
<td>47</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−1.7 (−2.8 to −0.6) for each 1 week delay in assessment</td>
<td></td>
<td></td>
<td></td>
<td>11.0%</td>
<td>0.002</td>
</tr>
<tr>
<td>Immunosuppression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FK506</td>
<td>11.8 (−6.0−29.5)</td>
<td>80</td>
<td>6</td>
<td>10.42%</td>
<td>0.0064</td>
</tr>
<tr>
<td>Cyclosporine A</td>
<td>16.5 (9.0–24.0)</td>
<td>675</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>27.8 (23.1–32.4)</td>
<td>1,410</td>
<td>69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NBS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple</td>
<td>14.0 (4.7–23.3)</td>
<td>473</td>
<td>22</td>
<td>5.0%</td>
<td>0.02</td>
</tr>
<tr>
<td>BBB</td>
<td>25.1 (21.0–29.1)</td>
<td>1,692</td>
<td>93</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pbio.1001738.t003
observed that prior differentiation of the implanted cells was associated with larger effects. Where the influence of cell differentiation was formally studied, a relationship with outcome was observed [53]. This suggests that optimal efficacy might be seen when cells have some lineage specificity but before final cell type commitment has occurred. For sensory outcome, studies where cells were delivered intravenously, rather than directly into the injured spinal cord, were associated with significantly larger effects. This suggests either that systemic changes may mediate the effects of stem cells or that local implantation may create additional injury that masks the benefit provided by stem cells.

We did not see a dose–response relationship for motor outcomes, even where we limited our analysis to a more homogenous subset of experiments. It may be that there is no dose–response effect or that the doses used in these experiments were all large enough to generate maximal responses. Where dose response was formally studied the authors found increasing benefit from doses as low as 10,000 implanted cells [54], and the median number of implanted cells in comparisons reporting motor outcomes was 250,000.

Immunosuppression with cyclosporine A was associated with increased efficacy in a systematic review of stem cells in focal cerebral ischaemia [28], and it is therefore interesting that in spinal cord injury both cyclosporine A and FK506 are associated with reduced efficacy. This suggests that any beneficial effect of immunosuppressants in promoting the survival of transplanted cells is outweighed by other factors, such as effects on stem cell biology or intrinsic repair mechanisms. Unfortunately, because of the univariate nature of our analyses we are unable to determine a “benefit–risk ratio” for the use of immunosuppression. However, there are studies that indicate that bone-marrow–derived stem cells are able to produce compartmentalised inflammatory lesions [55,56]. The mechanisms behind this observation are not understood, yet there are rising concerns that unwanted

Figure 4. Study characteristics that account for heterogeneity of motor data subanalysis when only data from rats implanted with allogeneic stem cells after injury created with an impactor at the midthoracic level and assessed by BBB. (A) Anaesthetic used, (B) immunosuppressant used, and (C) influence of additional behavioural testing on BBB. The shaded grey bar represents the 95% confidence limits of the global estimate. The vertical error bars represent the 95% confidence intervals for the individual estimates.
doi:10.1371/journal.pbio.1001738.g004
inflammatory-driven side effects, such as neuropathic pain, might limit the “usefulness” of gained motor function.

For motor outcome, the neurobehavioural test used (Figure 3A) accounted for most of the observed heterogeneity. The BBB locomotor rating scale was used in 70% of animals. In the more focussed analysis of rat allogeneic, midthoracic impact injury, using BBB as an outcome, studies that used other behavioural tests in addition to the BBB reported smaller effect sizes for the BBB. This may be a manifestation of outcome reporting bias; if the outcome on the BBB is smaller than expected, investigators might also report the outcome on other tests where the effect was larger; if the effect measured using the BBB was considered “sufficient,” there might be less motivation also to report outcomes using other measures, particularly if these were smaller than seen using the BBB.

Overall, there was no improvement in motor outcome where this was assessed using the staircase or mouse hind limb motor score tests. However, these accounted for a small proportion of the

<table>
<thead>
<tr>
<th>Table 4. Study characteristics accounting for heterogeneity of motor score—Randomised and blinded subset.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor Score</td>
</tr>
<tr>
<td>Pooled estimate</td>
</tr>
<tr>
<td>Method used to induce SCI</td>
</tr>
<tr>
<td>Aneurysm clip</td>
</tr>
<tr>
<td>Balloon compression</td>
</tr>
<tr>
<td>Compression, weight</td>
</tr>
<tr>
<td>Iridesectomy scissors</td>
</tr>
<tr>
<td>Impactor</td>
</tr>
<tr>
<td>Knife</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
<tr>
<td>Type of injury</td>
</tr>
<tr>
<td>Compression</td>
</tr>
<tr>
<td>Contusion</td>
</tr>
<tr>
<td>Hemisectomy</td>
</tr>
<tr>
<td>Transection</td>
</tr>
<tr>
<td>Location of injury</td>
</tr>
<tr>
<td>LowerThoracic/Lumbar</td>
</tr>
<tr>
<td>Midthoracic</td>
</tr>
<tr>
<td>NBS</td>
</tr>
<tr>
<td>Multiple tests</td>
</tr>
<tr>
<td>Tarlov</td>
</tr>
<tr>
<td>Dose</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>Both</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
<tr>
<td>Allocation concealment</td>
</tr>
<tr>
<td>Not concealed</td>
</tr>
<tr>
<td>Cell culture medium</td>
</tr>
<tr>
<td>Growth Factor</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
<tr>
<td>Cell manipulations</td>
</tr>
<tr>
<td>Diff.+Transfection</td>
</tr>
<tr>
<td>None</td>
</tr>
<tr>
<td>Transfection</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pbio.1001738.t004
Efficacy was strongly associated with both the location of and the methodology used to create the injury. The largest effect was seen with lower thoracic and lumbar lesions and when the spinal cord was lesioned by hemisection or transsection rather than contusion or compression.

The use of isoflurane anaesthesia at SCI induction was associated with substantial improvement in sensory outcome; in the overall motor analysis, there was no effect, but in the more homogenous restricted analysis, isoflurane was again associated with substantially larger effects. Again, this contrasts with findings in focal cerebral ischaemia and suggests that, despite interest in a general paradigm of “neuroprotection,” these conditions are in certain respects biologically very different. However, these findings are based on a small number of individual comparisons and should be interpreted with caution.

The sex of the experimental animal accounted for a large proportion of the observed heterogeneity in both the sensory and motor analyses. For the motor analyses, this seems to be the influence of abnormally high effect sizes reported in studies where either the sex of the animals used was not reported or where “both sexes” were used. For sensory outcome, studies using male animals led to significantly higher estimates of effect with no clear benefit seen with lower thoracic and lumbar lesions and when the spinal cord was lesioned by hemisection or transsection rather than contusion or compression.

<table>
<thead>
<tr>
<th>Sensory Outcome</th>
<th>Effect Size % (95% CI)</th>
<th>Number of Animals</th>
<th>Number of Comparisons</th>
<th>Adjusted R²</th>
<th>p &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled estimate</td>
<td>26.3 (7.9–44.7)</td>
<td>473</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell manipulation</td>
<td>Differentiation</td>
<td>79.9 (34.8–125)</td>
<td>156</td>
<td>3</td>
<td>61.27%</td>
</tr>
<tr>
<td></td>
<td>Transfection</td>
<td>12.1 (–23.1–47.3)</td>
<td>131</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>42.5 (12.7–72.3)</td>
<td>156</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>0.0 (–43.5–43.5)</td>
<td>30</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Anaesthetic</td>
<td>Isoflurane</td>
<td>81.0 (37.0–125.1)</td>
<td>161</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ketamine</td>
<td>12.9 (–11.3–37.1)</td>
<td>131</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pentobarbital</td>
<td>17.5 (–16.7–51.8)</td>
<td>104</td>
<td>5</td>
<td>42.79%</td>
</tr>
<tr>
<td></td>
<td>Halothane</td>
<td>−15.3 (–318.1–287.5)</td>
<td>24</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>17.7 (–53.7–89)</td>
<td>53</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Dose</td>
<td>27.8 (5.6–50.0)</td>
<td>31.72%</td>
<td>0.017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>39.7 (3.9–75.4)</td>
<td>347</td>
<td>11</td>
<td>21.48%</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>−0.3 (–29.5–28.9)</td>
<td>126</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Route of delivery</td>
<td>Intraspinal cord</td>
<td>20.4 (2.5–38.4)</td>
<td>428</td>
<td>21</td>
<td>19.25%</td>
</tr>
<tr>
<td></td>
<td>Intravenous</td>
<td>77.2 (21.6–132.8)</td>
<td>45</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Overall, our study suggests that the effects of stem cell therapy in SCI are more likely to be observed in animal studies than in clinical trials. The results of our study also imply that the predictability of stem cell-based therapy may be greater than that of other types of therapy, as the variability in outcomes is lower. However, the results of our study should be interpreted with caution, as the study was limited by the small number of studies included and the potential for bias in the selection of studies. Further research is needed to validate these findings and to determine the optimal conditions for the use of stem cell therapy in the treatment of SCI.
Figure 5. Study characteristics that account for heterogeneity in sensory score. (A) Type of manipulation of stem cells prior to implantation, (B) dose–response relationship, (C) route of stem cell delivery, (D) anaesthetic used, and (E) sex. The shaded grey bar represents the 95% confidence limits of the global estimate. The vertical error bars represent the 95% confidence intervals for the individual estimates. doi:10.1371/journal.pbio.1001738.g005
precise conditions are “unknown” that report the greatest effect. There is no obvious biological explanation for this. It may be that a failure to report such details is a surrogate indication that such work is generally of lower quality, and therefore at greater risk of bias.

Immunosuppression is no longer identified as accounting for a significant proportion of the heterogeneity. However, the effect size in cyclosporine-A-treated animals (mean, 24.3; 95% CI, 13.2–35.3) is the same as in animals where no immune suppression was used (mean, 24.9; 95% CI, 18.3–31.6). This appears to confirm that immune suppression offers no advantage in experiments using allogeneic implants to treat SCI.

Intriguingly, in the subanalysis a dose–response relationship does emerge. As the mean number of cells implanted is \(6.3 \times 10^6\) rather than \(7.4 \times 10^6\) in the full motor dataset, this is consistent with the hypothesis that such an effect was previously masked by a ceiling effect.

**Limitations of our approach.** Firstly, we were only able to include data from studies in the public domain and—for motor outcome at least—there is evidence of a publication bias in favour of studies with large effect sizes. Further, we found some evidence (in the motor BBB subanalysis) consistent with selective reporting of outcomes within individual publications. The true effect sizes are therefore likely to be lower than reported here. Secondly, for both study quality and study design features, we relied on published information. Where relevant information was not available (the sex of a cohort of animals, or the taking of measures to reduce bias), we have either analysed these as not known or inferred that things that were not reported did not occur. Thirdly, we present a series of univariate analyses; multivariate meta-regression or stepwise partitioning of heterogeneity might provide more robust insights, but these techniques are not well established. Similarly, for continuous variables, the meta-regressions reported here assumed a linear relationship between the independent and dependent variables, and this is likely that this represents an oversimplification, at least for some independent variables. Fourthly, we have observed the experiments of others rather than conducted experiments of our own, and this observational research should be considered as hypothesis generating only. Finally, we limited our analysis to neurobehavioural outcomes; the true effect size in cyclosporine-A–treated animals (mean, 24.3; 95% CI, 13.2–35.3) is the same as in animals where no immune suppression was used (mean, 24.9; 95% CI, 18.3–31.6). This appears to confirm that immune suppression offers no advantage in experiments using allogeneic implants to treat SCI.

In conclusion, stem cells appear to have substantial efficacy in SCI. Effects on sensory outcome appear more dependent on facets of stem cell biology; motor outcome appears to be more dependent on features of the animal modelling and the outcome scale used.

**Methods**

The study protocol is available at www.camarades.info/index_files/Protocols.html. A completed PRISMA checklist and flow diagram for this systematic literature review can be found in Text S1.

**Definitions**

We define a “publication” as a discrete piece of work (including abstracts); each publication may report data from a number of experiments. Each experiment may describe outcome in a number of different experimental cohorts, and the contrast between outcomes in a single treatment cohort with that in a control cohort we define as an “individual comparison.” We define “nesting” as combining the effect sizes from different functional outcomes measured in the same cohort of animals to give a single summary estimate of effect in that individual comparison (a nested individual comparison).

**Systematic Review**

Using prespecified inclusion and exclusion criteria we identified all publications reporting relevant experiments (see below) by searching (December 2011) three electronic databases (PubMed, EMBASE, and ISI Web of Science) using the search strategy “(stem cell OR stem OR haematopoietic OR mesenchymal) AND (spinal cord injury OR hemisection OR contusion injury OR dorsal column injury OR complete transaction OR corticospinal tract injury),” with search results limited to those indexed as describing animal experiments.

**Inclusion and Exclusion Criteria**

Two investigators (A.A. and E.S.) independently reviewed retrieved publications. We included experiments where functional outcome in a group of animals exposed to traumatic spinal cord injury and treated with allogeneic or autologous stem cells was compared with functional outcome in a control group of animals. We excluded individual comparisons that did not report (or where we could not calculate) the number of animals, the mean outcome, or its variance in each group. We excluded experiments where interventions such as growth factors were used to mobilise endogenous stem cells or where nontraumatic models of spinal cord injury were used.

**Data Extraction**

From each individual comparison we extracted data for reported outcomes. This included extraction of mean and variance data from each cohort exposed to an intervention (controls and active therapy) and from sham cohorts of normal (unlesioned and untreated) animals, and by imputation where the performance of a normal animal could be imputed from the description of the scoring scale. Stem cells were characterised as “autologous” when cells were extracted from an animal, might be manipulated in some way, then returned to the same animal; or “allogeneic” when embryonic or adult cells derived from a different animal were administered to a recipient animal. Where a publication reported more than one experiment, or where an experiment reported more than one individual comparison (for instance, increasing numbers of stem cells transplanted), we considered these separately and extracted data for each, correcting the weighting of these studies in meta-analysis to reflect the number of experimental groups served by each control group. Where different functional outcomes were reported in a single cohort of animals, we combined these outcomes using fixed effects meta-analysis (nesting), to give a summary estimate of functional outcome in that cohort, described here as a comparison. Where a test involved exposing the animal to increasing intensities of the same stimulus (for instance, in allodynia testing), we used data for the median intensity. For sensory tests, only data for stimulation distal to the lesion were included. Where functional outcome was measured at different times, we extracted data for the last time point reported.

Study quality was assessed using a checklist adapted from good laboratory practice guidelines for in vivo stroke modelling [59] and the CAMARADES quality checklist [60]. The checklist comprised (i) publication in a peer-reviewed journal, (ii) statements describing control of temperature, (iii) randomisation to treatment group, (iv)
allocation concealment, (v) blinded assessment of outcome, (vi) avoidance of anaesthetics with known marked intrinsic neuroprotective properties, (vii) sample size calculation, (viii) compliance with animal welfare regulations, and (ix) whether the authors declared any potential conflict of interest.

Analysis

For each individual comparison, we calculate a normalised effect size \([\text{normalised mean difference}]\) as the percentage improvement \(\left(\frac{\text{恥}-\text{恥}}{\text{恥}}\right)\) or worsening \(\left(\frac{\text{恥}-\text{恥}}{\text{恥}}\right)\) of outcome in the treatment group using the following formula:

\[
ES_i = 100\% \times \left(\frac{\bar{x}_c - \bar{x}_{\text{sham}}}{\bar{x}_c - \bar{x}_{\text{sham}}}\right),
\]

where \(\bar{x}_c\) and \(\bar{x}_{\text{sham}}\) are the mean reported outcomes in the control and treatment group, respectively, and \(\bar{x}_{\text{sham}}\) is the mean outcome for a normal (unlesioned and untreated) animal. In this calculation, the score achieved by the sham animals acts as the “fixed zero value” or baseline allowing the difference between the sham and treatment groups to be expressed as a ratio. This ratio takes into account differences in the “direction” of individual neurobehavioural scales.

Its corresponding standard error was calculated using:

\[
SE_i = \sqrt{\frac{SD^2_c}{n_c} + \frac{SD^2_{rx}}{n_{rx}}},
\]

where \(n_c\) refers to the number of animals in the control group and \(n_{rx}\) refers to the number of animals in the treatment group. \(SD^2_c\) and \(SD^2_{rx}\) are the normalised standard deviations for the control and treatment group, respectively. These were calculated using the formulae:

\[
SD^2_c = 100 \times \frac{SD_c}{\bar{x}_c - \bar{x}_{\text{sham}}}
\]

and

\[
SD^2_{rx} = 100 \times \frac{SD_{rx}}{\bar{x}_{rx} - \bar{x}_{\text{sham}}},
\]

where \(SD_c\) and \(SD_{rx}\) are the reported standard deviation for the control and treatment group, respectively.

We then used DerSimonian and Laird random effects weighted mean effect meta-analysis to calculate a summary estimate of effect size; results are presented as the percentage improvement in outcome and its 95% confidence intervals. The variability of the outcomes assessed is presented as the heterogeneity statistic \(I^2\) with \(n-1\) degrees of freedom.

The analysis was stratified according to (i) the approach to stem cell therapy (allogeneic, autologous, embryonic, source of cells, \textit{ex vivo} manipulation), (ii) biological factors (number of cells, time and route of administration, time of assessment of outcome), (iii) aspects of study design (anaesthesia, species of animal, immunosuppression, model and severity of spinal cord injury), and (iv) elements of study quality.

The extent to which study design characteristics explained differences between studies was assessed using meta-regression with the \textit{metareg} function of STATA/SE10, and the significance level was set at \(p<0.05\). The meta-regression was univariate rather than multivariate; and we calculated adjusted \(R^2\) values (a measure of how much residual heterogeneity is explained by the model) to explain the proportion of the observed variability in the observed effect size for a group of experiments explained by variation in the independent variable in question [61].

We sought evidence of publication bias using a funnel plot, Egger regression, and Trim and Fill [62]. A detailed description of the statistical methods used for meta-analysis and meta-regression can be found in [63].

Supporting Information

Table S1 Included studies. First author, publication year, stem cell used, species of host animal, number of animals, number of cells, time of treatment in relation to injury, anaesthetic used, type of injury, route of delivery, and outcome measure reported for studies included in the review.

Table S2 Quality of included studies/reporting of (1) publication in a peer-reviewed journal, (2) statement describing control of temperature, (3) randomisation to treatment group, (4) allocation concealment, (5) blinded assessment of outcome, (6) avoidance of anaesthetic with known marked intrinsic neuroprotective properties, (7) sample size calculation, (8) compliance with animal welfare regulations, and (9) statement of any potential conflict of interest.

Text S1 PRISMA 2009 Checklist.

Author Contributions

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: DWH MRM ESS. Analyzed the data: AA ESS JSL TEW PS MRM DWH. Contributed reagents/materials/analysis tools: ESS MRM DWH. Wrote the paper: AA ESS JSL TEW PS MRM DWH.

References