RESEARCH

Discrepancies in autologous bone marrow stem cell trials and enhancement of ejection fraction (DAMASCENE): weighted regression and meta-analysis

OPEN ACCESS

Alexandra N Nowbar *cardiovascular scientist*, Michael Mielewczik *postdoctoral biologist*, Maria Karavassilis *clinical medical student*, Hakim-Moulay Dehbi *statistician*, Matthew J Shun-Shin *academic clinical fellow in cardiology*, Siana Jones *cardiovascular physiologist*, James P Howard *physician*, Graham D Cole *BHF clinical research training fellow*, Darrel P Francis *professor of cardiology*, on behalf of the DAMASCENE writing group

International Centre for Circulatory Health, National Heart and Lung Institute, Imperial College London, London W2 1LA, UK

Abstract

Objective To investigate whether discrepancies in trials of use of bone marrow stem cells in patients with heart disease account for the variation in reported effect size in improvement of left ventricular function.

Design Identification and counting of factual discrepancies in trial reports, and sample size weighted regression against therapeutic effect size. Meta-analysis of trials that provided sufficient information.

Data sources PubMed and Embase from inception to April 2013.

Eligibility for selecting studies Randomised controlled trials evaluating the effect of autologous bone marrow stem cells for heart disease on mean left ventricular ejection fraction.

Results There were over 600 discrepancies in 133 reports from 49 trials. There was a significant association between the number of discrepancies and the reported increment in EF with bone marrow stem cell therapy (Spearman's *r*=0.4, P=0.005). Trials with no discrepancies were a small minority (five trials) and showed a mean EF effect size of –0.4%. The 24 trials with 1-10 discrepancies showed a mean effect size of 2.1%. The 12 with 11-20 discrepancies showed a mean effect of size 3.0%. The three with 21-30 discrepancies showed a mean effect size of 5.7%. The high discrepancy group, comprising five trials with over 30 discrepancies each, showed a mean effect size of 7.7%.

Conclusions Avoiding discrepancies is difficult but is important because discrepancy count is related to effect size. The mechanism is unknown but should be explored in the design of future trials because in the five trials without discrepancies the effect of bone marrow stem cell therapy on ejection fraction is zero.

Introduction

Autologous bone marrow stem cells offer an exciting opportunity for improvement of left ventricular function, reverse remodelling, and scar size reduction[1] in patients with ischaemic heart disease.[2] Results, however, have been conflicting. The reason for the differences between the various trials of effect on left ventricular function has so far not been identified. Meta-analyses have confirmed a significant positive effect on average but have found no clear explanation for the conflicts between individual trials.[1 3 4]

It has recently been discovered that some pioneering trials of autologous bone marrow stem cells have unexplained discrepancies that cast doubt on their validity.[5] It was not possible to report this directly in the journals that published the trials.[5] Discrepancies in reports have never been systematically

Correspondence to: A N Nowbar alexandra.nowbar09@imperial.ac.uk

Extra material supplied by the author (see http://www.bmj.com/content/348/bmj.g2688?tab=related#webextra)

Appendix 1: Design discrepancies
Appendix 2: Methods and baseline characteristics discrepancies
Appendix 3: Results discrepancies
Appendix 4: Trials included and links to reports
Appendix 5: Assessment of Cochrane risk of bias
Appendix 6: Calculation of ejection fraction effect size
Appendix 7: Discrepancies in safety study
Appendix 8: Funnel plot
Appendix 9: Univariate and multivariate analyses
Appendix 10: Meta-analysis
Appendix 11: Methods supplement

explored as a possible explanatory variable for the effect size of autologous bone marrow stem cells on ejection fraction.

We examined reports of the randomised controlled trials of bone marrow stem cell therapy for discrepancies of design, methods, or results and examine the relation between number of discrepancies and effect sizes reported. We defined a discrepancy as two (or more) reported facts that cannot both be true because they are logically or mathematically incompatible.

Methods

Search strategy and eligibility criteria

We searched Embase and PubMed (1966 to April 2013), using the following search strategy: ("bone marrow cell" OR "bone marrow cells" OR "stem cells" OR "stem cell" OR "progenitor cell" OR "progenitor cells") AND ("myocardial infarction" OR "coronary artery disease" OR cardiomyopathy OR "heart failure") AND random*.

We also manually searched citation lists[1 5] and PubMed links to related citations. We included trials that met the following criteria:

- Trial reporting the effect on mean ejection fraction of infusion of autologous stem cells derived from bone marrow in patients with acute or established cardiac disease

- At least one publication by the authors described it as randomised

- Available through our institution and in a language understood well by at least one investigator.

For each trial identified by this method, we used the international standard randomised controlled trial number registry (isrctn.org), ClinicalTrials.gov registry, PubMed, Google, and manual evaluation of references to search for other reports from that trial published until end of April 2013.

Data extraction

Two authors (ANN and SJ) extracted data from each trial, with disputes resolved by a third author (MJS). When the ejection fraction was measured by more than one imaging technique (magnetic resonance imaging (MRI), echocardiography, radionuclide imaging, left ventriculography), we used the data from the technique specified as the primary endpoint. If this was not defined, we used the technique that was highlighted in the abstract or (if the abstract was not specific) given priority in the conclusion or (if not mentioned in either) given priority in the results. When the ejection fraction effect size was reported at multiple time points, we used that of the longest follow-up.

We defined the ejection fraction effect size[6] as the change in ejection fraction in the active arm minus the change in ejection fraction in control arm. We used this if it was stated directly in the trial. If it was not directly stated, we calculated it from the changes provided in each arm or, when these were not provided, from the baseline and follow-up values in each arm.

We used the standard error of the effect size if it was stated explicitly in the trial. If the confidence interval of the effect size was given instead, we extracted the standard error. If only the standard errors or standard deviations or confidence intervals of the changes in each arm were provided, we then extracted the two standard deviations and sample sizes and used them to calculate the standard error of the estimate.

Detection of discrepancies

The trials were then examined for discrepancies, which were categorised into the following three types[5]:

1. Discrepancies in the design—for example, conflicting statements as to whether the study was randomised (tabulated in appendix 1)

2. Discrepancies in methods and baseline characteristics—for example, sample or subgroup sizes that could not be an integer number of patients (listed in appendix 2)

3. Discrepancies in results—for example, conflicts between tables and figures or impossible values (listed in appendix 3).

Eight authors (ANN, DPF, GDC, HD, JPH, MM, MJS, SJ) read all the reports, except those of the four trials for which the discrepancies had already been found and published.[5] Proposed discrepancies were discussed. A discrepancy was declared valid for inclusion in the study only if no member of the group could find a valid explanation.

Contradictions in numerical values were considered as discrepancies but errors in spelling or grammar were not. If the same conflicting statements appeared more than once (for example, a trial repeatedly described as randomised in one publication, and repeatedly described as accepter-rejecter in another), this was considered a single discrepancy.

Trials, and their reports, were coded with a "t" number or "r" number, respectively. Appendix 4 provides a decoded list of trials and reports with web links to the sources. Each discrepancy was numbered with a three digit code after the two digit "t" number. The first digit of the discrepancy code was allocated according to the type of discrepancy—that is, 1, 2, or 3, as listed above.

Of the trials, four had already undergone this process of identification and checking of discrepancies therefore for these we used the discrepancies as previously published by our group[5] (t07, t08, t21, and t49). Because the present publication focuses on counting discrepancies, where our previous report[5] listed more than one discrepancy on a single row, we now separated these on to individual rows.

Assessing risk of bias

Risk of bias in the included trials was assessed with the Cochrane Collaboration's risk of bias assessment tool (see appendix 5). Each trial was assessed by two independent observers and any differences resolved by a third observer.

Data analysis

We visualised the relation between the ejection fraction effect size and the discrepancy count with a scatter plot and quantified it with Spearman's rank correlation coefficient. It was further visualised with a histogram with trials grouped by the number of discrepancies in intervals of 10. Means were weighted by sample size. We constructed a funnel plot of the data and used Egger's test to assess asymmetry.[7]

A meta-analysis was conducted for trials that provided sufficient data to weight the effect size estimates by a function of the reciprocal of the square of the standard error of the effect size estimate. We pooled the data on ejection fraction effect size for this subset of trials using a random effects model and present them as weighted mean differences with 95% confidence intervals.

As part of an exploratory analysis we performed univariate and multivariate linear regression analyses including the discrepancy

count, sample size, and the five specific domains of the Cochrane risk of bias tool as predictor variables for the effect size. Any aspect of bias that was agreed to be "unclear" was treated as a "no." The multivariate model was built by stepwise backwards selection based on the Akaike information criterion.[8]

Data analysis was carried out with R (version 3.0.2, R Foundation),[9] the graphics package ggplot2 (version 0.9.0),[10] and the meta-analysis package metafor (version 1.9).

Results

Figure 1 shows how we identified trials⇓. We identified 49 randomised trials reporting the effect of bone marrow stem cells on ejection fraction for cardiac disease (appendix 6). Length of follow-up ranged from three months to 65 months, with modal duration of six months. Each trial had between one and 13 reports. There were 133 reports in total (see appendix 4 for the list of trials and reports). We identified one study during the search in which the focus was on safety alone (appendix 7).

Discrepancies in design, methods, or results

We identified 604 instances of discrepancy (appendices 1-3) within a trial report or between reports of that trial. We identified 44 discrepancies in the reports of the study on safety. There were many types of discrepancy, as shown by the examples in table 1⇓. Table 2 shows examples from the trials of phenomena that are unusual but were not counted as discrepancies⇓.

Many aspects of the reports contained discrepancies. Even the primary endpoint was not spared (t09/308, t14/301, t19/305). Sometimes the discrepancies seemed to affect whether the difference between trial arms was significant (t10/301). Effect size, defined as the increment in ejection fraction from bone marrow stem cell therapy, ranged from −3.9 to 14 percentage units. Numbers of discrepancies in individual trials across all their reports ranged from 0 to 89.

There was a significant correlation between the number of discrepancies and the reported ejection fraction effect size (Spearman's r=0.4, P=0.005, fig 2⇓). There were only five studies with no discrepancies, and these showed a mean effect size of −0.4%, with this average weighted by sample size. The 24 trials with one to 10 discrepancies showed mean effect size of 2.1%; the 12 with 11 to 20 discrepancies showed mean effect size of 3.0%; the three with 21-30 discrepancies showed mean effect size of 5.7%; and five high discrepancy trials, with over 30 discrepancies each, showed a mean effect size of 7.7% (fig 3⇓).

Publication bias

The funnel plot (appendix 8) did not show significant asymmetry (Egger's test P=0.4) that would suggest publication bias.

Discrepancies and risk of bias

The results of the exploratory univariate and multivariate analyses are presented in appendix 9. Only the number of discrepancies (P<0.001) and sequence generation (P=0.03) remained significant contributors to the effect size (adjusted R^2=0.38, P<0.001).

Meta-analysis of studies providing information on uncertainty of effect size

We could adequately extract the standard error of the effect size estimate in only 31 trials to allow a formal meta-analysis to be conducted using this information (appendix 10). The weighted

mean effect size was 0.0 (95% confidence interval −4.67 to 4.65) for trials with no discrepancies; 1.9 (0.30 to 3.57) for trials with 1-10 discrepancies; 4.6 (1.64 to 7.61) for trials with 11-20 discrepancies; 4.4 (−0.97 to 9.75) for trials with 21-30 discrepancies; and 10.4 (8.44 to 12.36) for trials with more than 30 discrepancies.

Discussion

Whenever we present scientific information we risk introducing conflicting statements that form discrepancies. Our study shows that scientists who achieve progressively better consistency of reporting find progressively smaller effects on ejection fraction of treatment with bone marrow stem cells. In trials with a discrepancy count of zero, the ejection fraction effect seems to be zero.

Study limitations

We were unable to blind ourselves to effect size because this was embedded within the report itself. There might be additional unidentified discrepancies. Our work involved developing newly derived mathematical limits on what is possible (appendix 11). There could be other such limits that have not yet been established. We invite readers to contribute either new discrepancies or new general methods for identifying the impossible.

Our method of counting discrepancies is imperfect because there is no universally accepted convention. We have tried to be consistent (appendices 1-3) but are open to suggestions from readers. Some readers might consider what we list as a single discrepancy to be several (for example, multiple repetitions of the same contradiction) or what we consider several to be just one (arguing, for example, that multiple discrepancies in a table might have been values from a different trial pasted into a manuscript accidentally).

We have taken a simple approach of including all trials, including some that we have previously identified as containing discrepancies and that showed a large effect of stem cells.[5] With exclusion of these four trials, Spearman's rank correlation coefficient in figure 2 is lower at 0.3 (P=0.03) but similar. We were able to examine only studies whose results have been reported. We examined clinical trial registries but many trials seem to have sped through to publication, with the registration step skipped.

In some cases it was not clear whether a trial was randomised. Previous meta-analysts have handled the confusion in contradictory ways, with some trials being classified as randomised in one meta-analysis and non-randomised in others.[1 15 16] Our policy considered a trial eligible if an author of the primary report stated at any stage that the trial was randomised. We recognise that other conventions for inclusion would also have been possible. This will remain a challenge as long as primary authors find the distinction puzzling.

Our main analysis is weighted simply by sample size because this was available in every trial. In formal meta-analysis it is ideal to weight by a function of the reciprocal of the square of the standard error of the effect size estimate, but more than a third of the trials did not provide this information. For those trials that did provide sufficient data to weight the effect size estimates in this way, we conducted a formal meta-analysis (shown in appendix 10).

We have not attempted to control for the fact that some trials issued more reports than others, and different reports gave information to different levels of depth. It is difficult to

satisfactorily control for this because sometimes the multiple reports are pure duplication, sometimes they cover non-overlapping information, and sometimes they are contradictory.

We excluded five reports because they were in Chinese language journals to which we had no access.[17-23] We do not know to what extent this access limitation might have biased our results. Although it might have been possible to use additional routes to obtain the full text, and then arrange a translation, the translation process could always be suspected as a source of imperfection.

Results in context of other similar work

There seem to be no other studies exploring the relation between effect size of bone marrow stem cell therapy and the number of discrepancies in the reports. Several meta-analyses have covered bone marrow stem cell therapy for increasing ejection fraction but have not discussed the discrepancies in the reports. They all concluded that the average effect was a significant increase in ejection fraction.[1 3] This was reiterated in a Cochrane review[4] and the recent MESS (meta-analysis of cardiac stem cell studies) meta-analysis.[24]

Our findings expand on these meta-analyses. Our study concurs that viewing all the studies together as a single entity, there is on average a positive effect on ejection fraction. However, we found that the positivity was not consistent across the spectrum of discrepancy count. The studies with the most discrepancies seem to be contributing most to the positivity, while the studies with no discrepancies show a zero effect. Averaging effect size across all studies might therefore not be wise because it does not reflect their varying factual accuracy.

Standard meta-analyses include quality assessment, but this does not seem to involve identifying or quantifying factual discrepancies.[5] Well conducted meta-analyses have somehow classed many studies with numerous discrepancies as high quality.[1 3 25-27]

Discrepancy count seems additive to a traditional assessment of risk of bias (appendix 9). If our findings are verified by other workers in other specialties, then addition of discrepancy checking, and ideally cross checking with raw data, might make meta-analysis more illuminating.

Possible explanations

We do not know the cause of the discrepancies. We have asked for resolution of over 150 discrepancies through journals.[5] None were resolved, although we found it triggered correspondence from lawyers.

One possibility is that authors might feel pressure for results to match expectations. One signal of a misguided desire to please is the phenomenon of directed editing of rounded percentages to force them to add up to 100%. In reality, correctly rounded percentages should often not add up to 100% when there are many categories.[28] The effect of even a little bias can be surprisingly dramatic.[29 30]

Secondly, exciting new treatments might be reported before full checking. One sign of this, in the neighbouring specialty of cardiomyocyte-derived stem cell therapy, is the insertion of the word "randomised" into the title of the journal publication[31] that was not present in the manuscript finalised by the authors[32] on Pubmed Central. There were seven controls in total, but after subtraction of the four who were not randomised and one who was randomised to stem cells but refused treatment, the number

of randomised controls was only two. For the *Lancet* this was a new low.

Thirdly, bone marrow stem cell therapy might be less effective when it is carried out in a rigidly standardised way. Centres with less attention to detail might incorporate an unnoticed contaminant that enhances the effect of treatment. These centres might produce reports with more discrepancies. In support of this, just over a fifth of the trials (t07, t08, t09, t11, t21, t27, t33, t35, t40, t43, t44, t49) showed ejection fraction effects of 7% or more, but these trials accounted for more than half of all the discrepancies.

The final possibility is that in the reports with the fewest discrepancies, the ejection fraction effect might also have been measured with least error. If so, the true effect of bone marrow stem cells on ejection fraction is zero.

Implications for correctness of values reported in trials

When trials provide full data, serious errors in reporting can come to light, such as omitted patients,[33] reclassification of causes of death,[34 35] or studies based on fictitious data.[36 37] As full data disclosure is rare, readers currently cannot estimate how many trial reports are incorrect. It is essential that there is open access to data.[38]

In our study, the reported standard deviation of the NYHA score (New York Heart Association classification for chronic heart failure) offers a unique window into correctness of reporting, which does not require raw data. If NYHA data for individual patients are fabricated, then the means and standard deviations will remain mathematically possible.

It is only when standard deviations are not correctly calculated from real NYHA values that mathematically impossible values can arise. Of 11 trials reporting a standard deviation of NYHA, the values in five (45%) are mathematically impossible.

The NYHA score is simple to measure, and the standard deviation is simple to calculate. Ejection fraction effect is more complex to measure and its statistical significance is more complex to calculate. We are concerned that, if simple calculations on simple variables are definitely incorrect in almost half of trials, then the more subtle statistical statements regarding more subtle variables might in most cases also be incorrect.

Implications for interpreting trial design

Readers need to know whether a trial is randomised or not, but the reports were sometimes vague or even contradictory. Some trials were initially reported as accepter-rejecter (non-randomised) and later as randomised[5] (t21, t41, t49). In one, a later publication recalled the existence of a placebo control group (t07r3). In this specialty, patients' voluntary choice is sometimes considered a form of randomisation, a policy accepted by some journals (t21/102).[5 39] Identical tables and identical figures have inexplicably been presented as results of different studies,[5] with different names, different designs (randomised versus accepter-rejecter), and different sample sizes (t07r1, t07r10).

Journals could resolve such discrepancies but currently do not consider this a priority[5 40] (t07r1, t07r10, t21r5, t21r11, t49r1, t49r3).

Implications for safety of bone marrow stem cell therapy

The safety of bone marrow stem cell therapy is underlined by a large report focusing on this.[41-44] Unfortunately it too contains

many discrepancies (appendix 7), including impossible percentages and conflicts between tables and figures, perhaps because the reassuring findings had to be made available with urgency.

Implications for clinicians and researchers

If patients ask for advice on which bone marrow stem cell trial to enter, we want to maximise the benefit to them, while maximising their contribution to reliable evidence for future patients. Unfortunately, these seem to be in conflict (fig 3⇓).

Sometimes researchers feel that only findings of positive effects indicate scientific success. But meticulously reported studies reporting neutral effects are vital contributions to science. It is more valuable to have a reliable report of a small improvement than an unreliable report of a large improvement. Error-free reporting is difficult to achieve, as we have found in our own experience.[45] Only 10% of these trials were reported without introducing discrepancies. We consider these to be the greatest scientific successes, even though the effect size was unfortunately zero.

Several lessons can be drawn for the design of future trials of bone marrow stem cells. Prior registration on a public clinical trial registry was not universal and would have been helpful in distinguishing unambiguously between trials that were multiply published or merely identical by coincidence.

We recommend that reports include a spreadsheet of all the data used for construction of the tables, so that incorrect values could be more easily identified. Readers should accept that authors cannot avoid errors; in turn authors should correct errors promptly and indicate clearly when later reports incorporate corrections. It should be remembered that the 604 discrepancies listed in appendices 1-3 are unlikely to be all the errors. They are only those detectable by us without any information beyond the published reports. Disclosing the individual patient data could help to correct more errors.

It is important for studies using change in ejection fraction as an endpoint to be properly designed to resist error and to have adequate sample size to combat the effects of biological variability. Left ventricular ejection fraction is a mutable variable, which in some modalities is easily manipulated innocently by clinicians who have prior beliefs on what a realistic value should be for a particular patient. Sample size planning can sometimes be erroneously omitted when clinicians are enthusiastic to "demonstrate the effectiveness" of a treatment seen as exciting.

Conclusions

It is difficult to avoid discrepancies in clinical trial reports. Trials with progressively fewer discrepancies tend to find progressively smaller effects on ejection fraction of bone marrow stem cell therapy. The reason for this association is unknown. The few trials for which the discrepancy count was zero had a stem cell effect size that was also zero.

Notes added at proof stage: The institution of t07, t08, t21, and t49 is recently reported to have identified evidence of misconduct[46] and has notified the city prosecutor. The institution of the SCIPIO trial[31] is recently reported to have requested that the publication be retracted.[47]

1 Jeevanantham V, Butler M, Saad A, Abdel-Latif A, Zuba-Surma EK, Dawn B. Adult bone marrow cell therapy improves survival and induces long-term improvement in cardiac parameters: a systematic review and meta-analysis. *Circulation* 2012;126:551-68.
2 Finegold JA, Asaria P, Francis DP. Mortality from ischaemic heart disease by country, region, and age: statistics from World Health Organization and United Nations. *Int J Cardiol* 2012;168:934-45.
3 Zimmet H, Porapakkham P, Porapakkham P, Sata Y, Haas SJ, Itescu S, et al. Short- and long-term outcomes of intracoronary and endogenously mobilized bone marrow stem cells in the treatment of ST-segment elevation myocardial infarction: a meta-analysis of randomized control trials. *Eur J Heart Fail* 2012;14:91-105.
4 Clifford DM, Fisher SA, Brunskill SJ, Doree C, Mathur A, Watt S, et al. Stem cell treatment for acute myocardial infarction. *Cochrane Database Sys Rev* 2012;2:CD006536.
5 Francis DP, Mielewczik M, Zargaran D, Cole GD. Autologous stem cell therapy in heart disease: silent discrepancies and scientific and editorial responses. *Int J Cardiol* 2013;168:3381-403.
6 Bouri S, Whinnett ZI, Cole GD, Manisty CH, Cleland JG, Francis DP. Definitions of outcome, response and effect in imaging research to avoid confusion. *JACC Cardiovascular Imaging* 2014;7:104-6.
7 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.
8 Bozdogan H. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika* 1987;52:345-70.
9 R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, 2013.
10 Wickham H. Ggplot2: elegant graphics for data analysis. [Version 0.9.0] Springer, 2009.
11 Arnesen H, Lunde K, Aakhus S, Forfang K. Cell therapy in myocardial infarction. *Lancet* 2007;369:2142-3.
12 Lonza. BioWhittaker X-VIVO Media Systems. http://bio.lonza.com/uploads/tx_mwaxmarketingmaterial/Lonza_BenchGuides_X-VIVO_Media_Information_Sheet.pdf.
13 Fraccarollo D, Galuppo P, Bauersachs J. Novel therapeutic approaches to post-infarction remodelling. *Cardiovasc Res* 2012;94:293-303.
14 Wikipedia. As I was going to St Ives. http://en.wikipedia.org/wiki/As_I_was_going_to_St_Ives.
15 Lipinski MJ, Biondi-Zoccai GG, Abbate A, Khianey R, Sheiban I, Bartunek J, et al. Impact of intracoronary cell therapy on left ventricular function in the setting of acute myocardial infarction: a collaborative systematic review and meta-analysis of controlled clinical trials. *J Am Col Cardiol* 2007;50:1761-7.
16 Charwat S, Gyongyosi M, Lang I, Graf S, Beran G, Hemetsberger R, et al. Role of adult bone marrow stem cells in the repair of ischemic myocardium: current state of the art. *Exp Hematol* 2008;36:672-80.
17 Abbasi M, Javan H, Alizadeh B, Afzalnia S. Can intracoronary stem cell injection permanently improve cardiac function after myocardial infarction? *Interact Cardiovasc Thorac Surg* 2011;12:229-31.
18 Zhang C, Sun A, Zhang S, Yao K, Wu C, Fu M, et al. Efficacy and safety of intracoronary autologous bone marrow-derived cell transplantation in patients with acute myocardial infarction: insights from randomized controlled trials with 12 or more months follow-up. *Clin Cardiol* 2010;33:353-60.
19 Lu MJ, Zhao SH, Liu S, Zhang PH, Jiang SL, Zhang Y, et al. Assessment of therapeutic effects of stem cell transplantation in heart failure patients with old myocardial infarction by magnetic resonance imaging. *Zhonghua Xin Xue Guan Bing Za Zhi* 2008;36:969-74.
20 Yao K, Huang RC, Ge L, Qian JY, Li YL, Xu SK, et al. Observation on the safety: clinical trial on intracoronary autologous bone marrow mononuclear cells transplantation for acute myocardial infarction. *Zhonghua Xin Xue Guan Bing Za Zhi* 2006;34:577-81.

What is already known on this topic

Autologous bone marrow stem cell therapy has been reported to substantially increase cardiac function

The trials have differed in the effect sizes they reported, for reasons that are not clear

What this study adds

Many reports of trials of bone marrow stem cell therapy contain factual discrepancies

The number of discrepancies in a trial was significantly associated with the reported effect size

Trials with over 30 discrepancies report a large effect size. Classes of trials with fewer discrepancies have found progressively smaller effect sizes, culminating in discrepancy-free trials reporting an effect size of zero

21 Huang RC, Yao K, Zou YZ, Ge L, Qian JY, Yang J, et al. Long term follow-up on emergent intracoronary autologous bone marrow mononuclear cell transplantation for acute inferior-wall myocardial infarction. *Zhonghua Yi Xue Za Zhi* 2006;86:1107-10.

22 Huang RC, Yao K, Li YL, Zhang YQ, Xu SK, Shi HY, et al. Transplantation of autologous bone marrow mononuclear cells on patients with idiopathic dilated cardiomyopathy: early results on effect and security. *Zhonghua Xin Xue Guan Bing Za Zhi* 2006;34:111-3.

23 Wang JA, Xie XJ, He H, Sun Y, Jiang J, Luo RH, et al. A prospective, randomized, controlled trial of autologous mesenchymal stem cells transplantation for dilated cardiomyopathy. *Zhonghua Xin Xue Guan Bing Za Zhi* 2006;34:107-10.

24 Gyongyosi M, Lemarchand P, Roncalli J, Lunde K, Tendera M, Wojakowski W, et al. Safety and efficacy of cardiac cell therapy: first results of the MESS (meta-analysis of cardiac cell studies) database including individual patient data of 13 randomized and 3 cohort studies. *J Am Col Cardiol* 2013;61:(10_S).

25 Van Vark LC, Bertrand M, Akkerhuis KM, Brugts JJ, Fox K, Mourad JJ, et al. Angiotensin-converting enzyme inhibitors reduce mortality in hypertension: a meta-analysis of randomized clinical trials of renin-angiotensin-aldosterone system inhibitors involving 158,998 patients. *Eur Heart J* 2012;33:2088-97.

26 Yui Y. Concerns about the Jikei Heart Study. *Lancet* 2012;379:e48.

27 Sawada T, Yamada H, Dahlof B, Matsubara H, KYOTO HEART Study Group. Effects of valsartan on morbidity and mortality in uncontrolled hypertensive patients with high cardiovascular risks: KYOTO HEART Study. *Eur Heart J* 2009;30:2461-9.

28 Mosteller F, Youtz C, Zahn D. The distribution of sums of rounded percentages. *Demography* 1967;4:850-8.

29 Shun-Shin M, Francis DP. Why are some studies of cardiovascular markers unreliable? The role of measurement variability and what an aspiring clinician scientist can do before it is too late. *Prog Cardiovasc Dis* 2012;55:14-24.

30 Francis DP. How easily can omission of patients, or selection amongst poorly-reproducible measurements, create artificial correlations? Methods for detection and implications for observational research design in cardiology. *Int J Cardiol* 2013;167:102-13.

31 Bolli R, Chugh AR, D'Amario D, Loughran JH, Stoddard MF, Ikram S, et al. Cardiac stem cells in patients with ischaemic cardiomyopathy (SCIPIO): initial results of a randomised phase 1 trial. *Lancet* 2011;378:1847-57.

32 Bolli R. Effect of cardiac stem cells in patients with ischemic cardiomyopathy: initial results of the SCIPIO trial. www.ncbi.nlm.nih.gov/pmc/articles/PMC3614010/.

33 Vedula SS, Li T, Dickersin K. Differences in reporting of analyses in internal company documents versus published trial reports: comparisons in industry-sponsored trials in off-label uses of gabapentin. *PLoS Med* 2013;10:e1001378.

34 Home PD, Pocock SJ, Beck-Nielsen H, Curtis PS, Gomis R, Hanefeld M, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. *Lancet* 2009;373:2125-35.

35 Marciniak TA. Cardiovascular events in the RECORD trial. www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/EndocrinologicandMetabolicDrugsAdvisoryCommittee/UCM218484.pdf.

36 Feringa HH, Elhendy A, Bax JJ, Boersma E, de Jonge R, Schouten O, et al. Baseline plasma N-terminal pro-B-type natriuretic peptide is associated with the extent of stress-induced myocardial ischemia during dobutamine stress echocardiography. *Coron Artery Dis* 2006;17:255-9 (retraction published *Coron Artery Dis* 2013;24:725).

37 Erasmus Medical Centre. Report on the 2012 follow-up investigation of possible breaches of academic integrity. 2012. http://cardiobrief.files.wordpress.com/2012/10/integrity-report-2012-10-english-translation.pdf.

38 Lehman R, Loder E. Missing clinical trial data. *BMJ* 2012;344:d8158.

39 Strauer BE, Yousef M, Schannwell CM. Corrigendum to "The acute and long-term effects of intracoronary Stem cell Transplantation in 191 patients with chronic heARt failure: the STAR-heart study". *Eur J Heart Fail* 2013;15:360.

40 Cole G, Francis DP. Comparable or STAR-heartlingly different left ventricular ejection fraction at baseline? *Eur J Heart Fail* 2011;13:234.

41 De Rosa S, Seeger FH, Honold J, Fischer-Rasokat U, Lehmann R, Fichtlscherer S, et al. Procedural safety and predictors of acute outcome of intracoronary administration of progenitor cells in 775 consecutive procedures performed for acute myocardial infarction or chronic heart failure. *Circ Cardiovasc Interv* 2013;6:44-51.

42 Leistner DM, Schmitt J, Palm S, Klotsche J, Estel S, Fink A, et al. Intracoronary administration of bone marrow-derived mononuclear cells and arrhythmic events in patients with chronic heart failure. *Eur Heart J* 2011;32:485-91.

43 Ehrlich JR. Stem cell therapy and proarrhythmia: a review of clinical experience. http://spo.escardio.org/eSlides/view.aspx?eevtid=45&fp=761.

44 De Rosa S, Seeger F, Honold J, Fischer-Rasokat U, Lehmann R, Fichtlscherer S, et al. P1676—Procedural safety and predictors of acute outcome of intracoronary administration of progenitor cells in 775 consecutive procedures performed for acute myocardial infarction of chronic heart failure. *Clin Res Cardiol* 2011;100:(suppl 1). www.abstractserver.de/dgk2011/ft/abstracts/P1676.htm.

45 Nijjer SS, Pabari PA, Stegemann B, Palmieri V, Leyva F, Linde C, et al. The limit of plausibility for predictors of response: application to biventricular pacing. *JACC Cardiovasc Imaging* 2012;5:1046-65.

46 Nature Blogs. Evidence of misconduct found against cardiologist. http://blogs.nature.com/news/2014/02/evidence-of-misconduct-cardiologist.html

47 Boston Globe. Brigham researcher facing new questions after retraction. www.bostonglobe.com/lifestyle/health-wellness/2014/04/11/harvard-investigation-leads-expression-concern-brigham-led-stem-cell-research/iTubGX14dxVXqVT4K1UzRP/story.html

Tables

Table 1| Spectrum of discrepancies in published reports of autologous bone marrow stem cell trials and enhancement of ejection fraction

Discrepancy	Examples
Studies that are reported by authors to be both accepter-rejecter and randomised	Authors published data as accepter-rejecter, and then later described similar (t41/101) or same (t07/101) data as randomised (t08/101, t21/102, t49/103)
Impossible numbers of patients, percentages, sums, or summary statistics	100% of 6 patients referred to as 5 patients (t01/205) or 50% of 9 patients on a drug (t38/208). 20 patients in 3 groups whose sizes add up to 19 (t16/201). Standard deviation (SD) impossibly wide (t43/308) or mathematically impossible (t34/301) (explanation in appendix 11). Large identical changes in mean in both recipients and control with no change in SD (t21/206)
Sex reclassification	Women present in early reports seem to have become men by later reports (t01/202, t21/203)
Zero or negative NYHA class (which can only be I, II, III, or IV)	NYHA class of 0 postoperatively, only in stem cell recipients, giving mean of 0.7 (t35/302). One patient seems to have NYHA of –5 (t07/334)
Indicating non-significant differences as significant, or significant differences as non-significant (actively or by omission)	Mean values of 2.38 (SEM 0.26) and 2.2 (SEM 0.20) reported as significantly different, but *t* test result not significant (t22/301). Groups that differ with P<0.001 described as comparable (t07/305, t49/202). Readers not informed of highly significant changes in control group, which if calculated show P values as low as $<1\times10^{-108}$ (t07/352)
Conflicts in protocol or follow-up	Patients who died or were lost to follow-up were still taking drugs, reporting symptoms, and undergoing tests (t41/201, t07/357, t07/358, t07/359, t07/360). Discrepancy over whether controls had sham injection or how injection could reach stated position (t35/201, t35/202). Of 41 patients, at 3 years, 12 had died or 10 had died, or perhaps none had died since all 41 reported their NYHA class at 3 years (t41/301, t41/302, t41/303)
Fiddly figures: contradictions	Conflicts between figures and numerical data (t40/302, t29/305). Measurement spread increases but SD shrinks (t46/301), or SD bars vary but SD stays same (t40/302, t40/303, t40/304, t40/305). More patients on graph showing individuals' EFs than were supposed to be in study (t42/302). Conflicts between tables in numbers of patients (t28/301) or means and SDs (t27/307)
Principal report is of significant effect, subsequent report (presumably a correction) shows effect had been smaller and non-significant	EF effect of +7.1 ("P=0.05"), but assembling effects in two subgroups shows overall effect of +6.5 (P>0.05) (t12/301)

NYHA=New York Heart Association; SEM=standard error of mean; EF=ejection fraction.

Table 2| Unusual phenomena not listed as discrepancies in published reports of autologous bone marrow stem cell trials and enhancement of ejection fraction

Phenomenon	Example
Randomised results presented intermingled with non-randomised	75 patients randomised between control and stem cells but results shown only as averages that include another 17 from uncontrolled cohort (t15r1). Significant stem cell effect was seen
Vanishing SPECT EF	All patients had radionuclide SPECT LVEF, but results not shown and indicated to be "similar trends" to echocardiographic EF, which showed significant stem cell effect (t42r1). In another trial all had radionuclide LVEF but results not shown; instead MRI LVEF substudy results are shown (t25r2). Significant stem cell effect was seen
Reversal of NYHA-mortality link	Of patients with NYHA III and IV, within one year almost all the NYHA III patients died and almost all the NYHA IV survived (t34r3)
Extraordinarily narrowly distributed EFs during follow-up	Given natural test-retest variability of EF in single individual, distribution of measurements across patients should be substantially wider, but this is not always reflected in trial data (t33r1). Large stem cell effect was seen
Delayed recollection of lack of blinding	One study initially published as double blinded, but subsequently authors issued corrigendum whose only effect was removal of words "double-blindedly" (t44r1). Large stem cell effect was seen
Controls received intracoronary injection of cell culture medium not licensed for use in humans	Control subjects received infusion into coronary arteries of cell culture medium X-VIVO 10 (t17r4) "designed to support the generation of Lymphokine Activated Killer cells"; manufacturer warns that it is "not approved for human or veterinary use, or for application to humans or animals."[11][12] Significant stem cell effect was seen
Unconventional informed consent process for randomised controlled trial	Consent for randomisation obtained from relatives rather than patients (t33r1), or not described at all (t40r2, t40r3, t44r1)
St Ives syndrome	Each patient can have many treatment episodes, each episode can be counted in more than one trial, each trial can have more than one report (with different names), each report can appear in more than one meta-analysis, and sum of patient counts from all meta-analysis can be totalled up, producing multiple levels of multiple counting.[6][13][14] (t49r1). Large stem cell effect was seen
Balls	One study reported randomisation "using a nonparticipant in the study to pick a red ball . . . or blue ball," which seems insufficient guarantee of bias-resistance (t35r1). Large stem cell effect was seen

SPECT=single photon emission computed tomography; LVEF=left ventricular ejection fraction; NYHA=New York Heart Association; SEM=standard error of mean; EF=ejection fraction; MRI=magnetic resonance imaging.
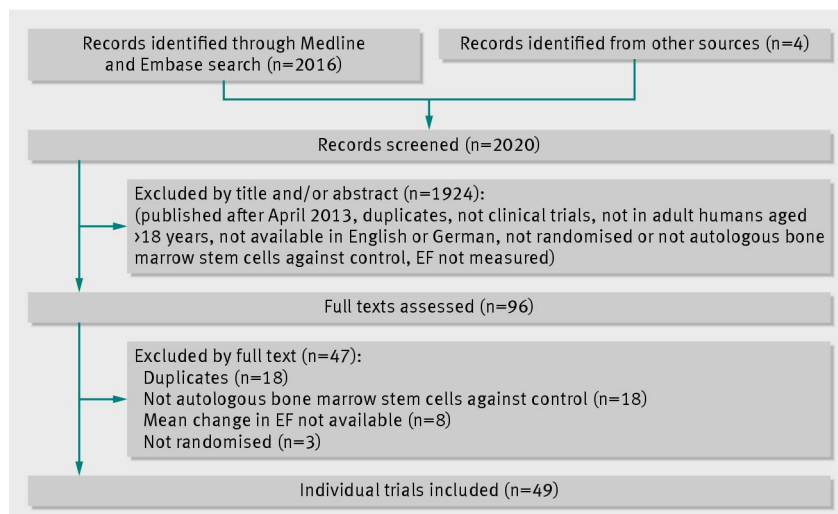
Figures

Fig 1 Identification of randomised controlled trials of autologous bone marrow stem cells for heart disease (EF=ejection fraction)
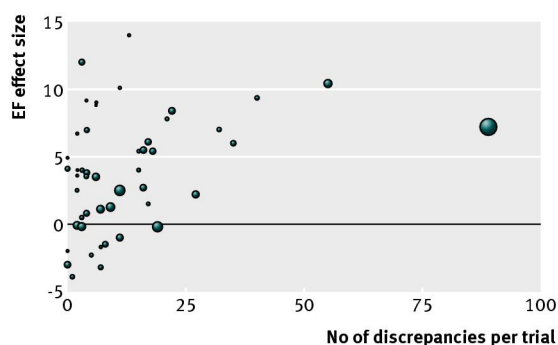

Fig 2 Correlation between number of discrepancies in trial's reports and ejection fraction (EF) effect size. Dot area is proportional to trial's sample size (Spearman's r=0.4, P=0.005)
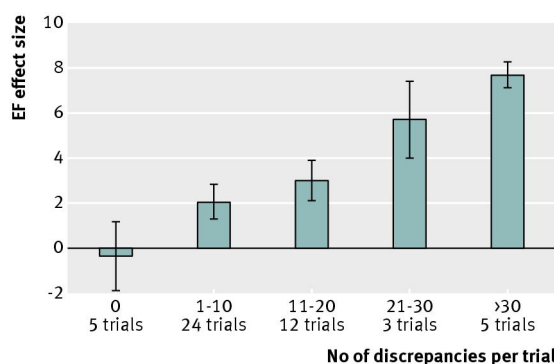

Fig 3 Mean ejection fraction (EF) effect size by number of discrepancies in trial's reports. Error bars here show only SE of mean effect size weighted for sample size across trials in each category. Formal meta-analytic confidence intervals, which fully integrate sample size and uncertainty within each trial, are available only for subset of trials (see appendix 10)