Cancer, meta-analysis and reporting biases: the case of erythropoiesis-stimulating agents

Thomy Tonia\textsuperscript{a}, Guido Schwarzer\textsuperscript{b}, Julia Bohlius\textsuperscript{a}

\textsuperscript{a} Institute of Social and Preventive Medicine, University of Bern, Switzerland
\textsuperscript{b} Institute of Medical Biometry and Medical Informatics Freiburg, Germany

Summary

Reporting and publication bias is a well-known problem in meta-analysis and healthcare research. In 2002 we conducted a meta-analysis on the effects of erythropoiesis-stimulating agents (ESAs) on overall survival in cancer patients, which suggested some evidence for improved survival in patients receiving ESAs compared with controls. However, a meta-analysis of individual patient data conducted several years later showed the opposite of our first meta-analysis, that is, evidence for increased on-study mortality and reduced overall survival in cancer patients receiving ESAs. We aimed to determine whether the results of our first meta-analysis could have been affected by publication and reporting biases and, if so, whether timely access to clinical study reports and individual patient data could have prevented this. We conducted a hypothetical meta-analysis for overall survival including all studies and study data that could have been available in 2002, at the time when we conducted our first meta-analysis. Compared with our original meta-analysis, which suggested an overall survival benefit for cancer patients receiving ESAs [hazard ratio (HR) 0.81, 95\% confidence interval (CI) 0.67–0.99], our hypothetical meta-analysis based on the results of all studies conducted at the time of the first analysis did not show evidence for a beneficial effect of ESAs on overall survival (HR 0.97, 95\% CI 0.83–1.12). Thus we have to conclude that our first meta-analysis showed misleading overall survival benefits due to publication and reporting biases, which could have been prevented by timely access to clinical study reports and individual patient data. Unrestricted access to clinical study protocols including amendments, clinical study reports and individual patient data is needed to ensure timely detection of both beneficial and harmful effects of healthcare interventions.

Key words: meta-analysis; publication bias; reporting bias; cancer; erythropoiesis-stimulating agents

Meta-analyses and the threat of reporting and publication biases

The large number of studies conducted in medical research and their sometimes conflicting results make it difficult for healthcare professionals and policy makers to reach well-informed decisions. Meta-analyses offer a possible solution, as they provide an overall effect estimate by pooling results from different studies dealing with the subject in question. In this way they may overcome problems that individual studies might have, such as small sample size and low statistical power, and may provide answers when the results of individual studies are conflicting. A correctly conducted meta-analysis should be preceded by a systematic review of the literature, which aims to identify all the available studies on the subject. It is conducted in a transparent and replicable manner, and the studies are selected and evaluated according to predefined criteria. Meta-analyses, however, might also yield misleading results. There are three main reasons why this could happen: (a) poor quality of the included studies; (b) heterogeneity between the studies; and (c) publication and reporting biases [1]. As each individual study affects the overall result of a meta-analysis, its quality should be assessed and taken into account. Study quality could be evaluated by using various criteria such as method of randomisation and allocation concealment, blinding of patients and outcome assessors, use of intention-to-treat analysis, etc. In addition, a high degree of heterogeneity between studies can also affect the overall result and researchers should seek to explain heterogeneity by stratifying results according to pre-specified groups, for instance according to type of cancer, or type of chemotherapy, etc.

The term “publication bias” refers to the fact that not all studies are published and that publication of studies follows a pattern that cannot be attributed to chance: studies with positive results tend to be published more often and earlier than studies with negative results [2, 3]. A longitudinal study of the publication outcome of clinical trial protocols submitted to five ethics committees showed that trials were 2.6 times more likely to be published if they reported statistically significant results [4]. Bourgeois and colleagues [5], on the basis of a sample of more than 500 registered clinical trials, found that studies funded by industry were more likely to be published if they reported positive results. Reporting bias refers to the selective reporting of outcomes in a published study. For instance, some outcomes that had been defined in the protocol might
not be reported [6], or the results reported are derived from per-protocol analyses rather than intention-to-treat analyses [7]. This kind of bias also seems to follow a pattern, whereby mainly the most statistically significant results or the ones meeting the authors’ assumptions are reported [8].

The case of erythropoiesis-stimulating agents in cancer patients

In 2002 we conducted a meta-analysis on the effects of erythropoiesis-stimulating agents (ESAs) in cancer patients. ESAs are recombinant human proteins resembling erythropoietin, which is a human hormone produced mainly in the kidney and which stimulates the production of red blood cells. ESAs were licensed to treat cancer-related anaemia in 1990. There are several short- or long-acting ESAs, with the most commonly used ones being epoetin-α or epoetin-β and darbepoetin. Since their introduction for the treatment of cancer-related anaemia, numerous randomised controlled trials and systematic reviews have been conducted, in order to assess their efficacy and their potential side effects [9–28]. Our first meta-analysis in 2002 [12] showed that besides reducing the risk and the need for red blood cell transfusions, ESAs might also improve overall survival (see fig. 1). The result was mainly driven by one large study conducted by Littlewood and colleagues in 2001 [29]. The study produced some evidence that patients with solid and haematological malignancies who received ESAs experienced improved overall survival compared with controls. The p-value of the trial was of borderline significance (p = 0.052). However, the endpoint of overall survival was introduced as a protocol amendment after the study had been started. Including adjusted data from this trial and data from another 18 trials with a total of 2,805 patients, our meta-analysis showed some evidence for improved overall survival in patients receiving ESAs (hazard ratio (HR) 0.81, 95% confidence interval (CI) 0.67–0.99 using adjusted survival estimates and HR 0.84, 95% CI 0.69–1.02 using unadjusted survival estimates from Littlewood et al. [29]). In other words, when we pooled the results from individual studies for our first meta-analysis, we found evidence for an overall survival benefit for cancer patients receiving ESAs. Of note, there was no evidence of a publication bias (p-value for regression test = 0.481) (see fig. 2). This finding was supported by biological hypotheses. There was evidence from animal models and studies in humans that anaemia might result in a poorer response to chemotherapy or radiotherapy owing to increased tumour hypoxia [30–34]. Consequently, researchers hypothesised that targeting cancer-related anaemia might also improve tumour response and overall survival. However, already in 2003 the findings of our first meta-analysis were contradicted by two large randomised controlled trials that showed reduced survival for the ESA group in patients with metastatic breast cancer receiving chemotherapy [35] and patients with head and neck cancers receiving radiotherapy [36]. These unfortunate findings were followed by several other trials showing detrimental survival results for cancer patients receiving ESAs, as compared with controls [37–41]. When updating our meta-analysis on overall survival in 2006, these findings from single studies were consolidated but still uncertain: the HR for overall survival was 1.08 (95% CI 0.99–1.18, 42 trials and 8,167 patients), making the overall survival benefit we had previously reported questionable. In this situation we decided to retrieve the individual patient data (IPD) for all randomised controlled trials (RCTs) of the effects of ESAs compared with standard care in cancer patients. We invited independent investigators and pharmaceutical companies (Johnson & Johnson, Roche, Amgen) to participate in this project and to contribute their trial data. In 2008 we received the IPD as requested, as well as corresponding study protocols, protocol amendments and clinical study reports for 53 RCTs including 13,933 patients [28]. Based on these data and documents we conducted a meta-analysis on the effects of ESAs on on-study mortality and overall survival, which was published in 2009 [28]. The use of IPD allowed us to overcome certain limitations of literature-based meta-analyses, such as the use of aggregated data, heterogeneous endpoints across studies and the inability to assess differences across subgroups. For this update we differentiated mortality during the active study period from overall survival, defined as deaths during the longest follow-up available. This update confirmed the previously noted trends: ESAs increased on-study mortality in cancer by a factor of 1.17 (HR 1.17, 95% CI 1.06–1.30) and there was some evidence for reduced overall survival (HR 1.06, 95% CI 1.00–1.12).
A hypothetical study

Introduction
To determine whether the results of our first meta-analysis [12] had been affected by publication and reporting biases, and if timely access to clinical study reports and IPD could have prevented this, we conducted a hypothetical meta-analysis for overall survival and on-study mortality including studies and study data which could have been available in 2002, at the time when we conducted our first meta-analysis.

Methods
In this hypothetical meta-analysis, we included studies that recruited patients until the end of 1999. We chose this timeframe because it closely corresponds to the studies we included in the first, literature-based meta-analysis, that is, studies published up to the beginning of 2002. We identified eligible studies on the basis of the meta-analyses conducted previously [12, 14, 28]. In these meta-analyses [12, 14, 28] we included randomised controlled trials comparing ESAs with placebo or best usual care in cancer patients who were anaemic or at risk of developing cancer-related anaemia, and who were or were not receiving anticancer treatment. We conducted systematic searches using electronic databases (Medline, Embase, Cochrane Library and databases of conference proceedings), as well as documents from Oncology Drug Advisory Committee (ODAC) hearings [42]. For the first, literature-based review, we also obtained unreported data from the investigators. For the IPD review we obtained raw study data, study protocols, their amendments and clinical study reports from manufacturers and independent investigators. For the literature-based meta-analyses we used overall survival defined as the longest follow up available. For the IPD meta-analysis we used overall survival and in addition on-study mortality defined as death from any cause during the active study period plus 30 days. Details of these meta-analyses are reported elsewhere [12, 14, 28].

For the hypothetical meta-analysis we extracted overall survival and on-study mortality data from these three meta-analyses [12, 14, 28]. Whenever we had more than one result for a given study the following hierarchy applied: (1.) data and results taken from IPD review; (2.) data and results provided by study authors for previous meta-analyses; (3.) data and results as reported in the literature. We compared results from the different reviews for a given study, discrepant data and results were noted and reasons for discrepancies investigated and reported. Overall survival and on-study mortality was assessed as time-to-event analysis using hazard ratios. We calculated effect estimates and 95% confidence intervals (CIs) using fixed effects models.

Results
Our first, literature-based meta-analysis evaluated 19 trials including 2,805 patients for overall survival [12]. We obtained overall survival data either from the literature [29, 43–45] or from study investigators [46–60]. We found a HR of 0.81 (95% CI 0.67–0.99, figs 1 and 2), favouring ESAs. For the hypothetical meta-analysis we identified 25 eligible studies (3,664 patients in total; 2,155 randomised to ESA; 1,509 randomised to control), which recruited patients up to the end of 1999. Of these, 19 trials had been included in the overall survival analysis of the first meta-analysis. Another six studies were not included in the first meta-analysis for the following reasons: One study [61] had been included in the first review [12], however, since survival data were not reported or provided by the study authors, the study was not included in the overall survival analysis. Another study [62] was published after 2002 and was therefore not identified for the first review [12]. Four studies [63–66] were completed at the time of the first review [12], but survival or any other study data had not yet been published. The survival results for these studies became available through ODAC reports published in 2003 [67] and the IPD meta-analysis published in 2009 [28]. Five of these six studies showed detrimental findings, that is, worse survival in patients receiving ESAs than in controls. For the 19 studies which had already been included in the first review [12] we compared the results reported in

---

**Figure 3**
Forest plot for overall survival based on data from our first meta-analysis [12] adding missing studies and using results from individual patient data review [27, 28].

**Figure 4**
Funnel plot for overall survival based on data from our first meta-analysis [12] adding missing studies and using results from individual patient data review [27, 28].
the different sources. For the six studies [43, 45, 49, 54, 55, 58], no additional data were retrieved for the IPD review because the studies were too small for inclusion [27, 28]. The results for these six studies [43, 45, 49, 54, 55, 58] are therefore identical in the first review and the hypothetical meta-analysis. For 13 studies [29, 44, 46–48, 50–53, 56, 57, 59, 60] we had results stemming from the first review [12] and the IPD meta-analysis [28]. For three studies the number of patients randomised and the number of events were identical and there were only minor differences in effect estimates (HRs or CIs) [29, 46, 48]. For the other ten studies we noted discrepant results, that is, the data reported in the literature or provided by the study authors for the first review [12] were not identical to the overall survival estimates generated in the IPD review [28]. Discrepancies concerned the number of patients randomised per study arm, the number of deaths per study arm and the reported HRs plus CIs, for details see attached data table 1. Differences were explained by differing definitions of end of study in three studies [47, 50, 52], longer follow up or updated study information available for IPD for four studies [44, 57, 59, 60]. For two studies IPD analysis was based on intention-to-treat whereas previously included results were based on a safety population including only patients who had received at least one dose of the drug [53, 56]. For another study differences are explained by use of different inclusion criteria in the first review and the IPD review [51]. The results obtained for the IPD review [28] were worse (a HR which was less favourable than the previously reported HR) in seven studies [44, 47, 50, 51, 56, 57, 59] and better (a HR which was more favourable compared to the previously reported HR) in three studies [52, 53, 60]. For the hypothetical meta-analysis we included the available data stepwise into the original review conducted in 2002. Firstly, we added published data for “late” publications [62] and IPD for unpublished outcomes from published [61, 63] and unpublished studies [64–66]. This resulted in the addition of six more studies, which means that in 2002 24% of the available studies were missing. Five out of these six studies were sponsored by pharmaceutical companies. By including these missing studies the hazard ratio changed from 0.81 (95% CI 0.67–0.99) to 0.88 (95% CI 0.73–1.05). Moving one step further, we replaced the published data of our first analysis [12] with IPD [27, 28]. This resulted in replacing data from 13 studies (or from 52% of all the included studies). The new HR was 0.97 (95% CI 0.83–1.12, figs 3 and 4), showing no overall survival benefit for patients receiving ESAs compared to controls. In a sensitivity analysis we conducted a similar analysis for on-study mortality and found similar results (data on file).

Discussion

Compared with our original meta-analysis, which suggested an overall survival benefit for cancer patients receiving ESAs (HR 0.81, 95% CI 0.67–0.99) [12], our hypothetical meta-analyses based on the results of all studies conducted at the time of the first analysis did not show evidence for a beneficial effect of ESAs on overall survival (HR 0.97; 95% CI 0.83–1.12). This suggests that our first meta-analysis [12] showed misleading overall survival benefit owing to publication and reporting biases, which could have been prevented by timely access to IPD and clinical study reports. Based on a series of meta-analyses with access to published reports and IPD, our study demonstrates the impact of publication and reporting biases on the perceived effects of ESAs on survival in cancer patients. We were able to show this effect by using a large number of studies. Compared with our original analysis we increased the number of studies by 25% and replaced 52% of the data with results generated in an IPD meta-analysis [27, 28]. Our case study, however, has limitations. For six studies [43, 45, 49, 54, 55, 58] we did not have IPD; however, these studies were small and a large effect on the overall results is unlikely. We investigated overall survival and on-study mortality, and we do not know whether the publication and reporting biases would affect other outcomes in a similar way. However, we recently updated several clinical outcomes for our Cochrane Review [14, 68] based on the published literature. With few exceptions study outcomes were underreported leading to overestimates of both beneficial and harmful effects of ESAs [68].

How to address publication and reporting biases?

Our case study adds another example to the large body of evidence on publication and reporting biases [2–4, 8, 69–71]. Funnel plots offer a way to explore the probability of publication bias; however, they only demonstrate but do not solve potential publication biases. They can also produce false negative results. For example, the funnel plot generated for our first meta-analysis did not indicate a publication bias. In 2004, the International Committee of Medical Journal Editors produced a statement requiring trials to have been registered in a registry before being considered for publication [72]. Two years after its implementation, more trials were registered than in the past [73]. However, this measure did not help to reduce selective reporting of outcomes leading to reporting bias, since this statement did not require reporting of study results. A study by Huic et al. [74] assessed about 150 reports from RCTs and found that almost 40% of the assessed trials had discrepancies between their registered and reported primary outcomes, and almost 65% had discrepancies in the secondary outcomes. Most of these discrepancies were due to post-hoc reporting of outcomes in the publications that had not been listed in the registry. There is some evidence that, specifically for adverse events, unpublished data might allow a more accurate estimate than the published data [75]. This kind of discrepancy has led to the recognition that not only registration of the trial but also complete reporting of study results needs to be ensured. The US FDA Amendments Act of 2007 requires the publication of the main trial results in a standardised summary report [76–79]. However, it does not require the publication of the entire clinical study report or the release of the IPD. The European Medicines Agency (EMA), recognising that the current state of disclosure of important study documents is no longer desirable [80], has changed its policy on document access [81]. With the new policy the EMA will release documents upon request provided that the EMA’s authorisation procedures on a given drug have been finalised.
There is an on-going debate as to whether the release of clinical study reports including tabulated listings of anonymised IPD is sufficient, or whether IPD in electronic databases suitable for re-analysis of the data are needed [80, 82]. Access to IPD permits researchers to define uniform study endpoints across studies, to conduct intention-to-treat analyses standardised across trials and to include additional information gained through longer follow up [83]. Access to well-reported clinical study protocols and reports in addition to the IPD are essential to fully understand the nature and quality of the IPD [84, 85]. Whether, in contrast, access to clinical study reports providing summary estimates would be sufficient to allow an adequate evaluation of the drug remains questionable and has to our best knowledge not yet been investigated.

### Is meta-analysis losing its credibility?

In a recently published opinion paper in Nature Medicine the authors pointed out that with a lack of robust evidence and a reasonable minimum amount of data, meta-analyses could lose credibility [86]. We agree that robust data are needed to generate meta-analyses that help to make well-informed decisions on healthcare interventions. We also agree that meta-analyses based on biased data sets can produce distorted and misleading results. However, we would like to emphasise that a lack of robust data, either due to a lack of high quality studies or due to reporting and publication biases, is not a problem of meta-analysis per se; it is a problem of healthcare and medical research. Meta-analyses together with funnel plots and regression tests are instruments that allow the assessment of the quality of the studies and detection of publication or reporting biases, actually making the absence of studies and study results visible. If the published data are so incomplete that we cannot use them for a meta-analysis how can we possibly use them?

### Table 1: Comparison of study data used for our first meta-analysis [87] on the effects of erythropoiesis-stimulating agents on cancer (conducted in 2002) and a hypothetical meta-analysis including all data which could have been available at the time of the first meta-analysis.

<table>
<thead>
<tr>
<th>Author, year of publication</th>
<th>ESA group n/N</th>
<th>Controls n/N</th>
<th>Hazard ratio [95% Confidence interval]</th>
<th>ESA group n/N</th>
<th>Controls n/N</th>
<th>Hazard ratio [95% confidence interval]</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abels 1993</td>
<td>13/65</td>
<td>13/59</td>
<td>0.89 [0.37, 2.10]</td>
<td>13/65</td>
<td>12/59</td>
<td>0.96 [0.44, 2.10]</td>
<td>Different end of study definition</td>
</tr>
<tr>
<td>Barnias 2003</td>
<td>NA</td>
<td>NA</td>
<td>7/72 [4.72]</td>
<td>16/132</td>
<td>12/127</td>
<td>1.53 [0.72, 3.26]</td>
<td>For IPD longer follow-up available</td>
</tr>
<tr>
<td>Boogaerts 2003</td>
<td>8/133</td>
<td>8/129</td>
<td>1.02 [0.38, 2.72]</td>
<td>10/81</td>
<td>9/76</td>
<td>1.11 [0.45, 2.73]</td>
<td>Number of patients and events identical, minor differences in HR and CI</td>
</tr>
<tr>
<td>Cascinu 1994</td>
<td>0/50</td>
<td>0/50</td>
<td>Could not be estimated</td>
<td>0/50</td>
<td>0/50</td>
<td>Not estimable</td>
<td>Same data</td>
</tr>
<tr>
<td>Case 1993</td>
<td>10/81</td>
<td>9/76</td>
<td>1.05 [0.40, 2.73]</td>
<td>10/81</td>
<td>9/76</td>
<td>1.11 [0.45, 2.73]</td>
<td>Number of patients and events identical, minor differences in HR and CI</td>
</tr>
<tr>
<td>Cazzola 1995</td>
<td>2/57</td>
<td>1/29</td>
<td>0.06 [0.00, 5.35]</td>
<td>5/114</td>
<td>3/29</td>
<td>0.69 [0.13, 3.57]</td>
<td>IPD includes 2 study groups with low dose ESA, which had been excluded from the 2002 review</td>
</tr>
<tr>
<td>Dammacco 2001</td>
<td>1/69</td>
<td>7/76</td>
<td>0.23 [0.05, 0.94]</td>
<td>1/69</td>
<td>7/76</td>
<td>0.23 [0.06, 0.90]</td>
<td>Number of patients identical, minor differences in CI</td>
</tr>
<tr>
<td>Del Mastro 1997</td>
<td>1/31</td>
<td>3/31</td>
<td>0.36 [0.05, 2.53]</td>
<td>1/31</td>
<td>3/31</td>
<td>0.36 [0.05, 2.53]</td>
<td>Same data</td>
</tr>
<tr>
<td>Dupny 1999</td>
<td>0/15</td>
<td>1/15</td>
<td>0.14 [0.00, 6.82]</td>
<td>0/15</td>
<td>1/15</td>
<td>0.14 [0.00, 6.82]</td>
<td>Same data</td>
</tr>
<tr>
<td>EPO-INT-3</td>
<td>–</td>
<td>–</td>
<td>9/135 [3/65]</td>
<td>7/67</td>
<td>10/65</td>
<td>0.68 [0.26, 1.77]</td>
<td>Different end of study definition</td>
</tr>
<tr>
<td>Henry 1995</td>
<td>8/67</td>
<td>10/65</td>
<td>0.75 [0.28, 2.01]</td>
<td>8/48</td>
<td>8/47</td>
<td>1.19 [0.44, 3.21]</td>
<td>Not available in 2002</td>
</tr>
<tr>
<td>Kurz 1997</td>
<td>0/23</td>
<td>0/12</td>
<td>Could not be estimated</td>
<td>0/23</td>
<td>0/12</td>
<td>Not estimable</td>
<td>Same data</td>
</tr>
<tr>
<td>Littlewood 2001</td>
<td>155/251</td>
<td>82/124</td>
<td>0.81 [0.62, 1.06] *</td>
<td>155/251</td>
<td>82/124</td>
<td>0.80 [0.61, 1.05]</td>
<td>Number of patients identical, minor differences in HR and CI</td>
</tr>
</tbody>
</table>

P value: 0.76 [0.58, 1.00]**

IPD based on ITT, published data based on safety

IPD based on ITT, published data based on safety

Longer follow-up; IPD based on ITT, published data based on safety

Not available in 2002

Not available in 2002

Not available in 2002

Different end of study definition

Different end of study definition

For IPD longer follow-up available

IPD contains updated study data

Same data

Same data

ESAs = erythropoiesis-stimulating agents; n = number of events; N = number of participants; * = unadjusted calculated based on p-value from Kaplan Meier estimate published in Littlewood 2001 [89]; ** adjusted HR as reported in Littlewood 2001. NR = not reported; NA = not available; study not published in 2002; IPD = individual patient data; HR = hazard ratio; CI = confidence interval; bold font indicates results which differed between the first review and the IPD review.
to make healthcare decisions for patients, with or without meta-analysis? Our study is yet another indication that publication and reporting biases are main obstacles to the efficient use of healthcare interventions. We strongly believe that these problems cannot be eliminated by the registration of trials and the standardised but limited reporting of trial results. Unrestricted access to clinical study protocols including amendments, to clinical study reports as well as to IPD is needed to ensure timely detection of both beneficial and harmful effects of health care interventions.

Funding / potential competing interests: No financial support and no other potential conflict of interest relevant to this article was reported.

Correspondence: Thomy Tonia, MSc, University of Bern, Institute of Social and Preventive Medicine, Finkenhubelweg 11, CH-3012 Bern, Switzerland, ttonia[at]isp.unibe.ch

References

Review article: Current opinion


78 Godlee F, Loder E. Missing clinical trial data: setting the record straight. BMJ. 2010;341:c5641.


Figures (large format)

Figure 1
Forest plot for overall survival based on data from our first meta-analysis [12].

Figure 2
Funnel plot for overall survival based on data from our first meta-analysis [12].
Figure 3
Forest plot for overall survival based on data from our first meta-analysis [12] adding missing studies and using results from individual patient data review [27, 28].
Figure 4
Funnel plot for overall survival based on data from our first meta-analysis [12] adding missing studies and using results from individual patient data review [27, 28].