Publication bias in reproductive research*

Johannes L.H.Evers
Research Institute GROW, Department of Obstetrics and
Gynaecology, Academisch Ziekenhuis Maastricht, and Maastricht
University, 6202 AZ Maastricht, The Netherlands.
E-mail: jev@sgyn.azm.nl

Publication bias is defined as any tendency on the part of
investigators or editors to fail to publish study results on
the basis of the direction or strength of the findings. This
may lead to overestimation of treatment effects in published
work. Inappropriate decisions about patient management
may result. We investigated what proportion of abstracts
at the European Society of Human Reproduction and
Embryology (ESHRE) annual meeting eventually reached
full publication, what was the time to publication, and
which factors might have affected publication. Among the
2691 abstracts of six ESHRE annual meetings, 151 (5.6%)
reporting randomized controlled trials (RCT) were
identified. Comprehensive searches of electronic databases
and handsearching of the two major journals in the field
yielded 79 full publications pertaining to these abstracts.
Kaplan–Meier analysis estimated 56% of RCT abstracts
to be eventually published in full, the median time to
publication being 32.5 months. Positive outcome (i.e. signi-
ficant results) did not affect the publication rate, and
neither did sample size, the subject category, or the native
language (English/non-English) of the country of origin.
Oral presentations resulted in eventual full publication
significantly more frequently (69%) than posters (42%). It
is concluded that a considerable publication deficit, but
not a publication bias, exists for RCT in reproductive
research.

Key words: abstract follow-up/abstract publication/ESHRE
annual meeting/publication bias/publication deficit

Introduction

Reports of properly conducted randomized controlled trials
(RCT) are the foundation of effective health care, but it has
been suggested that many are either not submitted or not
accepted for publication (Easterbrook et al., 1991). ‘Publication
deficit’ is defined as the difference between the proportion
eventually reaching full publication and 100%. ‘Publication
bias’ is the term used for statistically significant results being
more likely to reach publication than work with non-significant
results. Publication bias is defined as any tendency on the
parts of investigators or editors to fail to publish study results
on the basis of the direction or strength of the study findings
(Dickersin and Min, 1993). There is now considerable evidence
(Dickersin and Min, 1993) for a reporting bias throughout
the various means available for reporting research results:
acceptance of abstracts for presentation at meetings; subsequent
full publication of studies initially presented as abstracts; full
publication of initiated studies; and reporting of published
results by the lay press. If a trial fails to show any treatment
difference, the organizers may be inclined to lose interest in
writing up their results, or even in completing the study. Also,
even if they do produce a manuscript showing their negative
findings, journal editors may fail to publish, considering such
information less interesting for their readership. After all,
chances are only small that the paper will lead to any notable
medical progress. If a negative result trial finally does get
published, it is likely to be in a small specialist journal rather
than in one of the high profile journals (Pocock, 1983).

This overexposure of positive result trials in leading journals
and the hidden publication of negative result trials in low
profile journals (or the failure to get them published at all)
may lead to an overestimation of treatment effects in published
work. Inappropriate decisions about patient management may
result. In this age of rigid and meticulous meta-analysis, basing
conclusions on published work only may be misleading. The
problem is compounded by the fact that the results from
meta-analysis give the impression of being very precise and
convincing, and therefore are having a steadily increasing
impact on day-to-day clinical practice, on healthcare policy
making, and on planning of future research. For this reason,
the editors of over 100 medical journals around the world
have declared an ‘amnesty’ for unpublished RCT (Smith and
Roberts, 1997). The reason for this remarkable initiative was
that they recognized the serious implications of not reporting
negative result trials. They stress that the above-mentioned
undeserved overexposure of trials that show more promising
effects may prompt misleading conclusions about effectivenss.
Patients may thus be exposed to useless or even harmful
treatments. They also indicate that publication bias may reduce
the power of systematic reviews to detect moderate but

*A preliminary report of this work was presented at the 14th ESHRE
Annual Meeting in Gothenburg 1998.
clinically important treatment effects. Patients may thus be
denied effective forms of health care. Finally, patients may be
asked to participate in new studies designed to address ques-
tions that have already been answered (Savulescu et al., 1996).

Estimating how many studies are ‘out there’ that never
reach publication is next to impossible. One way to gain some
insight into the problem of publication bias in reproductive
research is to determine what proportion of abstracts of RCT
presented at the ESHRE annual meetings eventually will be
published as full-length articles. One might expect that RCT
would have a higher rate of publication than studies that used
other designs, simply because of the time and effort required
on the part of the investigators and participants (Scherer et al.,
1994). For the same reason, it does not seem likely that many
abstracts of RCT, once submitted, would be rejected for
presentation at a major meeting. Still, the present study can
only give an impression of part of the publication bias, since
the decision to submit an abstract for presentation at a con-
ference would be prone (but perhaps to a lesser degree) to
the same selection bias that is intended to be explored here.

Materials and methods

Before embarking upon this study, we estimated the proportion of
abstracts of RCT presented at the ESHRE annual meetings to be 5%
of all submissions. Furthermore, we decided that we would consider
publication bias to exist if positive result studies would reach a >25%
higher full publication rate than negative result studies, and that
the number of studies in our analysis should give us 80% power to detect
this difference in full publication rate of at least 25% in favour of
studies with statistically significant results compared to those without,
at the α = 0.05 level. If half of the RCT showed significant results
we would need 138 abstracts of RCT, or 2760 abstracts submitted in
total. We estimated that we should allow at least a 3 year follow-up
period for full-length publications to follow (oral or poster) presenta-
tion of the abstract.

The abstract books of the ESHRE annual meetings 1992 in The
Hague, 1993 in Thessaloniki, 1994 in Brussels, 1995 in Hamburg,
1996 in Maastricht, and 1997 in Edinburgh together contained 2691
abstracts of oral and poster presentations. These abstracts were
published in special supplements to Human Reproduction. A total
number of 151 abstracts of RCT was identified by handsearching the
six abstract books. Computerized free text word searching of the two
abstract volumes available on disk (i.e. from the annual meetings of
1996 and 1997) did not yield any additional studies. Abstracts were
selected for inclusion in this study if the abstract stated that the
results were from an RCT involving humans. Of the 151 abstracts
identified, six appeared to have been published before—but in the
same year as—the respective annual meeting already. These were
included in the first month of follow-up.

Sample size, outcome of the study (significant/non-significant
result) and presentation format (oral/poster) were recorded, as were
the primary language of the country of origin, and the category of
research the abstract belonged to: assisted reproductive technology
laboratory (n = 26), assisted reproductive technology clinics (n =
34), drug trials (n = 53), and the remaining abstracts, mostly
dealing with the general field of (in)fertility (n = 38). Abstracts were
classified as ‘significant’ if a P-value for any result was < 0.05, or
if results were stated to be significant. Sample size (number of
participants in the smallest arm of the study) was dichotomized as
being below or above the median (n = 31) for all included studies.

Results

The Cochrane database was checked and Medline and Embase
searches were performed (starting with those found in the same year
the abstract was presented and going forward to February 2000) to
study whether the abstract was followed by a formal publication as
a full-length article in a scientific journal. Publications for each
author of an individual abstract were examined, and if a full-length
publication was found that corresponded in content to the subject
matter and that had as authors a majority of the original abstract’s
authors, it was then assumed that the results of the respective abstract
had been published in full. Each abstract was coded as being published
in full or unpublished. In addition, a hand search was performed for
the same period of all issues of the two major journals in our field,
Human Reproduction and Fertility & Sterility. This revealed no
additional, as yet undiscovered publications.

Kaplan–Meier survival analysis was used to investigate the time
to publication to enable inclusion of abstracts where time since
presentation was insufficient for publication to have occurred. In
survival analysis, a study may be included for as long as it is
observed, and after that time (February 2000) it will be censored.
In this way each study contributes to estimating the eventual publication
rate for as long as it can be followed, and no information will be
lost. This offered the best estimate of the eventual publication rate
(Cheng et al., 1998). A Kaplan–Meier survival curve of all 151
reports was generated, the event being publication. Peto’s log-rank
was used to test for comparing two or more survival curves. This
test does not make any assumptions about the distributions of the
survival estimates that comprise the curves. The null hypothesis
that the risk of publication was the same in all groups was tested.

Of the 151 abstracts of RCT identified, 79 eventually appeared
as a formal publication, representing a crude publication rate of
52%. Of these, 38 (48%) were published in one of the
Human Reproduction journals, 17 (22%) in Fertility & Sterility,
and 24 (30%) in 19 other, different journals. The estimated
eventual full publication rates (Kaplan–Meier survival analysis)
are shown in Table I. The time to publication (cumulative
publication rate) is illustrated in Figure 1. The estimated
median time to publication was 32.5 months (range 0–79
months), the median sample size (smallest arm) was 31 (range
3–641). Survival estimates suggested that the propor-
tion published in the first year after the annual meeting is
17% [95% confidence interval (CI) 12–24%], at 2 years 35%
(18–43%), at 3 years 48% (40–57%). At 5 years, the full
publication rate was estimated at 53% (45–61%). Of 69
abstracts showing a significant outcome, 41 reached full
publication (59%), as compared to 38 of 82 without a significant
outcome (46%). The difference was not significant [P = 0.15;
odds ratio (OR) 1.70, 95% CI 0.89–3.24].

Log-rank testing showed no significant differences in
publication rate between studies with and studies without
significant results (χ^2 for equivalence of publication rates
= 3.0597; P = 0.08). nor between studies from English-speaking
countries and from non-English speaking countries (χ^2
= 1.573; P = 0.69), and neither between studies in which the
sample size of the smallest group was below the median as
compared to above the median (χ^2 = 0.0012; P = 0.97).
Also, no significant difference was found between the four
categories of abstracts recognized, laboratory assisted
Table 1. Estimated full publication rates (Kaplan–Meier survival analysis) of abstracts presented between 1992 and 1997 at ESHRE annual meetings [estimate of percentage full publication (%) and 95% confidence intervals (95% CI) of estimates are given]

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n</th>
<th>%</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>All abstracts</td>
<td>151</td>
<td>56</td>
<td>47–65</td>
</tr>
<tr>
<td>Outcome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significant</td>
<td>69</td>
<td>63</td>
<td>51–75</td>
</tr>
<tr>
<td>Non-significant</td>
<td>82</td>
<td>50</td>
<td>39–63</td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;median</td>
<td>75</td>
<td>53</td>
<td>42–65</td>
</tr>
<tr>
<td>&gt;median</td>
<td>76</td>
<td>60</td>
<td>46–75</td>
</tr>
<tr>
<td>Presentation*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral</td>
<td>72</td>
<td>69</td>
<td>58–80</td>
</tr>
<tr>
<td>Poster</td>
<td>79</td>
<td>42</td>
<td>31–56</td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>40</td>
<td>54</td>
<td>39–70</td>
</tr>
<tr>
<td>Non-English</td>
<td>111</td>
<td>56</td>
<td>46–67</td>
</tr>
<tr>
<td>Subject category</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ART laboratory</td>
<td>26</td>
<td>60</td>
<td>44–76</td>
</tr>
<tr>
<td>ART clinics</td>
<td>34</td>
<td>49</td>
<td>32–70</td>
</tr>
<tr>
<td>Drug trial</td>
<td>53</td>
<td>53</td>
<td>40–68</td>
</tr>
<tr>
<td>General (infertility)</td>
<td>38</td>
<td>60</td>
<td>42–80</td>
</tr>
</tbody>
</table>

*P = 0.0008 for difference between full publication rate of abstracts of oral versus poster presentation (Peto’s log-rank test). ART = assisted reproductive technology.

Discussion

Of all abstracts of RCT presented at six consecutive ESHRE annual meetings (1992–1997), 56% are estimated to eventually reach full-length publication in a scientific journal, a large proportion of which (48%) is expected to appear within 3 years. This is in agreement with the findings of a meta-analysis (Scherer et al., 1994) which found a 51% full publication rate (95% CI 45–57%) of abstracts presented at 11 surgical, cardiological, anaesthesiological, paediatric, perinatological, oncological and ophthalmological meetings. RCT accepted for oral presentation have a significantly greater chance of getting published (especially in high-profile journals), and do so significantly earlier than poster presentations. This can be explained by higher quality research being reflected in higher quality abstracts, which in turn qualify preferentially for oral presentation at the ESHRE annual meetings and by the same token for publication in the more prestigious journals. The outcome of the RCT (i.e. significant versus non-significant), nor its sample size or subject category affected publication or time to publication, and neither did the native language in the country of origin. When considering the results of the present study, it should be kept in mind that the decision to submit an abstract for presentation at a conference would be prone to the same selection bias that may be operative in determining whether to put the findings of a study onto paper and submit them in the form of a full manuscript to a scientific journal. Therefore, it is most likely there are even more unpublished RCT ‘out there’ than have been discovered by the present study.

In conclusion, we found a considerable publication deficit (44%), but not a publication bias for RCT in reproductive technology, clinical assisted reproductive technology, drug trials, and others ($\chi^2$ for equivalence of publication rates = 0.2046; $P = 0.9768$; $\chi^2$ for trend = 0.2382; $P = 0.6255$). A significant difference was found between eventual publication rates of abstracts accepted for oral as compared to poster presentation ($\chi^2 = 11.1579; P = 0.0008$): 69% of oral presentations eventually were estimated to reach full-length publication (95% CI 58–80%) as compared to only 42% of posters (95% CI 31–56%). Of the oral presentations reaching full-length publication, 52% (95% CI 42–62%) were published in a high profile journal (upper 25% quartile of Science Citation Index journal citation score rankings), which was significantly more (P < 0.05) than for poster presentations (29%; 95% CI 27–32%).

The estimated median times to publication (median survival times) did not differ significantly between significant result studies (29 months) and non-significant result studies (79 months), nor between abstracts from English-speaking (34 months) or non-English-speaking countries (49 months), and neither between small (37 months) and large studies (50 months). Also, the four subject categories of abstracts had comparable median times to publication: laboratory assisted reproductive technology (30 months), clinical assisted reproductive technology (>34 months), drug trials (49 months), others (25 months). Abstracts accepted for oral presentation reached eventual full-length publication significantly (P < 0.05) sooner (26 months) than poster abstracts (>79 months). (Whenever the > sign is used, an accurate estimate of the median time to publication could not be made, since the estimated full publication rate in the respective category was <50%).

Figure 1. Cumulative publication estimate of abstracts of randomized controlled trials presented at the ESHRE annual meetings of 1992–1997. □ median publication estimate, △ lower limit 95% confidence interval (CI), × upper limit 95% CI; Kaplan–Meier survival analysis.
research. For abstracts presented at an ESHRE annual meeting, quality rather than outcome tends to determine subsequent formal publication as a full-length paper in a scientific journal.

References

Received on February 17, 2000; accepted on June 15, 2000