Language bias in randomised controlled trials published in English and German

Matthias Egger, Tanja Zellweger-Zähner, Martin Schneider, Christoph Junker, Christian Lengeler, Gerd Antes

Summary

Background Some randomised controlled trials (RCTs) done in German-speaking Europe are published in international English-language journals and others in national German-language journals. We assessed whether authors are more likely to report trials with statistically significant results in English than in German.

Methods We studied pairs of RCT reports, matched for first author and time of publication, with one report published in German and the other in English. Pairs were identified from reports found in a manual search of five leading German-language journals and from reports published by the same authors in English found on Medline. Quality of methods and reporting were assessed with two different scales by two investigators who were unaware of authors' identities, affiliations, and other characteristics of trial reports. Main study endpoints were selected by two investigators who were unaware of trial results. Our main outcome was the number of pairs of studies in which the levels of significance (shown by p values) were discordant.

Findings 62 eligible pairs of reports were identified but 19 (31%) were excluded because they were duplicate publications. A further three pairs (5%) were excluded because no p values were given. The remaining 40 pairs were analysed. Design characteristics and quality features were similar for reports in both languages. Only 35% of German-language articles, compared with 62% of English-language articles, reported significant ($p<0·05$) differences in the main endpoint between study and control groups ($p=0·002$ by McNemar's test). Logistic regression showed that the only characteristic that predicted publication in an English-language journal was a significant result. The odds ratio for publication of trials with significant results in English was 3·75 (95% CI 1·25–11·3).

Interpretation Authors were more likely to publish RCTs in an English-language journal if the results were statistically significant. English language bias may, therefore, be introduced in reviews and meta-analyses if they include only trials reported in English. The effort of the Cochrane Collaboration to identify as many controlled trials as possible, through the manual search of many medical journals published in different languages will help to reduce such bias.

Lancet 1997; **350:** 326–29

Department of Social Medicine, University of Bristol, Canynge Hall, Bristol BS8 2PR, UK (M Egger MD); **Department of Social and Preventive Medicine, University of Bern, Switzerland** (T Zellweger-Zähner Med Pract, M Schneider MD, C Junker MD); **Department of Public Health and Epidemiology, Swiss Tropical Institute, Basle** (C Lengeler PhD); **and Department of Medical Biometry and Medical Informatics, University of Freiburg, Germany** (G Antes PhD)

Correspondence to: Dr Matthias Egger

Introduction

Systematic review of evidence about the benefits and risks of medical interventions can influence decision-making in clinical practice and public-health medicine, identify areas in which further research is needed, and guide allocation of resources.[1,2] The dissemination of medical evidence, including the publication of results from randomised controlled trials (RCTs), is influenced by several factors, however, that modify the probability of whether a trial is included in a meta-analysis. Publication bias—selective publication of significant findings and the non-publication of those without such findings—has been documented repeatedly.[3–5] (The word "significant" relates to statistical significance here and in the rest of the article.) Consequently, only biased samples of all the existing evidence are likely to be publicly available.

Several factors influence the probability of whether a published study is included in a systematic review.[6–8] One factor that has received little attention is the language in which a paper is published.[9] Investigators working in non-English-speaking countries publish some of their work in national journals. Authors may be more likely to report in an international, English-language journal results that are significant, whereas other findings are more likely to be published in local journals. English language bias could, therefore, be introduced in systematic reviews and meta-analyses that are based exclusively on reports written in English. Although concern has been expressed about this type of bias,[9–11] its significance in meta-analytic research is unclear at present.

In an attempt to identify all published controlled trials, the Cochrane Collaboration has embarked on an extensive manual search of many medical journals published in languages other than English.[12] We manually searched through five leading German-language general-medicine journals[13] and, at the same time, did a bibliographical study. Our objectives were to describe publication trends and quality features of RCTs done in German-speaking Europe and to assess whether trials with significant results are more likely to be published in an international English-language journal than trials without significant findings.

Methods

Five leading general-medicine journals published in German-speaking Europe were searched manually for RCTs. We searched thoroughly each issue of *Deutsche Medizinische Wochenschrift, Schweizerische Medizinische Wochenschrift, Schweizerische Medizinische Rundschau (Praxis), Wiener Klinische Wochenschrift,* and *Wiener Medizinische Wochenschrift* published between 1985 and 1994. A trial was defined as an RCT if assignment of participants to treatment and control groups was described as randomised by words such as "randomly", "random", and "randomisation". All RCTs found were reported to the Cochrane Controlled Trials Register,[14] but only RCTs from institutions in Germany, Switzerland, and Austria, and published in German were included in our study.

To examine trends in the language of publication, up to three Medline searches (Datastar, Knight Ridder Information Services, Berne, Switzerland) were done for each German-language report that fulfilled the inclusion criteria. RCTs published in an English-language journal by the same key authors during the same period

were identified from Medline. Key authors were defined as first, second, and last authors. The key authors' names and first initials, followed by a "wild card" for additional initials, together with appropriate restrictions for publication period and language were used as key words for searches. Authors were assumed to be identical if names and initials, as well as specialty and country, were the same on both reports. The abstracts and, if necessary, the full text of the articles were read to identify them as RCTs.

All German-language reports with first authors who were also first authors of at least one trial published in an English-language journal were identified. Reports were matched for first authors and year of publication, if possible, with a random-number generator by an investigator (CJ), who was unaware of trial results and quality. If no English-language report was published in the same year, reports published 1 year or, if necessary, within several years of the German-language report were identified and one of those was randomly selected. Duplicate publications that reported on the same trial with the same number of patients and the same endpoints in both languages were excluded. Articles reporting different endpoints of the same trial, or reporting separately preliminary and final results, were included in the main analysis but excluded from sensitivity analyses.

For each trial a main endpoint was selected independently by two investigators (TZZ and ME). Trial results were concealed from these investigators, who were given only the introduction and material and method sections of each article. Discrepancies in the choice of endpoint were resolved by consensus. Once an endpoint had been selected, the full version of the article was read to find the p value for the difference in the main endpoint between the randomised groups. Our main outcome was the number of discordant pairs of studies in which one paper had a significant and the other a non-significant p value, or one article had a lower p value than the other.

We used the quality scores described by Chalmers and colleagues[15] and Jadad and colleagues[16] to assess trial quality. The Jadad scale emphasises the quality of the reporting, whereas the Chalmers scale focuses on the quality of methods.[17] Because there is evidence that masked assessment produces more consistent scores,[17] authors' identities and affiliations, journal names, dates of publication, sources of financial support, and acknowledgments were omitted. Reports were then scored independently by two investigators (TZZ and MS). Discrepancies were resolved by discussion. For each article, the journal's impact factor for the year of publication was obtained from the *Journal Citation Reports*.[18]

The number of discordant pairs was analysed by McNemar's χ^2 test. Because exact p values were not generally reported, values were classified in four groups ($p \geq 0{\cdot}05$, not significant; $0{\cdot}01 \leq p < 0{\cdot}05$; $0{\cdot}001 \leq p < 0{\cdot}01$; and $p < 0{\cdot}001$). The *t* distribution was used to calculate 95% CIs for differences between quality scores. Finally, we used univariate and multivariate conditional logistic regression to identify factors associated with publication of reports in English. All analyses were done with the SAS software package, version 6.11, and EGRET software, version 0.26.

Results

255 trials were identified through the manual search. 32 (13%) reports were excluded (23 in French, five in English, two in German from Hungary, and one in German from the UK and Croatia). Therefore, 223 German-language reports published by 529 key authors fulfilled the inclusion criteria. Of these, the numbers of trials published each year decreased from 40 in 1985 to eight in 1994. The Medline search on the 529 key authors found 570 English-language reports. The average number of English-language reports published per key author for each year was 0·4 in 1985 and increased to 2·2 in 1994. Similarly, the average number of reports per key author and year published in either language increased from 0·8 in 1985 to 2·5 in 1994.

62 first authors had also been the first authors of English-language articles. 19 (31%) pairs had to be excluded because the only English-language reports available were duplicate publications of the same trial. A further three (5%) pairs were excluded because they gave no p values. The remaining 40 pairs were analysed.

	German language (n=40)	English language (n=40)
Design		
Parallel group	29 (72%)	29 (72%)
Crossover	11 (28%)	11 (28%)
Control intervention		
Placebo	13 (32%)	18 (45%)
Standard treatment	23 (58%)	19 (47%)
No treatment	4 (10%)	3 (8%)
Sample size		
Mean (SD)	63 (73)	59 (59)
Median (range)	43 (6–400)	40 (6–254)
Randomisation		
Central	4 (10%)	3 (8%)
Open	4 (10%)	4 (10%)
Not described	32 (80%)	33 (82%)
Masking		
Double-blind	18 (45%)	22 (55%)
Single-blind	2 (5%)	2 (5%)
Open	13 (32%)	14 (35%)
Not described	7 (18%)	2 (5%)
Withdrawals		
Adequate description*	25 (62%)	21 (52%)
Analysis by intention to treat	15 (38%)	15 (38%)
Quality (% of maximum score)		
Mean (SD) Jadad score	48·5 (21·7)	46·0 (20·4)
Mean (SD) Chalmers score	43·9 (18·3)	45·6 (18·6)

*Number of withdrawals by treatment group.

Table 1: **Characteristics of RCT pairs**

Among the German-language reports, 22 (55%) were published in *Deutsche Medizinische Wochenschrift*, seven (18%) in *Wiener Klinische Wochenschrift, five (13%) in Schweizerische Medizinische Wochenschrift,* four (10%) in *Wiener Medizinische Wochenschrift,* and two (5%) in *Schweizerische Medizinische Rundschau (Praxis)*. The English-language reports were published in 33 different journals, mostly internal medicine specialist journals. 11 (28%) were published in a journal supplement. 13 (33%) pairs of reports were about cardiovascular medicine, seven (18%) were about gastroenterology, and 16 (40%) were about various other internal medicine specialties. Two (5%) pairs were each about sports medicine and gynaecology. English-language reports were, on average, published 5 months later than German-language articles. 35 (88%) pairs consisted of two different trials. In three pairs, reports related to the same trials and endpoints, but the German-language report presented an interim analysis and the English-language report presented the final analysis. In one pair, papers reported on the same trial and participants but on different endpoints, and in another the same trial was reported with different endpoints and different groups of participants.

Table 1 shows design and quality features. There were no significant differences ($p > 0{\cdot}2$) between the two language groups. A parallel-group design was used in 72% of trials in both languages. Placebo-controlled trials were more frequent in the English-language sample, whereas control groups receiving standard treatments were more frequent among reports published in German. The median number of participants was around 40 in both groups. Results for quality were also similar. With both scales, the mean scores were about 45% of maximum in both language groups (mean difference German-English 2·5% [95% CI −5·8 to 10·8], Jadad score;[16] −1·7% [−8·9 to 5·5], Chalmers

p for main endpoint	German language (n=40)	English language (n=40)
p≥0·05	26 (65%)	15 (38%)
0·01≤p<0·05	8 (20%)	14 (38%)
0·01≤p<0·05	3 (8%)	4 (8%)
p<0·001	3 (8%)	7 (18%)

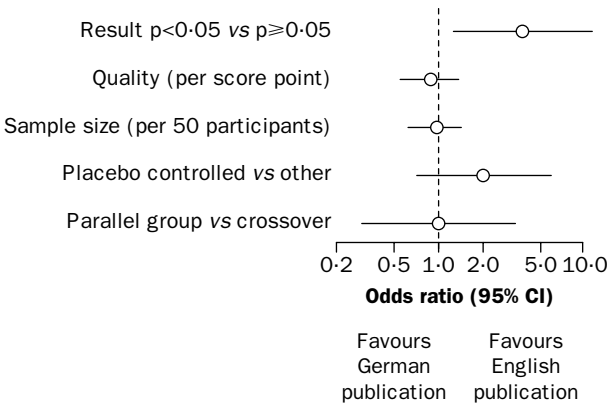
Table 2: **Distribution of p values among RCT pairs**

score[15]). Journal impact factors were available for 38 articles in either language. Median impact factors were 0·71 (range 0·08–1·29) and 1·23 (0·11–24·5) for reports published in German and English, respectively (p<0·0001).

Only 35% of German-language articles, but 62% of English-language articles reported a significant (p<0·05) difference between experimental and control groups in the main endpoint (table 2). 22 of the 27 discordant pairs showed a significant (or more significant) result for the English-language report and five pairs showed a significant (or more significant) result in the German-language report (p=0·002 by McNemar's test). This finding was confirmed by logistic regression. Because scores for quality were strongly correlated (r_s=0·47, p<0=0001), only the Jadad scale was initially included in these models. Univariate and multivariate analyses showed that a significant result was the only characteristic that predicted publication in an English-language journal. In univariate analysis, the odds ratio for publication in an English-language journal was 3·75 (95% CI 1·25–11·3) if the trial result was significant (figure). The association was slightly stronger when adjusted for reporting quality, sample size, use of a placebo, and use of the parallel design in a multivariate model (odds ratio 3·98 [1·20–13·2]). There was no evidence of statistical interactions between the variables included in this model.

Results were similar after exclusion of the five pairs of articles reporting on the same trial. 19 of 24 discordant pairs showed a significant (or more significant) result for the English-language report with five pairs showing the more significant result for the German-language report (p=0·008 by McNemar's test). After exclusion of 11 pairs that involved reports in journal supplements, there were 15 and four discordant pairs respectively (p=0·022 by McNemar's test). We included the Chalmers score instead of the Jadad score in multivariate logistic regression and found similar results (odds ratio 3·67 [1·14–11·8]).

A list of the 40 first authors, 80 main endpoints and p values, and 80 bibliographical references is available from the authors on request.



Characteristics of trial reports as predictors for language of publication
Results from univariate logistic regression models.

Discussion

English is the predominant language in contemporary medical research. Investigators outside the English-speaking world who want their work to be recognised have little choice but to attempt to publish in English. We found that clinical trials are more likely to be reported in an English-language journal if they contain significant results whereas other trials were published in national journals in German. In the USA[19] and the UK,[3] surveys have shown that investigators are reluctant to submit studies without significant results for publication, possibly in anticipation of rejection. Indeed, the instructions for authors of a leading specialist journal stated that "mere confirmation of known facts will be accepted only in exceptional cases; the same applies to reports of experiments and observations having no positive outcome".[20] Researchers whose work is not in English may, in this situation, decide to publish in a national journal rather than not to publish at all. The proportion of trials without significant results that are published in any language may, therefore, be higher for studies done outside the English-speaking world.

Moher and colleagues[11] compared reporting quality of trials published in seven English-language general-medicine journals with six journals published in other languages. They obtained 51% and 46% of the maximum quality score for English-language and non-English-language reports. Our findings were similar and confirmed that differences in quality between English and non-English language reports are small. There is ample room for improvement, independent of language of publication. The Consolidated Standards of Reporting Trials (CONSORT),[21] developed by an international panel of methodologists and medical-journal editors, and adopted by *The Lancet* and 32 other journals, provide clear guidelines for the appropriate reporting of RCTs. Unfortunately, the majority of journals have not adopted these guidelines.

Our findings show that if systematic reviews and meta-analyses ignore trials published in languages other than English, bias could be introduced. This has been called Tower of Babel bias,[10] but we suggest that the term English-language bias is more appropriate. English-language bias is a concern because many systematic reviews and meta-analyses are exclusively based on trials that were published in English. A survey of eight general-medicine journals showed that among 36 meta-analyses published between 1991 and March, 1993, 26 (72%) had restricted their search to studies reported in English.[10] The importance of trials published in languages other than English seems to depend on the topic. In ophthalmology, for example, 20% of trial reports identified by a manual search and Medline searches were in languages other than English,[7] compared with 50–80% of trials on homoeopathy or phytotherapy.[22,23] The latter often appeared in publications not indexed in Medline.[24] Embase (the Excerpta Medica database) contains more journals published in languages other than English than does Medline. The use of reference lists, however, seems to be more important for finding further trials than the choice of electronic database.[24]

Are trials that are published in German potentially relevant for systematic reviews? Omission of trials that did not test clinically relevant interventions and assess meaningful endpoints is justified. In our sample, at least half the German-language trials reported clinically relevant interventions and endpoints, such as cytostatic or adjuvant therapies and survival or remission rates among patients with malignant disease,[25,26] antibotic therapy and cure of

duodenal ulcers,[27] and immunotherapy and the number of pain-free joints in rheumatoid arthritis.[28]

The proportion of trials published in German-language general-medicine journals diminished over time. Trials were increasingly published in English-language journals, mostly specialist journals. During the same time, impact factors of German-language journals declined. In response, several of these journals started to publish some articles in English or changed the language of the publication entirely to English. *Klinische Wochenschrift* is an example. The journal, founded as *Berliner Klinische Wochenschrift* in 1864, had been renamed to *Klinische Wochenschrift* in 1921 because "it had become apparent that local medical journals were no longer attractive to authors seeking national or international recognition".[29] In 1992, the editor decided to relaunch the journal in English and changed its name to *The Clinical Investigator* and introduced a new section on molecular medicine.[29] 3 years later, the journal became the official journal of the German Working Group for Gene Therapy, and again changed its title to *Journal of Molecular Medicine*.[30] This transition can be seen as a reflection of publication trends in German-speaking Europe.

31% of first authors who had published in German and English reported the same results in both languages. The uniform requirements for manuscripts submitted to biomedical journals,[31] developed by the International Committee of Medical Journal Editors, state that multiple publication of the same findings is rarely acceptable. Secondary publication in another language is justified only if several conditions are met—in particular, informing readers about the primary publication. Among our sample of 19 duplicate publications, the earlier report was explicitly acknowledged or quoted in only five instances. Our findings confirm an earlier suggestion that multiple publications of the same trial results in different languages without cross-report reference is common.[8]

The omission of RCTs published in German from systematic reviews and meta-analysis will reduce precision and, more importantly, may introduce English-language bias. Studies similar to this one should ideally be done for each language in which RCTs are reported. We believe, however, that the same situation will arise when reports published in other languages are excluded, particularly other west European languages. The situation may be different for Russian, Japanese, and Chinese journals.[32] The laborious effort of the Cochrane Collaboration to identify as many RCTs as possible through manual searches of many medical journals published in different languages is important for the validity of systematic reviews. The Cochrane Controlled Trials Register[14] is likely to be the best single source of trials for inclusion in systematic reviews and meta-analyses.

Contributors

Matthias Egger was involved in study design and assessment of main endpoints, was responsible for overall supervision and statistical analysis, and had main responsibility for the writing of the article. Tanja Zellweger-Zähner was responsible for hand-searching Swiss and Austrian journals and reports identified from Medicine, and was involved in assessment of main endpoints, quality scoring, and writing up the article. Martin Schnieder was involved in the study design, assessing reports identified from Medline, and quality scoring. Christopher Junker was responsible for matching of articles and involved in study design and assessment of reports identified from Medline. Christian Lengeles contributed to the manual search of Swiss journals, and Gerd Antes manually searched *Deutsche Medizinische Wochenschrift* for articles, and did statistical analysis; both were involved in writing up this article.

Acknowledgments

We thank Iain Chalmers for helpful comments on this paper. Manual search of German-language journals was funded by the BIOMED 1 programme (grant number BMHI-CT94-1289) and the Bundesministerium für Gesundheit. Manual searches of Swiss and Austrian journals were supported by the Swiss Academy of Medical Sciences and the Department of Public Health and Epidemiology of the Swiss Tropical Institute. ME and CL were funded by the Swiss National Science Foundation.

References

1 Anon. Cochrane's legacy. *Lancet* 1992; **340:** 1131–32.
2 Mulrow CD. Rationale for systematic reviews. *BMJ* 1994; **309:** 597–99.
3 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; **337:** 867–72.
4 Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987; **6:** 11–29.
5 Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias in clinical trials. *Control Clin Trials* 1987; **8:** 343–53.
6 Gøtzsche PC. Reference bias in reports of drug trials. *BMJ* 1987; **295:** 654–56.
7 Egger M, Davey Smith G. Meta-analysis: bias in location and selection of studies. *BMJ* 1997 (in press).
8 Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *Lancet* 1996; **347:** 1024–26.
9 Anon. Evidence-based medicine, in its place. *Lancet* 1995; **346:** 785.
10 Grégoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemol* 1995; **48:** 159–63.
11 Moher D, Fortin P, Jadad AR, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996; **347:** 363–66.
12 Chalmers I, Dickersin K, Chalmers TC. Getting to grips with Archie Cochrane's agenda. *BMJ* 1992; **305:** 786–88.
13 Egger M, Zellweger T, Antes G. Randomised trials in German-language journals. *Lancet* 1996; **347:** 1047–48.
14 The Cochrane Controlled Trials Register. In: The Cochrane Library. CDROM and online. Cochrane Collaboration; issue 1. Oxford: Update Software, 1997.
15 Chalmers I, Adams M, Dickersin K, et al. A cohort study of summary reports of controlled trials. *JAMA* 1990; **263:** 1401–05.
16 Jadad AR, Moore RA, Carrol D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; **17:** 1–12.
17 Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995; **16:** 62–73.
18 Institute for Scientific Information. 1994 Science Citation Index. Journal citation reports: a bibliometric analysis of science journals in the ISI database. Philadelphia: Institute for Scientific Information, 1995.
19 Dickersin K, Yuan-IM, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *JAMA* 1992; **267:** 374–78.
20 Anon. Manuscript guideline. *Diabetologia* 1984; **25:** 4A.
21 Begg CB, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996; **276:** 637–39.
22 Kleijnen J, Knipschild P, ter Riet G. Clinical trials of homoeopathy. *BMJ* 1991; **302:** 316–23.
23 Kleijnen J, Knipschild P. Ginkgo biloba. *Lancet* 1992; **340:** 1136–39.
24 Kleijnen J, Knipschild P. The comprehensiveness of Medline and Embase computer searches. *Pharmaceutisch Weekblad Sci Edn* 1992; **14:** 316–20.
25 Jurincic-Winkler CD, Horlbeck R, von der Kammer H, Scheit KH, Klippel KF. Adjuvante Immuntherapie mit Keyhole Limpet Hemocyanin (KLH) beim Nierenzellkarzinom der Kategorie PT2 N+ und PT3-4, No-N+, Mo. *Wien Klin Wochenschr* 1994; **14:** 455–58.
26 Flamm J, Fischer M. Komplette Androgenblockade (Orchiektomie+Flutamid) versus Androgenblockade mit Zytostase (Orchiektomie+Estramustin) in der Behandlung des virginellen fortgeschrittenen Prostatakarzinoms. *Wien Klin Wochenschr* 1988; **100:** 589–92.
27 Bayerdörffer E, Kasper G, Pirlet T, Sommer A, Ottenjann R. Ofloxacin in der Therapie Campylobacter-pylori-positiver Ulcera duodeni. *Deutsch Med Wochenschr* 1987; **112:** 1407–11.
28 Lemmel EM, Bach GL, Bolten W, et al. Immunodulierende Therapie der chronischen Polyarthritis mit Thymopentin. *Deutsch Med Wochenschr* 1988; **113:** 172–76.
29 Zöllner N. The Clinical Investigator. *Clin Invest* 1992; **70:** 1–2.
30 Ganten D. The Journal of Molecular Medicine: tradition, continuity and renaissance. *J Mol Med* 1995; **73:** 1–3.
31 International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. *N Engl J Med* 1991; **324:** 424–28.
32 Gasel B, Weng Y. Medical journals in China. *Ann Intern Med* 1990; **112:** 70–72.