DISCUSSION PUBLISHERS
Research Center · Open Access Edition

Local Literature Bias in Genetic Epidemiology: An Empirical Evaluation of the Chinese Literature

Zhenglun Pan, Thomas A Trikalinos, Fotini K. Kavvoura, Joseph Lau, John P. A. Ioannidis

Discussion Research ID	DPR-2026-1194
Year	2005
Field	Decision Sciences
DOI	10.1371/journal.pmed.0020334
Author ORCID	0000-0003-1932-2619
Licence	CC-BY
Original source	PLoS Medicine
Permanent link	https://discussionpublishers.com/research/57be7d43-71a4-4a14-b511-6bd6923d990e
Publisher	Discussion Publishers discussionpublishers.com

Abstract

BACKGROUND: Postulated epidemiological associations are subject to several biases. We evaluated whether the Chinese literature on human genome epidemiology may offer insights on the operation of selective reporting and language biases. METHODS AND FINDINGS: We targeted 13 gene-disease associations, each already assessed by meta-analyses, including at least 15 non-Chinese studies. We searched the Chinese Journal Full-Text Database for additional Chinese studies on the same topics. We identified 161 Chinese studies on 12 of these gene-disease associations; only 20 were PubMed-indexed (seven English full-text). Many studies (14-35 per topic) were available for six topics, covering diseases common in China. With one exception, the first Chinese study appeared with a time lag (2-21 y) after the first non-Chinese study on the topic. Chinese studies showed significantly more prominent genetic effects than non-Chinese studies, and 48% were statistically significant per se, despite their smaller sample size (median sample size 146 versus 268, $p < 0.001$). The largest genetic effects were often seen in PubMed-indexed Chinese studies (65% statistically significant per se). Non-Chinese studies of Asian-descent populations (27% significant per se) also tended to show somewhat more prominent genetic effects than studies of non-Asian descent (17% significant per se). CONCLUSION: Our data provide evidence for the interplay of selective reporting and language biases in human genome epidemiology. These biases may not be limited to the Chinese literature and point to the need for a global, transparent, comprehensive outlook in molecular population genetics and epidemiologic studies in general.

Republished by Discussion Publishers under the original CC-BY licence. Copyright remains with the original authors; please cite the original work.

Cite as: Zhenglun Pan, Thomas A Trikalinos, Fotini K. Kavvoura, Josep (2005). Local Literature Bias in Genetic Epidemiology: An .

Local Literature Bias in Genetic Epidemiology: An Empirical Evaluation of the Chinese Literature

Zhenglun Pan[1], Thomas A. Trikalinos[2,3], Fotini K. Kavvoura[2], Joseph Lau[3], John P.A. Ioannidis[2,3,4*]

1 Department of Rheumatology, Shandong Provincial Hospital, Jinan 250021, Shandong, China **2** Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, **3** Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America, **4** Biomedical Research Institute, Foundation for Research and Technology–Hellas, Ioannina, Greece

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

Author Contributions: The original idea for this study arose in an electronic conversation between ZP and JPAI. JPAI drafted the protocol that was critically evaluated by all other authors. ZP performed all the database searches. ZP performed the data extraction that was duplicated by FKK and JL. TAT performed the statistical analyses. JPAI wrote the final draft. All authors interpreted the results, commented critically on the manuscript, and approved its final version.

Academic Editor: Marta Gwinn, U.S. Centers for Disease Control and Prevention, United States of America

Citation: Pan Z, Trikalinos TA, Kavvoura FK, Lau J, Ioannidis JPA (2005) Local literature bias in genetic epidemiology: An empirical evaluation of the Chinese literature. PLoS Med 2(12): e334.

Received: May 9, 2005
Accepted: August 15, 2005
Published: November 22, 2005

DOI: 10.1371/journal.pmed.0020334

Copyright: © 2005 Pan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

*To whom correspondence should be addressed. E-mail: jioannid@cc.uoi.gr

ABSTRACT

Background

Postulated epidemiological associations are subject to several biases. We evaluated whether the Chinese literature on human genome epidemiology may offer insights on the operation of selective reporting and language biases.

Methods and Findings

We targeted 13 gene-disease associations, each already assessed by meta-analyses, including at least 15 non-Chinese studies. We searched the Chinese Journal Full-Text Database for additional Chinese studies on the same topics. We identified 161 Chinese studies on 12 of these gene-disease associations; only 20 were PubMed-indexed (seven English full-text). Many studies (14–35 per topic) were available for six topics, covering diseases common in China. With one exception, the first Chinese study appeared with a time lag (2–21 y) after the first non-Chinese study on the topic. Chinese studies showed significantly more prominent genetic effects than non-Chinese studies, and 48% were statistically significant per se, despite their smaller sample size (median sample size 146 versus 268, $p < 0.001$). The largest genetic effects were often seen in PubMed-indexed Chinese studies (65% statistically significant per se). Non-Chinese studies of Asian-descent populations (27% significant per se) also tended to show somewhat more prominent genetic effects than studies of non-Asian descent (17% significant per se).

Conclusion

Our data provide evidence for the interplay of selective reporting and language biases in human genome epidemiology. These biases may not be limited to the Chinese literature and point to the need for a global, transparent, comprehensive outlook in molecular population genetics and epidemiologic studies in general.

Discussion Publishers edition DPR-2026-1194 https://discussionpublishers.com/research/57be7d43-71a4-4a14-b511-6bd6923d990e

Introduction

Research conducted in non-English-speaking countries may be published either in English-language journals that are usually indexed in major international bibliographic databases or in domestic journals, many of which are not indexed in international databases. There is some empirical evidence that the decision to publish in international versus domestic journals may be influenced by the nature of the results: Significant results may be published in international journals, while nonsignificant results appear in the local literature, resulting in language bias (the "tower of Babel" bias) [1,2]. The opposite phenomenon, a reverse tower of Babel bias, nevertheless has also been described [3] in which most of the locally produced and published literature is spuriously statistically significant. Moreover, other investigators have questioned whether the inclusion or not of non-English studies makes any meaningful difference in the overall picture of the evidence [4].

The available evidence on these biases stems from the literature of randomized controlled trials. However, there are other fields in which language biases may be particularly important to appreciate. Genetics poses some special challenges. There are millions of polymorphisms in the human genome, and an exponentially increasing number of studies are trying to associate genetic polymorphisms with the risk of common diseases or treatment outcomes [5]. The risk conferred by each one of these genetic markers is usually small [5], with odds ratios between 1.1 and 1.4. Therefore, selective publication of studies with different results may potentially invalidate the overall picture about genetic risk factors. Moreover, there is major debate on whether there are differences in the strength of the genetic effects across people of different "racial" descent [6–8]. Language-related biases would tend to affect predominantly literature that refers to populations of specific "racial" descent, thus affecting the larger debate on "racial" descent differences.

The Chinese literature is a prominent example of possible bias, because a plethora of domestic scientific journals are not cataloged in international databases. China accounts for one-fifth of the world population, and this research is of major importance not only for China, but also internationally. It has been estimated that overall, for each internationally indexed publication from China, there are 18 publications in local nonindexed journals [9]. The consequences of potential selective publication and language biases for human genome epidemiology research and for biomedical research in general are unknown. Here we aimed to evaluate the extent to which genetic association studies are published in local Chinese journals not indexed in PubMed. We tried to understand whether the results of the Chinese literature differs from the results of the non-Chinese literature and what the implications would be for the total evidence on postulated epidemiological associations and inherent biases.

Methods

Definitions

The primary comparison addressed the results of Chinese versus non-Chinese studies. "Chinese studies" refers to studies performed in the People's Republic of China, regardless of the language of publication. All of them have been performed in people of Chinese descent. Chinese studies are further classified according to whether they are indexed in PubMed or not. "Non-Chinese studies" refers to studies performed outside of China, regardless of the language of publication and regardless of the "racial" descent of the studied populations. Non-Chinese studies are further classified according to whether they evaluated people of Asian or non-Asian descent.

Database of Meta-Analyses of Gene-Disease Associations

We used published meta-analyses of gene-disease associations with binary outcomes and unrelated subjects. Whenever a publication provided data on more than one "racial" descent group, these were split and counted as separate studies. We started from a dataset of 55 meta-analyses previously used in an evaluation of differences between small and larger genetic association studies with binary outcomes [10]. The exact search strategy and eligibility criteria for these meta-analyses have been described previously [10,11]. For each one of them, we updated searches until December 2004, in order to identify more recent meta-analyses on exactly the same topic and containing more studies. More comprehensive meta-analyses replaced the older ones. Then we focused only on meta-analyses in which at least 15 non-Chinese studies were already available. We took this approach because there is evidence that the early literature on gene-disease associations often provides unreliable, inflated results [11,12]. Moreover, Chinese studies may not appear for at least a few years after the appearance of the first non-Chinese studies, thus meta-analyses with few non-Chinese studies may not have had any Chinese studies published yet. Meta-analyses were selected regardless of whether or not they already included any studies from China or individuals of Chinese ethnic descent. None of these meta-analyses had access to Chinese journals not indexed in PubMed, and all included studies were PubMed-indexed.

Search for Additional Studies from the Chinese Literature

For each of the eligible meta-analyses, we searched the national Chinese database of biomedical literature (last search December 2004) for potentially additional gene-disease association studies published in local Chinese journals that would fulfill the eligibility criteria of the original meta-analysis. The Chinese Journal Full-Text Database covers 8,000 journals since 1994, and it is accessible with username and password via the Web site of Tsinghua University. We excluded family-based studies, since they are based on linkage analyses, and these have also been excluded from the original meta-analyses as well.

The search strategy for each topic used the name of each genetic marker (using the abbreviated name of the gene, the expanded name of the gene, and the polymorphism) in combination with terms pertaining to the disease and/or outcome of interest. Retrieved abstracts and articles were further screened for eligibility by the same native Chinese investigator (ZP) who performed the literature search. When in doubt, two other investigators (JPAI and JL), one of them Chinese-speaking (JL), decided on the study's eligibility.

Data Extraction

From each eligible Chinese study, we recorded the name of the first author, journal of publication, year of publication,

Discussion Publishers edition DPR-2026-1194 https://discussionpublishers.com/research/57be7d43-71a4-4a14-b511-6bd6923d990e

ethnic descent, and data on the 2×2 table for the association (data necessary to derive the crude odds ratio and standard error thereof for the probed association). For consistency, the same genetic contrast was used as in the original meta-analysis. We also recorded whether the study was also indexed in PubMed.

We also examined, in each Chinese article, whether the disease was defined with specific criteria, whether any effort was described to ensure that the controls were indeed disease-free or otherwise appropriate, whether it was specified that genotyping was performed blinded to the clinical status, whether there was any mention that the disease-free controls were tested for conformity to Hardy-Weinberg equilibrium, whether any authors were involved from countries other than China, and whether the article was published in an international or national versus a local journal.

Data extraction was performed by a native Chinese investigator (ZP). Key data were independently verified by a second investigator (FKK) whenever tables in English were available or from another Chinese-speaking investigator (JL) otherwise. The few discrepancies were discussed and consensus was reached with a third arbitrator.

Data Analysis

Descriptive statistics summarized the number of studies, total sample size, number and percentage of studies with statistically significant results on their own, and year of publication. Sample sizes were compared between groups of studies with the Mann-Whitney U test and with median regression adjusted for topic (bootstrap p-values). The proportion of studies with statistically significant results was compared with the χ^2 test.

For each meta-analysis and for each group of studies, we estimated the summary odds ratio with inverse variance random effects models, which allow for between-study heterogeneity and incorporate it in the calculations [13]. We tested for between-study heterogeneity with the χ^2-distributed Q statistic (considered significant at $p < 0.10$) [13], and estimated its extent with the I^2 statistic. I^2 ranges between 0% and 100% and represents the proportion of between-study variability that can be attributed to heterogeneity rather than chance (considered large for values of 75% and higher) [14]. Given the prominent differences in effect sizes between different groups of studies, it was considered inappropriate to obtain an overall summary effect including all of them. Instead, we estimated whether the results of different groups of studies differed between themselves beyond chance. A standardized z-score statistic was employed, as previously described [15].

For Chinese studies, for each study we estimated the probability that it would have a formally statistically significant result at the $\alpha = 0.05$ level, conditional on the sample size of its case and control groups, the genetic marker frequency in the controls, and the summary genetic effect seen across Chinese studies. This calculation was performed as a regular power calculation for a case-control study. The sum of these probabilities across Chinese studies (the expected number of studies with statistically significant results) was then compared to the observed number of statistically significant findings using a χ^2 test.

We also compared PubMed-indexed versus not PubMed-indexed Chinese studies as to all the other study and quality characteristics mentioned in the Data Extraction section above.

Analyses were conducted in Intercooled Stata 8.2 (Stata Corp., College Park, Texas, United States) using the *metan* module. p-Values are two-tailed.

Results

Data on Chinese and Non-Chinese Studies

Thirteen published meta-analyses were found with at least 15 non-Chinese studies [16–26]. Data on any Chinese studies could be retrieved for 12 of those, and these 12 topics are considered from now on (for the association of *DRD2* TaqIA polymorphism with alcoholism [26], no Chinese study was identified; Table 1). Overall, there were 161 eligible Chinese studies, only 20 of which were indexed in PubMed. Of the 20 Chinese studies indexed in PubMed (two on ID1, two on ID2, one on ID3, two on ID4, five on ID10, one on ID11, and seven on ID12; Table 1), only six had already been included in the published meta-analyses (one on ID11 and five on ID12), while the others were more recent; only seven of the 20 were published in full-text English journals. Of the 309 non-Chinese studies already included in the published meta-analyses, 44 pertained to populations of Asian descent (Japan, $n = 25$; Korea, $n = 7$; Chinese people outside of China, $n = 5$; Taiwan, $n = 4$; Malaysia, $n = 2$; and Singapore $n = 1$), and 265 to people of non-Asian descent (Figure 1).

For six topics we retrieved an extensive Chinese literature from the Chinese database (14–35 studies for each), while for the other topics the Chinese studies were sparse (four or fewer studies per topic) (Table 1). Chinese data were typically sparse if the disease was relatively uncommon in China compared with other countries, e.g., bladder cancer (bladder cancer is almost 10-fold less common in China than in Europe or the United States) [27] and alcoholism (at least until the early 1990s) [28]; or if the disease was not very common globally (e.g., systemic lupus erythematosus and schizophrenia). Chinese studies were plentiful if the disease was common (e.g., cancer in general, lung cancer, coronary heart disease, and diabetic nephropathy), with the exception of the postulated association of the *ITGB3* gene with coronary artery disease, for which only one Chinese study was available.

With one exception, where the first Chinese study was published in the same year as the first non-Chinese study, the first Chinese study always appeared with a considerable time lag compared with the remaining world literature (2–21 y; Table 1).

Study Sample Sizes

The sample size for Chinese studies was significantly smaller than for non-Chinese studies ($p < 0.001$ both by U test and topic-adjusted median regression; Figure 1). Although non-Chinese studies of non-Asian descent populations overall seemed to be larger than studies on non-Chinese studies of Asian descent populations ($p < 0.001$ by U test), the difference was lost after adjusting for topic ($p = 0.72$). Chinese studies indexed or not indexed in PubMed did not differ in sample size ($p = 0.79$ by U test, $p = 0.55$ by median regression; Figure 1).

Statistically Significant Results

Overall, 78 (48%) of the 161 Chinese studies had formally statistically significant results. There was some heterogeneity

Discussion Publishers edition DPR-2026-1194 https://discussionpublishers.com/research/57be7d43-71a4-4a14-b511-6bd6923d990e

Table 1. Eligible Meta-Analyses

ID	Disease/Outcome	Gene (Polymorphism)	Genetic Contrast	Studies (Total Sample)		Studies with $p < 0.05$ (%)		First Year Published	
				Chinese	Non-Chinese	Chinese	Non-Chinese	Chinese	Non-Chinese
1	Myocardial infarction	*ACE* (insertion/deletion)	DD versus DI + II	27 (4,514)	15 (18,664)	17 (63)	6 (40)	1996	1992
2	Ischemic heart disease	*ACE* (insertion/deletion)	DD versus DI + II	35 (6,586)	17 (21,876)	20 (57)	4 (23)	1998	1994
3	Cancer	*HRAS/HRAS1* (rare alleles)	Rare versus common alleles	23 (1,559)	24 (8,542)	7 (30)	8 (33)	1994	1985
4	Bladder cancer	*NAT2* (slow acetylation alleles)	Slow/slow versus others	3 (417)	20 (5,836)	1 (33)	6 (30)	2000	1979
5	Diabetic nephropathy	*ACE* (insertion/deletion)	II versus ID + DD	25 (3,857)	20 (5,393)	8 (32)	6 (30)	1997	1994
6	Coronary artery disease	*ITGB3* (L33P)	A2A2 versus A1A2 + A1A1	1 (152)	31 (17,315)	0 (0)	4 (13)	1999	1996
7	Bladder cancer	*GSTM1* (gene deletion)	Null/null versus others	2 (400)	20 (5,795)	1 (50)	7 (35)	2002	1992
8	SLE nephritis	*FCGR2A* (R131H)	RR versus RH + HH	1 (86)	24 (2,801)	0 (0)	2 (8)	2003	1995
9	SLE	*FCGR2A* (R131H)	RR versus RH + HH	2 (261)	21 (4,708)	2 (100)	4 (19)	2000	1995
10	Coronary heart disease	*MTHFR* (677C/T)	TT versus CC	14 (1,778)	40 (23,922)	8 (57)	7 (18)	1998	1996
11	Schizophrenia	*DRD3* (Bal1)	Ser/Ser + Gly/Gly versus Ser/Gly	4 (1,527)	39 (8,556)	0 (0)	1 (2)	1993	1993
12	Lung cancer	*GSTM1* (gene deletion)	Null/null versus others	24 (5,909)	38 (16,119)	14 (58)	3 (6)	1997	1991

ACE, angiotensin converting enzyme; *DRD2/DRD3,* dopamine receptor D2/D3; *FCGR2A,* low-affinity receptor of the Fc domain of immunoglobulin G; *GSTM1,* glutathione S-transferase M1; *HRAS,* Harvey rat sarcoma viral oncogene homolog; *ITGB3,* platelet glycoprotein receptor IIIa; *MTHFR,* methylenetetrahydrofolate reductase; SLE, systemic lupus erythematosus.

DOI: 10.1371/journal.pmed.0020334.t001

in this proportion across topics (exact $p = 0.041$). Conversely, only 57 (18%) of 309 non-Chinese studies had significant results, despite the larger sample size, and the percentage differed greatly across the 12 topics (exact $p < 0.001$). As shown in Figure 1, the proportion of formally statistically significant studies differed between PubMed-indexed Chinese studies, non-PubMed-indexed Chinese studies, non-Chinese studies of Asian-descent populations, and non-Asian studies (65%, 46%, 27%, and 17%, respectively; $p < 0.001$ by χ^2). None of the five studies on Chinese-descent people living outside of China had statistically significant results.

Changes in Study Sample Sizes and Significant Results over Time

The sample size of Chinese studies increased over time (Spearman correlation coefficient for publication year and sample size, 0.32, $p < 0.001$), while this was not seen for non-Chinese studies (correlation coefficient 0.00, $p = 0.95$). As for non-Chinese studies, the proportion of Chinese studies with formally significant results did not increase over time; if anything, there was a trend towards decrease (47/89 [53%] in 1993–2000 versus 41/72 [43%] in 2001–2004; $p = 0.27$).

Comparison of Genetic Effects

Table 2 summarizes the genetic effect sizes. As shown, whenever there was a sizeable literature of Chinese studies, the gene-disease association was always formally significant in both non-Chinese and Chinese studies, but Chinese studies always showed a larger genetic effect than the non-Chinese studies (Figure 1). In five of the six topics the observed difference was even beyond chance ($p < 0.05$ on the z-score). Even with limited data, Chinese studies suggested larger estimates than non-Chinese studies also in the other three topics where there was some overall evidence for the presence of a gene-disease association; the genetic effect difference was beyond chance in one of the three topics (Table 2).

PubMed-indexed Chinese studies were too few for formal comparisons, but the available data suggested that they often tended to provide extreme estimates of genetic effects (Figure 2). In three of the five topics where at least two such studies were available, their summary estimate was the most extreme observed compared with any other group of studies (non-PubMed Chinese, non-Chinese Asian, and non-Chinese non-Asian).

Non-Chinese studies of Asian descent populations were available for eight topics. In seven of the eight cases, the estimated genetic effect size was stronger in these Asian-descent studies than in the non-Chinese non-Asian descent studies (Table 3). The difference was beyond chance in two topics (the associations of *MTHFR* C677T polymorphism with coronary heart disease [ID10], and of *GSTM1* gene deletion

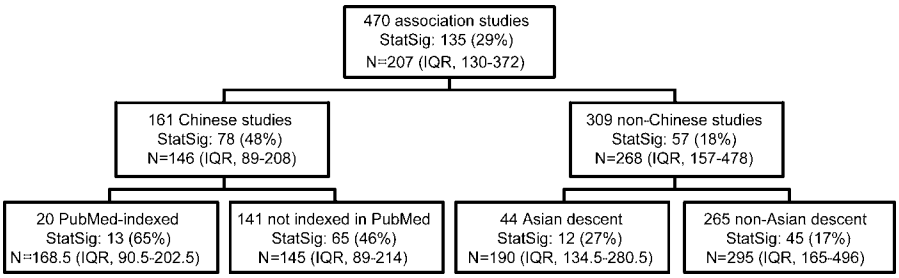


Figure 1. Categorization of the Examined Genetic Association Studies

IQR, interquartile range; N, sample size (as median and interquartile range); StatSig, statistically significant at the 0.05 level.

DOI: 10.1371/journal.pmed.0020334.g001

Discussion Publishers edition DPR-2026-1194 https://discussionpublishers.com/research/57be7d43-71a4-4a14-b511-6bd6923d990e

Table 2. Genetic Effects in Chinese and Non-Chinese Studies

ID	Random Effects Odds Ratio (95% CI)		Discrepancy in Effect (*p*-Value)	I^2 for Heterogeneity (%)	
	Chinese	Non-Chinese		Chinese	Non-Chinese
1	2.21 (1.84–2.66)[a]	1.28 (1.09–1.50)[a]	−4.42 (<0.01)	35	65
2	2.02 (1.66–2.47)[a]	1.20 (1.06–1.36)[a]	−4.36 (<0.01)	60	54
3	7.66 (4.51–13.0)	1.84 (1.54–2.21)	−4.99 (<0.01)	0	7
4	3.06 (0.95–9.87)[a]	1.43 (1.20–1.71)[a]	−1.26 (0.21)	81	48
5	0.53 (0.41–0.69)	0.68 (0.55–0.84)[a]	1.43 (0.15)	66	50
6	0.83 (0.02–44.4)	1.10 (0.99–1.21)[a]	0.14 (0.89)	NP	45
7	1.58 (1.02–2.43)	1.44 (1.25–1.67)	−0.38 (0.70)	0	27
8	0.83 (0.34–2.03)	1.11 (0.88–1.41)	0.63 (0.53)	NP	26
9	2.87 (1.50–5.50)	1.29 (1.10–1.52)	−2.33 (0.02)	0	19
10	2.03 (1.44–2.86)[a]	1.14 (1.01–1.30)[a]	−3.09 (<0.01)	59	50
11	0.95 (0.76–1.17)	1.11 (1.01–1.22)	1.33 (0.18)	0	8
12	1.63 (1.38–1.92)[a]	1.14 (1.07–1.22)	−3.54 (<0.01)	52	0

The ID numbers correspond to Table 1. The discrepancy between the Chinese and non-Chinese studies is expressed as a *z*-score and the corresponding *p*-value.
[a]Significant between-study heterogeneity ($p < 0.10$ for the *Q* statistic)
CI, confidence interval; NP, not pertinent (only one study available).
DOI: 10.1371/journal.pmed.0020334.t002

with lung cancer [ID12]). In topics for which several studies of different groups were available, the non-Chinese studies of Asian-descent populations seemed to have effect sizes somewhere between the effect sizes of Chinese studies and non-Asian studies (see Figure 1).

Expected versus Observed Significant Findings in Chinese Studies

Power calculations based on asymptotic statistical testing suggested that even if the large summary genetic effects claimed by the Chinese studies were genuine, one would expect 56.6 formally statistically significant studies, substantially fewer than the 78 observed in the database ($p < 0.001$). Based on exact statistical testing, one would expect 61.1 significant studies instead of the 81 observed ($p = 0.001$).

Qualitative Comparison of Chinese Studies Indexed versus Not Indexed in PubMed

PubMed-indexed Chinese studies did worse than Chinese studies not indexed in PubMed in defining disease with specific criteria (17/20 [85%] versus 137/141 [97%], respectively; exact $p = 0.042$), and in ascertaining the eligibility of controls (13/20 [65%] versus 129/141 [92%], respectively; exact $p = 0.003$). However, the only three Chinese studies mentioning blinding of the genotyping personnel to disease status were PubMed-indexed (exact $p = 0.002$). There was no difference in testing for violations of the Hardy-Weinberg law in the controls (5/20 [20%] versus 38/141 [27%], respectively; exact $p = 1.00$). Only PubMed-indexed Chinese studies had any representation of authors from countries other than China (4/20 [20%] versus 0/141 [0%], respectively; exact $p < 0.001$). As expected, almost all (19/20 [95%]) PubMed-indexed Chinese studies were published in international or national Chinese journals, while only 22 (16%) of the 141 studies not indexed in PubMed were published in national journals.

Discussion

This empirical evaluation reveals a large Chinese literature on human genome epidemiology that deserves more attention from the international community. The vast majority of this literature does not reach PubMed. Chinese studies usually appear with a time lag of several years after an epidemiological association is first postulated in the world literature, but many such studies are published, especially when the disease is perceived to be common in China. Chinese studies typically suggest much stronger genetic effects than non-Chinese studies, and this may be even more prominent for the few studies that reach PubMed. Although Chinese studies are smaller than non-Chinese studies and thus even more underpowered [5], surprisingly half of them reach formal statistical significance for the evaluated gene-disease association. This exaggeration is seen across very diverse topics.

The larger genetic effects in Chinese studies are unlikely to reflect genuine heterogeneity in the effects of genetic risk factors across various "racial" descent populations [8]. Heterogeneity due to ancestry should not have led always to larger effect sizes in all probed gene-disease associations. Therefore, the most likely explanation is publication bias against "negative" results [29–32] or other selection biases in the chase for statistically significant findings [33]. This explanation is further supported by our analysis of the expected number of statistically significant findings. Even if the average genetic effects in the Chinese studies were indeed as large as those observed, one would expect far fewer Chinese studies to have reached formal statistical significance on their own, given their small sample sizes. The alternative explanation that Chinese investigators may be targeting high-risk populations with particularly strong genetic effects is unlikely given these data.

Language may be a marker for other confounding characteristics of these studies, or even of the whole research and publication milieu. Moreover, even within the English-language studies, strong biases may occasionally operate in the confirmation process. Cultural issues may also be involved with unstated pressures to find positive results for various reasons in different settings around the globe. Various compromises of research quality may ensue.

We focused on gene-disease associations for which a considerable number of studies have been published in the English language. It is possible that there could be a

Discussion Publishers edition DPR-2026-1194 https://discussionpublishers.com/research/57be7d43-71a4-4a14-b511-6bd6923d990e

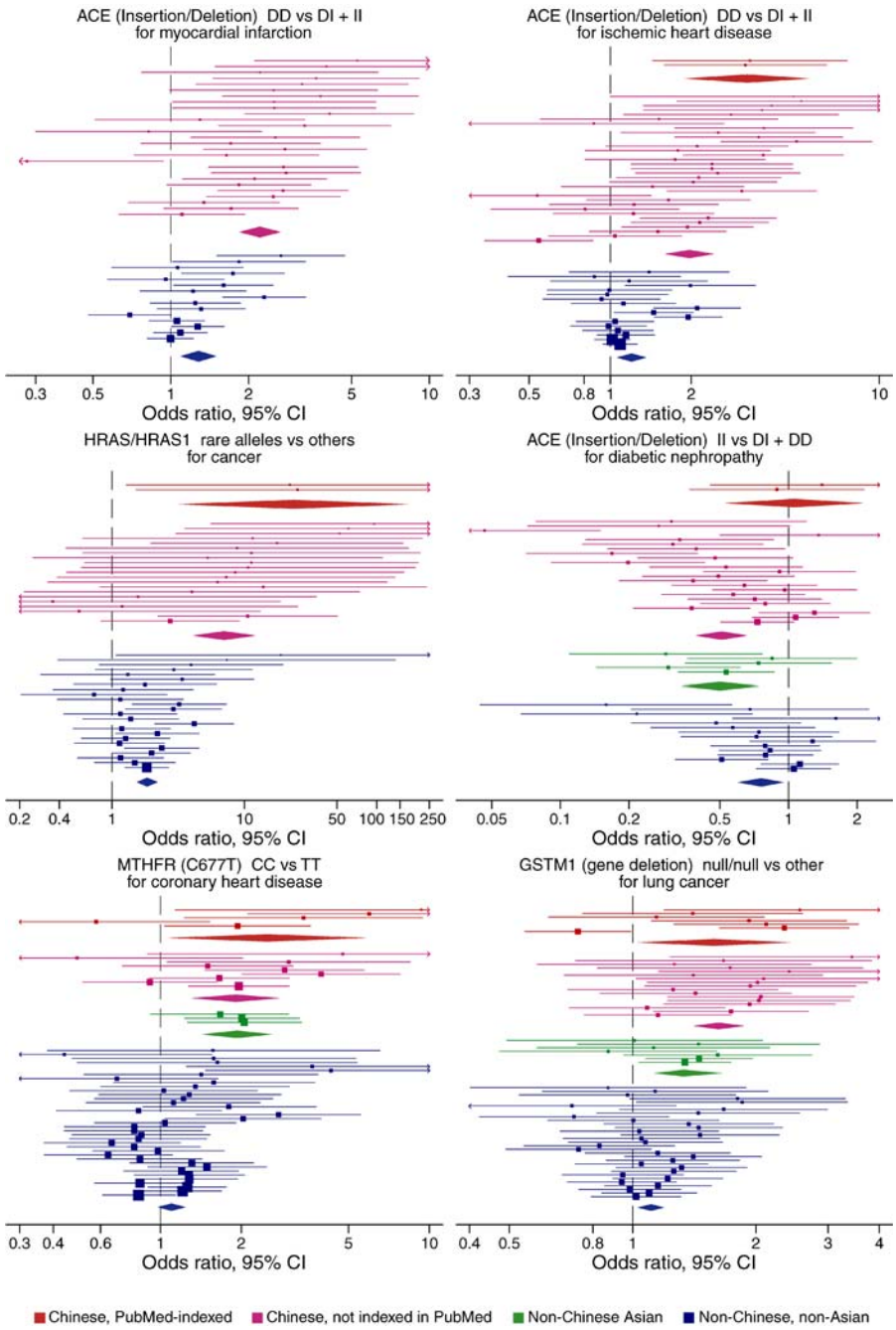


Figure 2. Meta-Analyses of Gene-Disease Associations in a Large Number of Both Non-Chinese and Chinese Studies

Each study is shown by its odds ratio and 95% confidence intervals (CIs). The box of the point estimate is proportional to the study weight. Also shown are summary estimates by random effects calculations (diamonds). Summary estimates are obtained separately for Chinese studies indexed in PubMed (red), Chinese studies not indexed in PubMed (pink), non-Chinese studies of Asian descent populations (green), and studies of persons of non-Asian descent (blue). An odds ratio of 1 means no genetic effect, odds ratios larger than 1 mean genetic predisposition, and odds ratios less than 1 mean genetic protection.

DOI: 10.1371/journal.pmed.0020334.g002

reluctance to submit and publish "negative" or inconclusive results when a large body of English-language literature has shown the presence of genetic effects. Also attempts to confirm multiply supported findings may be more likely to be made, especially with limited resources. However, such pressure for unilateral confirmation destroys the independence and thus also the importance of confirmation.

Our observation is reminiscent also of the randomized trial literature on acupuncture, where studies from China, Russia, Japan, Hong Kong, and Taiwan almost always yielded statistically significant results, in contrast to studies performed in other countries [3]. A predilection for the dissemination of statistically significant results in some non-English speaking countries has also been suspected in other fields, such as lung cancer chemotherapy trials [34]. To our knowledge, there has been no prior documentation of this phenomenon in molecular medicine. Given the rapid pace of production of information in molecular genetics and other

Discussion Publishers edition DPR-2026-1194 https://discussionpublishers.com/research/57be7d43-71a4-4a14-b511-6bd6923d990e

Table 3. Comparison of Genetic Effects in Non-Chinese Studies of Asian-Descent Populations and Non-Asian-Descent Studies

ID	Non-Chinese Asian Studies (Total Sample)	Random Effects Odds Ratio (95% CI)		I^2 for Heterogeneity (%)	
		Non-Chinese Asian	Non-Asian	Non-Chinese Asian	Non-Asian
4	3 (503)	2.09 (1.13–3.85)	1.39 (1.15–1.67)[a]	8	50
5	5 (762)	0.50 (0.34–0.74)	0.76 (0.60–0.95)[a]	31	45
7	4 (1,232)	1.76 (1.34–2.32)	1.38 (1.16–1.63)[a]	0	33
8	7 (709)	1.21 (0.70–2.08)	1.10 (0.84–1.44)[a]	0	37
9	7 (1,339)	1.01 (0.70–1.46)	1.37 (1.13–1.67)	0	32
10	3 (2,002)	1.93 (1.43–2.60)	1.08 (0.95–1.22)[a]	0	41
11	8 (1,862)	1.12 (0.93–1.35)	1.11 (0.99–1.24)	0	20
12	7 (2,484)	1.33 (1.13–1.56)	1.11 (1.03–1.19)	0	0

The ID numbers correspond to Table 1. For ID numbers 1, 2, 3, and 6, there were no non-Chinese studies of Asian descent populations.
[a]Significant between-study heterogeneity ($p < 0.10$ for the Q statistic)
CI, confidence interval.
DOI: 10.1371/journal.pmed.0020334.t003

modern disciplines, this bias may become a serious problem in the appraisal of cutting-edge science and may jeopardize the credibility of molecular discovery research.

We also found some evidence that superimposed language bias [2] is also operating in this literature. Typically, the few PubMed-indexed Chinese studies showed the most extreme genetic effects, and two-thirds of them reached formally statistically significant results on their own, even though their sample size was very small. Therefore, analyses limited to PubMed-indexed studies may sometimes yield spurious results, if the summary estimates are driven by these extreme findings. PubMed-indexed Chinese studies had worse quality ratings in case and control definitions and ascertainment than Chinese studies not indexed in PubMed. Language bias may not be limited to China, but may also be pertinent to other Asian countries with considerable scientific production, and beyond. We found that non-Chinese Asian studies also tend consistently to show relatively larger genetic effects than non-Asian studies, although data were too sparse to be definitive. The relative extent of selective reporting, publication bias, and language bias is difficult to disentangle here and may vary across topics and across local literatures. It would also be useful to analyze the local literatures for Japanese and Korean studies, where a considerable number of local journals also exist.

The Chinese literature is essential for the evaluation of evidence on genetic risk factors. China is making rapid scientific progress in this field, as in many others. It is already participating in the Human Genome project, and the Southern China National Genome Research Center established in Shanghai in 2001 creates new frontiers for gene-disease association studies. Evidence on population genetics, as well as for any other field pertinent to population health, is extremely important to obtain for China from a global perspective. Moreover, it is unlikely that biases are limited to China, as we discussed above. Also, European and American studies are not necessarily unbiased. There is strong evidence that early-published European and American studies that appeared in the most prestigious journals tended to have inflated results [11,35,36].

Here we did not update further the existing non-Chinese data from the published meta-analyses. Our investigation focused on the Chinese literature, and we tried to focus on meta-analyses with a large number of included studies that should hopefully have reached a stable effect estimate. Nevertheless, for at least two of the postulated associations examined here (the associations of *ACE* with cardiac outcomes), a very large study [37] conducted after the meta-analysis found absolutely no effect, while the previous studies had found modest, but statistically significant, genetic effects. Thus not only was the discrepancy against the Chinese studies even larger than what was found in our analyses, but the evidence from the earlier European-descent studies, in particular the smaller ones [16], had also been biased.

Since most effect sizes in genetic epidemiology, and most other molecular medicine fields, are small or modest, one wonders whether many of the postulated associations are generated from the interplay of various reporting and local literature biases that leave no country immune. In some of these postulated associations, the observed effect sizes may simply be estimates of the prevailing bias [38].

One might argue that the inclusion of poor-quality research may contaminate the better literature rather than provide a more accurate, comprehensive picture. Large-scale aggregate evidence may arrive at erroneous conclusions if studies are automatically included without some critical appraisal. However, it is unfair to judge the quality of research on the basis of its regional origin. Chinese studies may often be as good as or even better than many or most studies from countries publishing routinely in the English language [39–41]. Efforts to improve the quality of research around the globe should run in parallel with enhanced access to global research results.

Our findings have two broad implications. First, language bias may be important to consider in meta-analyses of observational studies in general, and its impact may be as large as or larger than its impact on randomized evidence. Second, human genome epidemiology in particular is a global enterprise, and a critical and comprehensive global view is important to decipher artifacts from true genetic effects. Large studies are useful to validate postulated gene-disease associations [12]. However, such studies are difficult to conduct, they are not completely immune from biases, and their targets must be carefully selected given the plethora of test hypotheses in molecular genetics [5,42]. Besides large studies, registration of investigators and data collections is

Discussion Publishers edition DPR-2026-1194 https://discussionpublishers.com/research/57be7d43-71a4-4a14-b511-6bd6923d990e

useful to consider [42,43]. In contrast to randomized trials [44], study registration is impractical in molecular medicine, since investigators would be reluctant to share their hypotheses in public. However, if all investigators working on the genetics of a specific disease were registered in a common network, then it would be easier to trace additional unpublished or non-indexed data. Common networks would also, hopefully, help to improve the quality of research. Such networks should aim for a global, inclusive outlook. The Chinese research output, as well as the output of other non-English-speaking countries, should be appropriately captured. Failure to maintain a global outlook may result in a scientific literature that is driven by the opportunistic dissemination of selected results.

Acknowledgments

TAT and FKK were funded by PENED grants from the General Secretariat for Research and Technology, Greece, and the European Commission. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. ■

References

1. Gregoire G, Derderian F, Le Lorier J (1995) Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? J Clin Epidemiol 48: 159–163.
2. Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, et al. (1997) Language bias in randomised controlled trials published in English and German. Lancet 350: 326–329.
3. Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. Control Clin Trials 1998;19: 159–166.
4. Juni P, Holenstein F, Sterne J, Bartlett C, Egger M (2002) Direction and impact of language bias in meta-analyses of controlled trials: Empirical study. Int J Epidemiol 31: 115–123.
5. Ioannidis JP (2003) Genetic associations: False or true? Trends Mol Med 9: 135–138.
6. Thomas DC, Witte JS (2002) Population stratification: A problem for case-control studies of candidate-gene associations? Cancer Epidemiol Biomarkers Prev 11: 505–512.
7. Wacholder S, Rothman N, Caporaso N (2002) Counterpoint: Bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiol Biomarkers Prev 11: 513–520.
8. Ioannidis JP, Ntzani EE, Trikalinos TA (2004) "Racial" differences in genetic effects for complex diseases. Nat Genet 36: 1312–1318.
9. International Working Party to Promote and Revitalise Academic Medicine. Academic medicine: The evidence base. BMJ 329: 789–792.
10. Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG (2003) Genetic associations in large versus small studies: An empirical assessment. Lancet 361: 567–571.
11. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001) Replication validity of genetic association studies. Nat Genet 29: 306–309.
12. Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG, Ioannidis JP (2004) Establishment of genetic associations for complex diseases is independent of early study findings. Eur J Hum Genet 12: 762–769.
13. Lau J, Ioannidis JP, Schmid CH (1997) Quantitative synthesis in systematic reviews. Ann Intern Med 127: 820–826.
14. Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. Stat Med 21: 1539–1558.
15. Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, et al. (1996) Large trials vs meta-analysis of smaller trials: how do their results compare? JAMA 276: 1332–1338
16. Agerholm-Larsen B, Nordestgaard BG, Tybjaerg-Hansen A (2000) *ACE* gene polymorphism in cardiovascular disease. Meta-analyses of small and large studies in whites. Arterioscler Thromb Vasc Biol 20: 484–492
17. Di Castelnuovo A, de Gaetano G, Donati MB, Iacoviello L (2001) Platelet glycoprotein receptor IIIa polymorphism PLA1/PLA2 and coronary risk: A meta-analysis. Thromb Haemost 85: 626–633
18. Karassa FB, Trikalinos TA, Ioannidis JPA, the FcgRIIa-SLE Meta-Analysis Investigators (2002) Role of the Fcg receptor IIa polymorphism in susceptibility to systemic lupus erythematosus and lupus nephritis. Arthritis Rheum 46: 1563–1571
19. Jonsson EG, Flyckt L, Burgert E, Crocq MA, Forslund K, et al. (2003) Dopamine D3 receptor gene Ser9Gly variant and schizophrenia: association study and meta-analysis. Psychiatr Genet 13: 1–12.
20. Benhamou S, Lee WJ, Alexandrie AK, Boffetta P, Bouchardy C, et al. (2002) Meta- and pooled analyses of the effects of glutathione S-transferase M1 polymorphisms and smoking on lung cancer risk. Carcinogenesis 23: 1343–1350
21. Klerk M, Verhoef P, Clarke R, Blom HJ, Kok FJ, et al. (2002) MTHFR 677C–>T polymorphism and risk of coronary heart disease: a meta-analysis. JAMA 288: 2023–2031.
22. Engel LS, Taioli E, Pfeiffer R, Garcia-Closas M, Marcus PM, et al. (2002) Pooled analysis and meta-analysis of glutathione S-transferase M1 and bladder cancer: A HuGE review. Am J Epidemiol 156: 95–109.
23. Tarnow L, Gluud C, Parving HH (1998) Diabetic nephropathy and the insertion/deletion polymorphism of the angiotensin-converting enzyme gene. Nephrol Dial Transplant 13: 1125–1130
24. Marcus PM, Vineis P, Rothman N (2000) NAT2 slow acetylation and bladder cancer risk: A meta-analysis of 22 case-control studies conducted in the general population. Pharmacogenetics 10: 115–122
25. Krontiris TG, Devlin B, Karp DD, Robert NJ, Risch N (1993) An association between the risk of cancer and mutations in the HRAS1 minisatellite locus. N Engl J Med 329: 517–523
26. Noble EP (1998) The D2 dopamine receptor gene: A review of association studies in alcoholism and phenotypes. Alcohol 16: 33–45
27. Parkin DM, Bray F, Ferlay J, Pisani P (2005) Global cancer statistics, 2002. CA Cancer J Clin 55: 74–108.
28. Li HZ, Rosenblood L (1994) Exploring factors influencing alcohol consumption patterns among Chinese and Caucasians. J Stud Alcohol 55: 427–433.
29. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR (1991) Publication bias in clinical research. Lancet 337: 867–872.
30. Dickersin K, Min YI (1993) Publication bias: The problem that won't go away. Ann N Y Acad Sci 703: 135–148.
31. Munafo MR, Clark TG, Flint J (2004) Assessing publication bias in genetic association studies: Evidence from a recent meta-analysis. Psychiatry Res 129: 39–44.
32. Agema WR, Jukema JW, Zwinderman AH, van der Wall EE (2002) A meta-analysis of the angiotensin-converting enzyme gene polymorphism and restenosis after percutaneous transluminal coronary revascularization: Evidence for publication bias. Am Heart J 144: 760–768.
33. Cohen J (1994) The earth is round (P-less-than .05). Amer Psychol 49: 997–1003.
34. Ioannidis JP, Polycarpou A, Ntais C, Pavlidis N (2003) Randomised trials comparing chemotherapy regimens for advanced non-small cell lung cancer: Biases and evolution over time. Eur J Cancer 39: 2278–2287
35. Gelernter J, Goldman D, Risch N (1993) The A1 allele at the D2 dopamine receptor gene and alcoholism. A reappraisal. JAMA 269: 1673–1677
36. Ioannidis JP (2005) Contradicted and initially stronger effects in highly cited clinical research. JAMA 294: 218–228.
37. Keavney B, McKenzie C, Parish S, Palmer A, Clark S, et al. (2000) Large-scale test of hypothesised associations between the angiotensin-converting-enzyme insertion/deletion polymorphism and myocardial infarction in about 5000 cases and 6000 controls. International Studies of Infarct Survival (ISIS) Collaborators. Lancet 355: 434–442.
38. Ioannidis JP (2005) Why most published research findings are false. PLoS Medicine 2: e124.
39. Salanti G, Amountza G, Ntzani EE, Ioannidis JP (2005) Hardy-Weinberg equilibrium in genetic association studies: An empirical evaluation of reporting, deviations, and power. Eur J Hum Genet 13: 840–848.
40. Bogardus ST Jr, Concato J, Feinstein AR (1999) Clinical epidemiological quality in molecular genetic research: The need for methodological standards. JAMA 281: 1919–1926.
41. Attia J, Thakkinstian A, D'Este C (2003) Meta-analyses of molecular association studies: Methodologic lessons for genetic epidemiology. J Clin Epidemiol 56: 297–303.
42. Ioannidis JP, Rosenberg PS, Goedert JJ, O'Brien TR (2002) International Meta-analysis of HIV Host Genetics. Commentary: Meta-analysis of individual participants' data in genetic epidemiology. Am J Epidemiol 156: 204–210.
43. Ioannidis JP, Bernstein J, Boffetta P, Danesh J, Dolan S, et al. (2005) A network of investigator networks in human genome epidemiology. Am J Epidemiol 162: 302–304.
44. De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, et al. (2004) Clinical trial registration: A statement from the International Committee of Medical Journal Editors. N Engl J Med 351: 1250–1251

Discussion Publishers edition DPR-2026-1194 https://discussionpublishers.com/research/57be7d43-71a4-4a14-b511-6bd6923d990e

Patient Summary

Background There are many different places that medical research can be published. However, some research is never published, which leads to so-called publication bias. One of the biggest divides is between English and non-English research. Research done in non-English-speaking countries can be published in English journals that are usually indexed in major international databases, but more often is published in domestic journals, many of which are not indexed in international databases. This selective publication is called language bias. Scientists have questioned what difference the inclusion or not of non-English studies makes to the total evidence. Publication bias is of concern, especially in genetics, which is a very fast-moving area of research.

Why Was This Study Done? China is a prominent example of a nation with many domestic journals that are not indexed in the international databases. This study looked at Chinese genetic association studies. Understanding the quality and findings of genetics research in China, home to one-fifth of the world's population, is essential for the evaluation of evidence on genetic risk factors. The authors hoped this study would help understand more about selective reporting and language biases.

What Did the Researchers Do and Find? They found that there were many studies (14–35 per topic) for each of the topics they chose to assess, which covered diseases common in China. Generally, Chinese studies appeared a considerable time (two to 21 years) after the first non-Chinese study on the topic. Chinese studies showed stronger genetic effects than non-Chinese studies, despite in some cases having smaller sample sizes. The largest genetic effects were often seen in Chinese studies indexed in Western databases.

What Do These Findings Mean? It seems that there is a combination of selective reporting of studies with interesting findings, and language biases in human gene studies. The main reason most studies didn't appear in the international literature was probably a combination of publication bias and superimposed language bias. It is important to note that such biases are not limited to Chinese literature. Researchers should consider language bias when doing analyses of groups of studies, and efforts to improve the quality of research around the globe should run in parallel with enhanced access to global research results.

Where Can I Get More Information Online? The Cochrane Collaboration has a small section explaining the different types of publication bias:
http://www.cochrane-net.org/openlearning/HTML/mod15–2.htm
BMJ has a presentation on publication bias by a BMJ editor:
http://bmj.bmjjournals.com/talks/publication_bias/index.htm

Discussion Publishers edition DPR-2026-1194 https://discussionpublishers.com/research/57be7d43-71a4-4a14-b511-6bd6923d990e