Peer Review of Social Science Research in Global Health: A View Through Correspondence Letters to *The Lancet*

Victoria Fan, Rachel Silverman, David Roodman, and William Savedoff

Abstract

In recent years, the interdisciplinary nature of global health has blurred the lines between medicine and social science. As medical journals publish non-experimental research articles on social policies or macro-level interventions, controversies have arisen when social scientists have criticized the rigor and quality of medical journal articles, raising general questions about the frequency and characteristics of methodological problems and the prevalence and severity of research bias and error. Published correspondence letters can be used to identify common areas of dispute within interdisciplinary global health research and seek strategies to address them. These letters can be seen as a "crowd-sourced" (but editor-gated) approach to public peer review of published articles, from which some characteristics of perceived bias and error can be gleaned. In December 2012, we used the online version of *The Lancet* to systematically identify relevant correspondence in each issue published between 2008 and 2012. We summarize and categorize common areas of dispute raised in these letters. The five most-cited concerns are: measurement error (51% of papers); omitted variables and confounding (45%); implausibility and lack of external validity (43%); missing or low-quality data (32%); and lack of transparency of methods (30%). We offer several recommendations, including the use of checklists and guidelines to facilitate better documentation of areas of potential bias; deployment of econometric-specific reviewers where appropriate; and explicit online linkage between all correspondence letters and the original articles to which they refer. Most importantly, we recommend that *The Lancet* adopt the replication standard, whereby the data and the coding used to produce the estimates are provided at least to the journal, for reviewers to analyze and replicate the estimates reported by the authors, and ideally to the public, as the leading economics journals now require.

Center for Global Development
www.cgdev.org

Working Paper 371
June 2014

Peer Review of Social Science Research in Global Health: A View Through Correspondence Letters to *The Lancet*

Victoria Fan
Center for Global Development

Rachel Silverman
University of Cambridge

David Roodman
Independent Consultant

William Savedoff
Center for Global Development

CGD is grateful for contributions from The William and Flora Hewlett Foundation and the UK Department for International Development in support of this work.

Victoria Fan, Rachel Silverman, David Roodman, and William Savedoff. 2014. "Peer Review of Social Science Research in Global Health: A View Through Correspondence Letters to *The Lancet*." CGD Working Paper 371. Washington, DC: Center for Global Development.
http://www.cgdev.org/publication/peer-review-social-science-research-global-health-view-through-correspondence-letters

Center for Global Development
2055 L Street NW
Washington, DC 20036

202.416.4000
(f) 202.416.4050

www.cgdev.org

Contents

I. Introduction

In recent years, the interdisciplinary nature of global health has blurred the lines between medicine and social science. With seemingly increasing frequency, medical journals have published non-experimental research articles on social policies or macro-level interventions. Yet this trend has generated controversy. Economists and social scientists have criticized medical journal articles over methodological practices[1,2], and some observers have argued that medical journals are ill-equipped to referee social science research.[3] Accusations of "dubious social science"[1] have been leveled at a variety of medical journals, including top-tier ones. These disputes raise general questions about the frequency and characteristics of methodological problems and the prevalence and severity of research bias and error.

As a first step towards addressing this controversy, we use letters published in *The Lancet* to identify common criticisms of quantitative social science research published in medical journals and seek strategies to address them. To some extent, these letters can be seen as a "crowd-sourced" (but editor-gated) approach to peer review of published articles, from which some characteristics of perceived bias and error can be gleaned. Thus the methodological issues raised by published letters are those that have "slipped through" the formal peer review process, and can therefore inform recommendations for strengthening peer review in medical publishing of social science research.

While many medical journals include published correspondence sections, we chose *The Lancet* as an appropriate and instructive case study for analysis. As a top-tier medical journal that frequently publishes social science research, *The Lancet* has been central to the emerging debate on interdisciplinary studies and peer review. Unlike many lower-tier journals, *The Lancet* is fortunate to have a large pool of high-quality submissions and distinguished peer reviewers, both of which are critical to assessing the rigor of interdisciplinary articles. *The Lancet*'s prestige also invites greater scrutiny; as a premier journal, its published articles have disproportionate impact on policy and research, receive more extensive news coverage, and are more likely to be cited by researchers. For these reasons, its articles are also more likely to attract the attention of critical readers, who can voice their concerns within *The Lancet*'s correspondence section.

By analyzing correspondence letters, this paper identifies the more common criticisms associated with studies using quantitative social science methods that are published in *The Lancet*. We use this information to reflect on differences between the peer review processes

of medical and economics journals, and we conclude with recommendations for strengthening the processes of inquiry, debate, publication, and learning.

II. Methods

To study criticism of social science studies published in medical journals, we generated two datasets: one for content and one for context. The content dataset was used to investigate the nature of criticism published in letters. The context dataset was used to assess the relative frequency with which different kinds of articles were the subject of letters.

The content dataset comprises 139 letters and the 53 original research articles associated with them. To generate this dataset we used the online version of *The Lancet*[4] to systematically identify and read relevant correspondence in each issue published between 2008 and 2012. For each letter within the correspondence section, we determined (1) whether the letter expressed concerns about one or more articles previously published in *The Lancet* since 2008; (2) whether those concerns related to quantitative methods or their interpretation in the original article; and (3) whether the content of the original article was pertinent to our inquiry, i.e. using multi-disciplinary or econometric techniques to evaluate a non-clinical research question. This dataset was generated in December 2012 and corresponds to information available on the website at that time.

Next, we downloaded and reviewed all correspondence, noting (1) the number of replies generated by each paper; (2) whether the original authors responded to the critiques raised in the correspondence; (3) the points of dispute that were raised; and (4) basic characteristics of the article including the study type, the unit of observation, the number of observations or countries, the main outcome, and the quantity of interest. We then categorized the critiques raised by correspondence into one or more of seventeen distinct categories (listed in Table 2). We did not attempt to assess the validity of these critiques, but merely categorized and summarized them. Two reviewers separately reviewed and coded each correspondence letter.

We created a second dataset to provide context for these papers by assessing whether global health articles were more or less likely than other articles to be the subject of criticism in letters or comments. We used PubMed to extract citation data on all research articles, letters, and comments published in *The Lancet* between 2008 and 2012. This information was obtained in March 2014. All of the citations (including articles, letters, and comments) were classified as 'global health or social science' (GHSS) articles if they included at least one of

the following keywords: Global, World, Developing Countries, International, Nations, Middle East, Africa, Asia, America, Social, Economic, Political. The GHSS classification encompasses more than just quantitative social science studies in global health that are the subject of the content dataset; however, this classification provided a rough division of items between those that are likely to use quantitative social science methods and those that are not. We then used this context dataset to see whether GHSS articles were more or less likely than other articles to be the subject of criticism in letters.

III. Results

Analysis of criticism in correspondence letters

Our content analysis found 139 letters that met our inclusion criteria. These corresponded to 53 original research articles, representing an average of 2.6 letters per article (not including authors' responses) or 3.5 letters per article (when counting authors' replies) (Table 1).

A majority of the criticism in these letters pertained to non-randomized or observational studies rather than randomized trials. The authors of the original research article published their own letter responding to the critique in 91% of cases (Table 1).

Table 1: Characteristics of the Content Dataset

Number of Correspondence Letters (which met inclusion criteria)	139
Number of Research Articles Identified With Correspondence Letter(s)	53
Average Number of Letters (Excluding Authors' Reply)	2.6
Average Number of Letters (Including Authors' Reply)	3.5
% Where Author(s) of Research Article Replied to Letter Critiques	91%
Type of study (share of 53 articles)	
Cluster Randomized Trials	9%
Cross-Country (or similar)	32%
Cross-Sectional	15%
Prospective (or similar)	15%
Retrospective (or similar)	29%

Source: See text for explanation of the content dataset. All items were published in *The Lancet* between 2008 and 2012.

Table 2 shows the distribution of methodological criticism for quantitative social science studies published in *The Lancet*'s between 2008 and 2012. The five concerns most frequently cited in letters are: measurement error (51% of papers); omitted variables or confounding (45%); implausibility or lack of external validity (43%); missing or low-quality data (32%); and lack of transparency of methods (30%).

Table 2: Prevalence of Methodological Criticism in Quantitative Social Science Studies Published in *The Lancet*, 2008-2012

Category of Methodological Dispute	Number	Percentage
Measurement Error	27	51%
Omitted Variable & Confounding	24	45%
Implausibility or External Validity	23	43%
Missing or Low-Quality Data	17	32%
Transparency	16	30%
Misinterpretation or Overstatement	14	26%
Reverse Causation & Dynamics	7	13%
Randomization Failure	6	11%
Calculation Error	6	11%
Not Robust to Alternative Specifications	5	9%
Clusters and/or Weighting	4	8%
Attrition/ Loss to Follow Up	3	6%
Invalid Instruments	2	4%
Autocorrelation	2	4%
Hawthorne Effect	2	4%
Conflict of Interest	2	4%
Spillovers	1	2%

Source: Authors' calculations from content dataset described in text, comprising 53 articles.
Note: Percentages do not sum to 100% because each study can have multiple critiques.

In Table 3, we further disaggregate our results by type of study. Among cross-country time-series studies, disputes relating to instruments, autocorrelation, and other issues of dynamics and reverse causation are relatively common. Separately, whereas studies using individual data were regular about reporting their sample size, cross-country studies were irregular in reporting the total number of country or country-time observations used (only 6 of 17). The data used to generate these tables accompanies this paper. This list of methodological critiques is familiar to anyone engaged in social science quantitative and econometric research.

Table 3: Frequency of Methodological Disputes in Identified Studies (of 53 Total Studies)

Study Type / Methodological Critique	Cluster Randomized Trials		Cross-Country (or Similar)		Cross-Sectional		Prospective (or Similar)		Retrospective (or Similar)		All Studies	
	N	%	N	%	N	%	N	%	N	%	N	%
Omitted Variable & Confounding	2	40%	4	24%	5	63%	3	38%	10	67%	24	45%
Reverse Causation & Dynamics	2	40%	2	12%	1	13%	2	25%			7	13%
Invalid Instruments	1	20%	1	6%							2	4%
Autocorrelation	1	20%	1	6%							2	4%
Clusters and/or Weighting	2	40%	1	6%	1	13%					4	8%
Spillovers	1	20%									1	2%
Attrition/ Loss to Follow-Up	2	40%					1	13%			3	6%
Measurement Error	1	20%	11	65%	3	38%	4	50%	8	53%	27	51%
Hawthorne Effect	2	40%									2	4%
Missing or Low-Quality Data	2	40%	9	53%	1	13%	1	13%	4	27%	17	32%
Randomization Failure	2	40%			1	13%			3	20%	6	11%
Transparency	3	60%	5	29%	1	13%	4	50%	3	20%	16	30%
Implausibility or External Validity	2	40%	9	53%	4	50%	4	50%	4	27%	23	43%
Calculation Error	1	20%	3	18%	1	13%			1	7%	6	11%
Not Robust to Alternative Specifications	1	20%	1	6%			1	13%	2	13%	5	9%
Misinterpretation or Overstatement of Results	1	20%	3	18%	1	13%	5	63%	4	27%	14	26%
Conflict of Interest	1	20%					1	13%			2	4%
Total Papers by Study Type	5		17		8		8		15		53	

Source: Authors' calculations from content dataset described in text, comprising 53 articles.

Notes: N refers to the number of papers with the methodological critique. Each paper can receive multiple critiques and hence numbers do not sum to 100%. The % refers to the percentage of papers of that study type with that methodological critique (N divided by the Total Papers by Study Type).

Do social science articles generate more controversy?

Although the content analysis focused specifically on articles using quantitative social science methods in global health, we wanted to assess whether these articles were more or less likely to be the subject of criticism in letters and comments than other kinds of articles published in the journal. Thus, the context dataset provides information for making a *rough* judgment on this question, because the selection of articles is less precise (we used a keyword identification process rather than reading the articles). We contend that articles classified as 'GHSS' are more likely to have used quantitative social science methods than those that did not. However, we recognize that this method will generate a substantial number of 'false positives' and 'false negatives', which will probably bias our estimates of differences toward zero.

The context dataset, generated from PubMed, contains citations for 7529 items in *The Lancet* between 2008 and 2012. Of these, 3095 were letters (most about 250 words) or comments (around 700 words), and the remaining 4434 were original research articles. Among these research articles, 1337 were classified as GHSS (30.2%). Among the 3095 correspondence letters, 21% were classified as GHSS.

GHSS articles appear to be less likely to receive comments in letters than other articles. Among GHSS articles, 20% were the subject of at least one letter or comment compared to 25% for other articles (two-sample t-test that difference is not equal to zero, p=0.0001). Looking at the volume of commentary, GHSS articles averaged 0.35 letters and comments compared to 0.46 for non-GHSS articles (including authors' replies) (p=0.0007).

IV. Discussion and Recommendations

To our knowledge, this is the first study that has used letters published in journals as a unit of analysis. This information provides new perspectives on methodological controversies and an opportunity to reach more objective assessments of problems that arise in the review and publication process. Nevertheless, using published letters also entails a number of important qualifications.

First, and most importantly, letters do not provide a representative sample of scholarly opinion. To be published in a medical journal's correspondence section, potential critics must read the original research article; formulate and articulate their views; choose to

formalize and submit a letter; and have the letter selected for publication by the journals' editors. More passionate readers are far more likely to write critiques, and the journal itself selectively publishes letters based on editors' judgments regarding the significance of the criticism, methodological validity, general interest, authors' prestige and profile, and probably the journal's reputation. Though the potential for bias at each step is clear, the direction of the net bias is not. A second qualification is that we accepted the criticisms published in letters at face value and did not independently assess their validity. A third qualification is that we may have misclassified letters and articles. In the contextual analysis, the classification we applied is only as good as our selection of keywords. By contrast, in the content analysis, we sought to minimize classification error by having two reviewers categorize each article.

As a final qualification, we want to prevent misinterpretation by making clear that this study makes no claims about the differences between medical and economics journals in terms of the quality of research or the prevalence of methodological problems. Our data simply cannot address that question. Most economics journals do not include correspondence sections, making it impossible, on the basis of analyzing letters, to make such a judgment.

Our main findings relate to the relative frequency of controversy for GHSS articles and the prevalence of methodological issues raised by critics. First, GHSS studies appear to generate somewhat less criticism in published letters than other studies published in the same medical journal. Second, among published letters that met our inclusion criteria, the most common concerns raised by correspondents involved measurement error; omitted variables and confounding; implausibility and lack of external validity; missing or low-quality data; and lack of transparency regarding methods.

These methodological concerns would be less likely to arise if people with appropriate expertise were involved in review, if working paper versions were available for public comment and if data and associated computer code were in the public domain. Prior to publication, authors could be alerted to problems that they have overlooked when making presentations at conferences, receiving comments on working papers or getting reviewers' feedback. This pre-publication dialogue provides opportunities to assess the validity of the concerns and the robustness of findings prior to final publication.

Other criticism raised in letters – for example, the Hawthorne effect, invalid instruments, and reverse causation – are unlikely to be affected by simple transparency policies, though

some could be ameliorated through registration of research designs. We interpret this list as providing guidance in setting priorities for strengthening peer review for which we make recommendations that fall into two categories: (1) improving the review process itself; and (2) encouraging earlier and more open channels for public debate.

To improve the review process for quantitative social science research, journals can ensure that reviewers with the appropriate expertise are engaged. Papers utilizing econometrics would clearly benefit from peer review by individuals with econometric (and not only statistical) expertise, especially if it were possible for them to check computer codes and replicate results. Medical journals might even consider incorporating a dedicated economics editor to help recruit qualified referees and evaluate reviews if that expertise is lacking on the editorial board.

The review process could also be improved by adopting clear standards for quantitative social science research. Whereas the medical field has clear guidelines for reporting randomized trials (e.g. CONSORT), such criteria within the social sciences tend not to be formalized. In particular, accepted norms for econometric reporting are not explicitly codified and are thus not available to reviewers of econometric papers in medical (or economic) journals. Thus, one approach would be to form a CONSORT-like committee to collectively compose guidelines on the reporting standards for cross-country time-series analysis, regression discontinuity, or other types of econometric designs, which could then be distributed to authors and reviewers in medical and economics journals alike. One early effort in this vein has been the Gates Reference Case, a "principle-based standardized methodology for good practice in the planning, conduct and reporting of economic evaluation…", recently launched by the UK's NICE International.[5]

In the interim, however, journals could identify articles in particular methodological categories where standards and recommendations have been codified. For example, a standardized checklist for generalized method of moments panel studies is possible based on reporting standards proposed in Roodman (2009).[6]

Medical journals could also improve the quality of the quantitative social science research they publish by encouraging earlier and more open channels for public debate. In this regard, the medical field may have something to learn from economics journals. In medicine, papers are never released before publication in a medical journal. In contrast, economics papers are frequently published as working papers and almost always online (e.g. National Bureau of

Economic Research). Such papers are widely circulated among peers, often in public, and subjected to revisions before final publication in a journal.

This more open approach promotes transparency and engages more people in review, whereas the medical approach precludes discussion prior to publication. There are exceptions in cases where *ex ante* protocol registration is required (as with clinical trials and systematic reviews), but this has not been applied as a standard requirement for quantitative social science studies. The longer a paper is circulated and among more people, the more likely that errors will be found and corrected. In this way, greater openness can improve scientific rigor and integrity.

This open approach is not without its problems. Double-blind peer review is less possible when working paper versions are publicly available (though in highly specialized fields, true double-blinding is difficult even under the best of circumstances anyway). Another concern is that openness puts findings into circulation where they may be widely cited, even if refuted by a later version, and that citations of ungated working papers may persist after the eventual publication of a gated final version.[7] However, cloaking work until it is peer-reviewed by a journal clearly does not assure that published articles are free of errors.[8] On balance, we believe the advantages of openness to uncover errors and facilitate review outweigh the disadvantages that have been raised. Thus even if medical research culture does not currently include prepublication distribution, it is not obvious why medical journals, including *The Lancet,* impose a ban on submissions that were previously published as working papers – especially in the case of work from researchers in non-medical disciplines. Accordingly, we recommend allowing the publication of studies with prior working paper versions or providing a compelling justification for how this policy furthers the science.

So long as medical journals continue to prohibit circulation of articles in draft versions prior to publication, ensuring the quality of these studies rests almost entirely on the journal's own peer review process, and letters will be the main channel for the public to express criticism. If so, journals can improve the integrity of the evidence base by making better use of these letters. Correspondence letters formalize a critique that (in most cases) requires the authors' response (with research article authors responding to letters 91% of the time in this sample), but authors are rarely compelled to correct or revise their papers. Indeed, letters rarely lead to changes in the paper or admission of error, and it is not immediately clear what purpose the letter serves beyond raising suspicions on paper quality. Further, original articles are rarely organized online in such a way as to alert readers to all relevant correspondence. Thus

researchers may cite or use study results without understanding the caveats raised by published correspondence. Requiring corrections, when appropriate, and closely linking all correspondence to articles when they are searched would help avoid such errors.

Medical journals should also consider adopting requirements that quantitative social science articles make their associated data and computer code publicly available. This standard is increasingly being adopted by leading economic journals, such as the *American Economic Review*.[9] This approach tends to improve research quality by imposing a discipline on researchers themselves that helps them to identify errors. It improves peer review by giving reviewers an opportunity to replicate estimates and test robustness. It also furthers the development of knowledge by giving later researchers an opportunity to re-assess data in light of new evidence or methods, as well as providing an excellent learning opportunity for students. As one of the world's leading medical journals, *The Lancet* in particular has an opportunity to pioneer the improvement of interdisciplinary work by embracing such a requirement for publication of data and computer code. Replicability, after all, is a *sine qua non* of science.

As a final reflection, we note that economists and other social scientists have much to learn from the medical field about improving research quality. For example, in recent years social scientists themselves have increasingly embraced randomized control trials (RCTs) without necessarily learning about best practices pioneered by medical researchers. A forthcoming review by Eble and Boone finds that RCTs published in top economics journals "fall far short of the recommendations for reporting and conduct put forth in the medical literature… suggesting risk of exaggerated treatment effects in the economics literature"[10] and the economics literature has only recently addressed the importance of registering trials[11]. Research in both fields, however, suffers from low compliance with voluntary registration.[12]

In sum, this paper argues that medical journals that publish social science research should adopt review processes appropriate to the methodologies and disciplines of those studies and to encourage public debate earlier and more systematically. We recommend the use of relevant checklists and guidelines to address the most common methodological problems that arise in quantitative social science studies, and recruitment of reviewers with specific expertise in econometrics to facilitate appropriate review. We recommend that medical journals adopt standards requiring that data and computer code are placed in the public domain.

V. References

1. Easterly W, Freschi L (2011). Development social science in medical journals: diagnosis is caveat emptor. Aid Watch. Accessed 20 March 2013 at http://aidwatchers.com/2011/05/from-shaky-research-into-solid-headlines-via-medical-journals/
2. Ozler B (2013). Study design and sub-group analysis in HIV trials. Development Impact. Accessed 20 March 2013 at http://blogs.worldbank.org/impactevaluations/study-design-and-sub-group-analysis-in-hiv-trials
3. Wagstaff A (2012). Should you trust a medical journal? Let's Talk Development. Accessed 20 March 2013 at http://blogs.worldbank.org/developmenttalk/should-you-trust-a-medical-journal-0
4. The Lancet (2008-2012). Assortment of article and commentary published in the journal. Accessed online between December and February 2013. Full list of over 150 included studies and correspondence [see Web Appendix 2].
5. NICE International (2014). Bill and Melinda Gates Foundation Methods for Economic Evaluation Project (MEEP) Final Report. A partnership between Bill and Melinda Gates Foundation, NICE International, the Health Intervention and Technology Assessment Program (Thailand), and the University of York, Centre for Health Economics. Accessed 16 June 2014 at http://www.nice.org.uk/aboutnice/niceinternational/projects/MethodsForEconomicEvaluationProjectAndTheGatesReferenceCase.jsp?domedia=1&mid=6788DC7D-F05F-BEEB-52ED184BE008E0C6
6. Roodman, D (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. Stata Journal, StataCorp LP, 9(1): 86-136, March.
7. Ozler B (2011). Working papers are NOT working. Development Impact. Accessed 21 March 2013 at http://blogs.worldbank.org/impactevaluations/working-papers-are-not-working
8. Godlee, F., Gale, C. R., & Martyn, C. N. (1998). Effect on the quality of peer review of blinding reviewers and asking them to sign their reports. JAMA: the journal of the American Medical Association, 280(3), 237-240.
9. The American Economic Review: Data Availability Policy. Website. Accessed 4 October 2013. http://www.aeaweb.org/aer/data.php
10. Eble A, Boone P. (2012) Risk and evidence of bias in randomized controlled trials in economics. Mimeo, 2013, Brown University.
11. American Economics Association Randomized Control Trial Registry. Website. Accessed 4 October 2013. https://www.socialscienceregistry.org
12. All Trials Registered | All Results Reported. Website. Accessed 4 October 2013. http://www.alltrials.net