

BMJ Open Does a research article's country of origin affect perception of its quality and relevance? A national trial of US public health researchers

M Harris,¹ J Macinko,² G Jimenez,³ M Mahfoud,⁴ C Anderson⁵

To cite: Harris M, Macinko J, Jimenez G, *et al.* Does a research article's country of origin affect perception of its quality and relevance? A national trial of US public health researchers. *BMJ Open* 2015;**5**:e008993. doi:10.1136/bmjopen-2015-008993

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-008993>).

Received 5 June 2015

Revised 5 November 2015

Accepted 10 November 2015



CrossMark

For numbered affiliations see end of article.

Correspondence to

Dr M Harris;
m.harris@imperial.ac.uk

ABSTRACT

Objectives: The source of research may influence one's interpretation of it in either negative or positive ways, however, there are no robust experiments to determine how source impacts on one's judgment of the research article. We determine the impact of source on respondents' assessment of the quality and relevance of selected research abstracts.

Design: Web-based survey design using four healthcare research abstracts previously published and included in Cochrane Reviews.

Setting: All Council on the Education of Public Health-accredited Schools and Programmes of Public Health in the USA.

Participants: 899 core faculty members (full, associate and assistant professors)

Intervention: Each of the four abstracts appeared with a high-income source half of the time, and low-income source half of the time. Participants each reviewed the same four abstracts, but were randomly allocated to receive two abstracts with high-income source, and two abstracts with low-income source, allowing for within-abstract comparison of quality and relevance

Primary outcome measures: Within-abstract comparison of participants' rating scores on two measures—strength of the evidence, and likelihood of referral to a peer (1–10 rating scale). OR was calculated using a generalised ordered logit model adjusting for sociodemographic covariates.

Results: Participants who received high income country source abstracts were equal in all known characteristics to the participants who received the abstracts with low income country sources. For one of the four abstracts (a randomised, controlled trial of a pharmaceutical intervention), likelihood of referral to a peer was greater if the source was a high income country (OR 1.28, 1.02 to 1.62, $p<0.05$).

Conclusions: All things being equal, in one of the four abstracts, the respondents were influenced by a high-income source in their rating of research abstracts. More research may be needed to explore how the origin of a research article may lead to stereotype activation and application in research evaluation.

Strengths and limitations of this study

- First study at the national level in the USA, to determine the impact of country-of-origin on the rating of healthcare research abstracts.
- All core faculty members (full, associate and assistant professors) of every Council on the Education of Public Health-accredited Schools and Programmes of Public Health in the USA were invited to participate in the study.
- Participants were blinded to the purpose of the study and randomised to receive high-income or low-income source abstracts.
- Abstracts were rated on strength of the evidence and likelihood of referral to a peer.
- Although 899 full, associate and assistant professors participated in the study, this corresponded to a 9.8% response rate.

BACKGROUND

Ideally, research findings ought to be judged on the strength of the evidence and their relevance. However, there is subjectivity involved in interpreting research.¹ Research certainly does not 'speak for itself'—we give it a voice—and how we judge whether one piece of research constitutes evidence or not is complex and messy. Common standards for assessing the internal validity of research do not account for the potential cognitive biases in the consumption and interpretation of research *post*publication, and each of us may reach a different conclusion as to whether the research presents strong evidence and whether we consider the research useful. In practice, we see many idiosyncracies. A rigorous randomised controlled trial (RCT) may convince a surgeon to change a certain practice, but may not have the same effect on a primary care physician.² Government regulators consider the reliability (the degree to which an innovation is communicated as being consistent in its

results) of an innovation more positively than do industrial scientists.³ Clinicians are more likely to adopt an innovation if they believe it has come from current users with similar professional, cultural and socioeconomic backgrounds.⁴ A legitimate source is important for innovation diffusion,^{5 6} but little is known about how legitimacy is defined or perceived. From the marketing literature, Bilkey and Nes⁵ showed that consumers tend to rate products from their own countries more favourably and that consumer preferences are positively correlated with the degree of economic development of the source country, probably evoked by the lower price cue of low income country products.⁷ Up to 30% of the variance of consumer product ratings can be attributed to the product's country-of-origin.⁸

In healthcare research, typically, the first pieces of information to be provided in a research article are the author's name, and the institution and country the research has been conducted in. Understanding anchoring to be a feature of heuristic thought,^{9–13} it follows that we should examine the extent to which the source affects our interpretation of that research. If one possesses a prior-held belief or attitude towards the source, how does this influence one's subsequent view of the research? All things being equal, would research conducted in Ethiopia be viewed in the same way as identical research conducted in the USA?¹⁴

The income and development level of the source country certainly seems to determine whether a manuscript is selected for publication.¹⁵ The number of publications from low income countries is significantly lower than the number from developed countries, in various research fields.^{15 16} In psychiatry, only 6% of the literature is published from regions that represent 90% of the global population.¹⁷ Similar under-representation exists in cardiology, HIV research and epidemiology.^{18 19} One argument for this is that research from low income countries (LICs) lacks the quality to meet publication criteria.²⁰ Others argue that there are systematic selection biases. Editorial board members of international biomedical journals are more likely to come from high income countries (HICs).^{21–23} Reviewers from OECD (Organization for Economic Cooperation and Development) countries view articles from their own country more favourably than they do articles from other countries.^{22 24 25} Studies recruiting participants from the USA are more likely to be published.^{21 23} In Peters and Ceci's²⁶ controversial experiment, only one of the nine articles that were initially published in a highly regarded American journal was accepted on resubmission to the same journal after fabricating the name of the original institutions. Kliever *et al*²⁷ demonstrated that articles from outside of North America were less likely to be accepted for publication. It seems that source matters.

The major obstacle to this research question is that there are no controlled studies to ascertain the impact of the source of the research *post*-publication. To fill this

research gap, we present here a randomised trial of Public Health research faculty in the USA. This national survey invites respondents, most of whom are experienced healthcare researchers and peer reviewers, to rate identical, typical healthcare research abstracts. To ascertain the impact of the source (institution and country) of the abstracts, we ensured that the abstracts that the respondents received were identical in every respect, however, we fictionalised the sources into either HIC or LIC and randomised the respondents to receive either type. We then compared their responses to two simple questions for each abstract—whether they think the evidence in the abstract is strong, and whether they would recommend the abstract to a peer. Under the null hypothesis, there should be no difference in the distribution of responses to the two types of abstract.

METHODS

Survey design

We used a web-based survey using a Qualtrics survey platform. The survey was divided into two sections, the first to collect demographic and professional data, and the second for the respondent to read and respond to four research abstracts. Each abstract was followed by the same two questions—first, how strong is the evidence presented in this abstract? And second, how likely are you to recommend this abstract to a colleague? Responses were on a scale (1–10) with 1 as the least (ie, not at all strong, not at all likely) and 10 as the most (extremely strong, extremely likely). The time taken to read and respond to each abstract was measured by the survey platform. Each question was forced response to avoid the problem of missing data. Recipients were randomly allocated to one of two possible surveys. In the first, abstracts 1 and 4 were fictionalised to HIC sources (UK and Germany) and abstracts 2 and 3 were fictionalised to LIC sources (Malawi and Ethiopia). These sources were reversed in the second survey. Therefore, each survey (survey A and survey B) had two abstracts from LIC sources and two from HIC sources (figure 1).

In order to ensure that the abstracts were of sufficient quality and internal validity, we purposively selected abstracts of papers that had been included in Cochrane Reviews and that were also likely to be of at least some interest to most public health academics and health service researchers. Each abstract had therefore already been vetted for sources of bias prior to publication, using the Cochrane risk of bias tool, and we only selected abstracts that had a high internal validity for the type of study that it was describing. There is a trade off between choosing abstracts of interest to all potential respondents and the length of the survey. We decided to choose four abstracts—one randomised controlled pharmaceutical trial, one randomised controlled service intervention, one pharmaceutical intervention of cross-sectional design and one service intervention of cross-sectional design—to give a balance in terms of content

	Abstract 1	Abstract 2	Abstract 3	Abstract 4
Original article title	Lay health worker intervention with choice of DOT superior to standard TB care for farm dwellers in South Africa: a cluster randomized control trial.	The use of routine monitoring and evaluation systems to assess a referral model of family planning and HIV service integration in Nigeria.	C-reactive protein lowering with rosuvastatin in the METEOR study.	Profiles of self-reported HIV-risk behaviors among injection drug users in methadone maintenance treatment, detoxification, and needle exchange programs.
Original first author	Clarke M	Chabikuli NO	Peters SA	Mark HD
Original journal	International Journal of Tuberculosis and Lung Disease	AIDS	Journal of Internal Medicine	Public Health Nursing
Original source	Sweden	Nigeria	The Netherlands	USA
Source Cochrane Review	Lay health workers in primary and community health care for maternal and child health and the management of infectious diseases	Integration of HIV/AIDS services with maternal, neonatal and child health, nutrition, and family planning services	Statins for the primary prevention of cardiovascular disease	Oral substitution treatment of injecting opioid users for prevention of HIV infection
Year	2005	2009	2010	2006
Degree of internal validity from Cochrane Review	+++++	++++	+++++	+++++
Fictional Source Survey A	Faculty of Medicine, University of Freiburg, Freiburg, Germany RSH	College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia XSL	Faculty of Health Sciences, Mzuzu University, Mzuzu, Malawi RPL	Department of Public Health and Primary Care, Oxford University, Oxford, United Kingdom XPH
Fictional Source Survey B	College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia RSL	Faculty of Medicine, University of Freiburg, Freiburg, Germany XSH	Department of Public Health and Primary Care, Oxford University, Oxford, United Kingdom RPH	Faculty of Health Sciences, Mzuzu University, Mzuzu, Malawi XPL

*R=Randomized Controlled Trial; X=Cross-sectional design; S=Service delivery; P=Pharmaceutical; H=High Income; L=Low Income

Figure 1 List of abstracts used in the survey and the fictionalised sources and institutions.

and design. All four abstracts were of similar length and complexity. The abstracts were presented as found in their PubMed format, with all technical content preserved and in a format familiar to any healthcare researcher, however, for each abstract, the institution and country of origin was fictionalised to one of four different high-income or low-income sources. For one abstract, the trial acronym was removed to avoid the possibility that some respondents would recognise the research. High-income source countries were selected from the top 10 countries by gross domestic product (GDP) per capita (>\$36 000 per capita), and OECD membership. Low-income source countries were selected from the bottom 10 countries by GDP per capita (<\$1046 per capita). The institutional affiliation was fictionalised to one of the top-five universities that also had a medical or healthcare faculty, in the respective countries. We used the 2014 Times Higher Education World rankings (<http://www.timeshighereducation.co.uk/world-university-rankings/2014-15/world-ranking>) for the HIC sources, and the <http://www.4icu.org> website for international rankings of institutions for the low-income sources.

We ensured that the source of the abstract was equally visible in each abstract and was mentioned in at least three locations throughout the abstract—the title, under the title and in the abstract itself. To avoid a possible order effect, the order in which the abstracts were presented in the survey was randomised for each participant. Neither the original nor the fictionalised journals were included in the source in order to avoid respondents reacting to the reputation of the publication type. Furthermore, in order to not influence the responses,

the survey was described as a Speed Reading survey, designed to examine whether the time taken to read an abstract influences the interpretation of the information within it. The survey platform enabled us to measure the time taken to respond to the entire survey, and each abstract, and this information was provided to the respondent at the end of the survey to heighten the ‘psychological realism’ of the survey. The survey was pilot-tested with Masters in Public Health students at Imperial College London and some faculty members at New York University, to ensure face validity of the questions and also to ensure that the design and flow of the survey were straightforward.

Participants and survey management

We included all core faculty members of Schools and Programmes of Public Health—located in a US state—that had publically available contact information and that were accredited by the Council on the Education of Public Health (CEPH—<http://ceph.org/accredited>) (159 institutions; see online supplementary appendix 1 for full listing). We excluded administrators, managers, adjunct faculty members and visiting faculty members, and faculty members from our own institution. From this universe of potential respondents (n=9421 once duplicates were removed), we randomised them to receive either survey A or survey B and sent them an invite to take the survey. Block randomisation within respective institutions was used, with four, six and eight sequences, from a web-based randomisation service (<http://www.sealedenvelope.com>, seed 13752665 5595533).

The survey was designed so that only the email recipient could open the link to the survey and that it could be taken only one time. The survey could not be sent anonymously, and was inaccessible to search engines. The survey was active only within the specified time frame (20 January–4 February 2015, chosen so that faculty members were highly likely to be present at their institution), and two email reminders were sent on day 7 and day 14 following the first email invite (20 January 2015). Panel members did not receive prior invitation to participate in the survey, however, our email invite indicated clearly that all responses were to be de-identified and analysed in aggregate form only, and solely for the purposes of this research. It also indicated that there was no obligation to participate but by choosing to participate consent to use the response for research is implied. We offered participants entry into a lottery draw for a \$500 Amazon voucher as an incentive to complete the survey. The study protocol, including the non-harmful deception around the ulterior motive of the study, was reviewed by the New York University Committee on Activities Involving Human Subjects and deemed exempt from full ethical review (#14-10332).

Statistical analysis and power calculation

Data were retrieved via Qualtrics in CSV format and analysed using Stata/SE V.13 (Statacorp, College Station, Texas, USA). We used demographic covariates (age, sex), professional experience covariates (research exposure, peer review experience, educational attainment) and institutional covariates (region, CEPH accreditation type and Ivy League status) to explain variation in the outcomes of interest. We grouped respondent age into categories based on a presumed mid-year birth and survey completion date of 31 January 2015. Educational

attainment was categorised into two groups—Academic and Clinical Academic—based on the completed qualifications provided in the survey responses. We used a generalised ordered logit model for the multivariable analysis and two-tailed t tests to compare the differences in mean responses as well as for the descriptive characteristics of the survey samples. We also explored high and low cut points for the outcome variables in bivariate analysis and illustrate the distribution of scores as proportions of respondents at the high (≥ 8) and low (≤ 3) ends of the distribution, using a univariate logistic regression model containing the binary outcome (ie, above/below a certain threshold) and a binary indicator of the abstract's country of origin. The corresponding test is a Wald test of the β coefficient for the abstract country of origin.

We calculated that sample sizes of 400 respondents for each survey would provide enough power (80%) to detect a statistically significant (95% confidence level) difference of 0.35 in mean scores between the two groups.²⁸

RESULTS

After randomisation, 4711 potential respondents received email-invites for survey A, and 4710 received email-invites for survey B. Fifty-one and 61 invitations bounced, respectively. A total of 567 started survey A and 594 started survey B. Of these, 433 completed survey A and 466 completed survey B. This corresponds to a response rate of 9.2% for survey A and 9.9% for survey B. Institutional characteristics (region and Ivy League representation) of responders and invitees were not significantly different, although there was a small over-representation of responders from CEPH-accredited Programmes in Public Health. The demographic

Table 1 Respondent characteristics for survey A and survey B

	All respondents (n=899)	Survey A (n=433)	Survey B (n=466)
Males, %	42.05	42.49	41.63
Age, mean	50.26	50.35	50.17
Academic credentials only*, %	84.58	84.69	84.48
Clinical credentials†, %	15.42	15.31	15.52
US born‡, %	81.65	82.68	80.69
Reads research daily§, %	60.07	61.20	59.01
CEPH Programme of Public Health¶, %	35.48	34.64	36.27
Ivy League university**, %	12.46	12.93	12.02
Region Northeast %	28.03	26.79	29.18
South %	42.05	43.42	40.77
Midwest %	18.24	17.32	19.1
West %	11.68	12.47	10.94

*For example, BSc, BA, MSc, MPH, PhD.

†For example, MD, MBBS, MBChB.

‡Versus non-US born.

§Versus reads research less than daily.

¶Versus CEPH School of Public Health.

**Versus non-Ivy League institution.

CEPH, Council on the Education of Public Health.

characteristics of the respondents of both surveys were equal, suggesting that randomisation performed as was expected (table 1). Ninety per cent of respondents of both survey types serve as peer reviewers for academic journals.

On average, respondents spent between 72.5 and 109.9 s on each abstract with no significant differences between the groups. table 2 shows the mean (SD) ratings for strength and referral for the four abstracts by the type of source. Referral to a peer for abstract 3 (randomised controlled trial of a pharmaceutical intervention) was significantly more likely if the source was from a HIC. There were no other significant differences between the abstracts based on the source. The findings were unchanged when using a proportion rating higher than eight or lower than three. As might be expected, strength rating for abstracts that described a more robust research design, specifically RCTs (abstract 1 and 3), scored higher for strength than abstracts 2 and 4, which were of a cross-sectional design. Also, as might be expected, the disposal of these abstracts also correlated well with respondents' view of the strength of the evidence contained within them. Correlation between the scores given for strength of evidence and subsequent referral was high (Spearman correlation coefficients varied between 0.71 and 0.85).

Tables 3 and 4 show the results of the multivariable analysis. Controlling for individual and institutional covariates, high-income source was a significant predictor of referral for abstract 3 only (OR 1.28, 1.02 to 1.62). For some abstracts, the time spent reviewing the abstract was negatively associated with the rating given to it for strength of evidence (abstract 1 OR 0.49, 0.34 to 0.71; abstract 3 OR 0.65, 0.46 to 0.92) or referral to a peer (abstract 1 OR 0.50, 0.35 to 0.72; abstract 2 OR 0.61, 0.44 to 0.84; abstract 3 OR 0.66, 0.44 to 0.84). However, rating for abstract 4 (both strength of evidence (OR 1.63, 1.06 to 2.51) and referral to a peer (OR 1.55, 1.01 to 2.38) improved when more time was spent on it. Individuals affiliated to CEPH Programmes of Public Health were significantly more likely to rate the strength of the evidence for this abstract higher (OR 1.38, 1.07 to 1.78) and to refer it to colleagues (OR 1.67, 1.30 to 2.15) than individuals affiliated to Schools of Public Health.

DISCUSSION

Two sinister issues may be occurring if the source of the research affects one's judgement of it. First, poor research may be given undue significance in part because of the perceived legitimacy of its source. The MMR scandal in the UK may have been a painful example of this. In this case, vaccination rates for the MMR immunisation plummeted when a study published by a high profile research group in a prestigious journal claimed a tenuous (and later discredited) connection between the immunisation and rates of autism.²⁹

Table 2 Abstract rating for strength and referral

Source	Abstract 1			Abstract 2			Abstract 3			Abstract 4		
	High	Low	All	High	Low	All	High	Low	All	High	Low	All
Strength												
Mean (SD)	5.77 (2.30)	5.78 (2.11)	5.77 (2.20)	4.92 (1.95)	4.90 (2.04)	4.91 (1.99)	6.92 (2.02)	6.76 (2.03)	6.84 (2.02)	3.95 (2.14)	4.05 (2.06)	4.00 (2.10)
≥8 (%)	27.61	24.78	26.15	10.13	12.06	11.06	47.63	43.16	45.47	6.96	4.74	5.81
≤3 (%)	22.04	18.10	20.00	27.59	30.63	29.05	8.19	9.05	8.60	48.49	45.91	47.15
Referral												
Mean (SD)	5.14 (2.54)	5.38 (2.36)	5.27 (2.45)	4.50 (2.21)	4.56 (2.26)	4.53 (2.23)	6.05* (2.40)	5.68 (2.45)	5.87 (2.43)	3.79 (2.23)	3.96 (2.21)	3.88 (2.22)
≥8 (%)	21.58	23.71	22.68	10.34	11.60	10.95	32.97	27.61	30.39	7.66	7.33	7.49
≤3 (%)	30.63	24.78	27.60	36.64	37.35	36.98	17.46	21.81	19.55	51.74	46.77	49.16
Mean time(s) (SD)	87.4 (68.4)	87.4 (118)	87.4 (97.3)	109.9 (169)	103.0 (200)	106.2 (186)	109.8 (131)	97.3 (304)	103.8 (237)	72.5 (56.4)	79.4 (146)	76.0 (112)

*p<0.05.
Abstract 1, RCT/Service; Abstract 2, Cross-sectional/Service; Abstract 3, RCT/Pharmaceutical; Abstract 4, Cross-sectional/Pharmaceutical.

Table 3 Predictors of abstract strength ratings †

	Abstract 1 OR 95% CI	Abstract 2 OR 95% CI	Abstract 3 OR 95% CI	Abstract 4 OR 95% CI
High vs low country origin	1.03 0.82 to 1.30	1 0.80 to 1.26	1.16 0.92 to 1.46	0.94 0.74 to 1.18
Male (vs female)	0.93 0.73 to 1.18	0.87 0.68 to 1.10	0.97 0.76 to 1.23	0.87 0.68 to 1.10
41–50 years (vs 21–40)	0.81 0.59 to 1.12	0.71* 0.52 to 0.97	1.01 0.73 to 1.39	0.87 0.64 to 1.20
51–60 years (vs 21–40)	0.79 0.57 to 1.09	0.74 0.54 to 1.03	1.14 0.82 to 1.58	0.77 0.56 to 1.06
61+ years (vs 21–40)	0.76 0.54 to 1.06	0.85 0.60 to 1.19	1.12 0.80 to 1.57	0.82 0.58 to 1.15
Clinical academic credentials (vs academic only)	0.83 0.60 to 1.14	0.65** 0.47 to 0.89	0.95 0.68 to 1.32	0.78 0.57 to 1.08
US born (vs not)	1.06 0.78 to 1.44	0.83 0.62 to 1.13	0.94 0.69 to 1.28	0.89 0.66 to 1.21
Reads research daily (vs <daily)	1.03 0.81 to 1.31	0.94 0.74 to 1.20	0.85 0.67 to 1.08	1.14 0.89 to 1.45
CEPH programme (vs school)	1.12 0.87 to 1.45	1.06 0.82 to 1.36	1.03 0.80 to 1.32	1.38* 1.07 to 1.78
Ivy League institution (vs others)	0.78 0.50 to 1.21	0.67 0.43 to 1.06	1.14 0.73 to 1.78	1.08 0.69 to 1.68
South region (vs Northeast)	0.71 0.50 to 1.00	1.08 0.77 to 1.52	0.84 0.59 to 1.18	1.05 0.74 to 1.47
Midwest region (vs Northeast)	0.82 0.55 to 1.23	1.17 0.78 to 1.74	1.14 0.76 to 1.71	1.07 0.72 to 1.59
West region (vs Northeast)	0.93 0.59 to 1.46	1.11 0.72 to 1.74	1.05 0.66 to 1.67	0.89 0.56 to 1.40
60–<120 s spent reading (vs <60 s)	0.67** 0.51 to 0.87	0.87 0.66 to 1.15	0.98 0.74 to 1.28	1.33* 1.04 to 1.70
120+ s spent reading (vs <60 s)	0.49*** 0.34 to 0.71	0.77 0.56 to 1.07	0.65* 0.46 to 0.92	1.63* 1.06 to 2.51
N‡	895	895	895	895

*p<0.05; **p<0.01; ***p<0.001.

†Generalised ordered logit model controlling for all variables in each column.

‡Only survey responses with no missing data included in the multivariate analysis.

Abstract 1, RCT/Service; Abstract 2, Cross-sectional/Service; Abstract 3, RCT/Pharmaceutical; Abstract 4, Cross-sectional/Pharmaceutical. RCT, randomised controlled trial.

Second, good research from an unexpected source may be discounted early on, resulting in missed opportunities to learn from important innovations. LICs have developed novel innovations and there are multiple opportunities to learn from LICs, for example, around improved surgical procedures,³⁰ improved long-term outcomes in mental illness,^{31–35} improved skill mix with scaled use of community health workers.^{36–38} However, there are strikingly few examples where these innovations have been adopted in HICs.³⁹ Even in Health Links, where HICs and LICs collaborate explicitly and reciprocally, there are surprisingly few examples of attempts to adopt LIC innovations in high-income settings—HIC volunteers learn a lot personally and professionally, however, this does not translate into changes in their own healthcare systems, and the learning and exchange of expertise is predominantly directed from the HICs towards the LICs.^{40–43} The Reverse Innovation ‘movement’ sets out to unpack the barriers to adopting LIC

innovations in HIC contexts. It is motivated in part by the rapidly changing global health landscape, and has gained interest in the USA and UK because the unsustainable growth in healthcare expenditure means that there is likely to be a genuine need to learn from LICs.⁴⁴

We know already from the Diffusion of Innovation literature that healthcare professionals perform poorly when it comes to adopting innovations or evidence from ‘elsewhere.’^{2 7} The not-invented-here culture prevails. However, we also know that innovations are more likely to diffuse if actors perceive the source to be similar to their own. Health professionals are homophilous.⁴ We might ask, therefore, whether health professionals are even more discriminating when presented with research from very ‘unlikely’ sources? Do they discriminate against sources that they might perceive to be so different from their own, or perceive to be so unlikely to produce good research, that the evidence is discounted early on?

Table 4 Predictors of abstract referral ratings†

	Abstract 1 OR 95% CI	Abstract 1 OR 95% CI	Abstract 3 OR 95% CI	Abstract 4 OR 95% CI
High vs low country origin	0.85 0.67 to 1.07	0.94 0.75 to 1.19	1.28* 1.02 to 1.62	0.9 0.71 to 1.13
Male (vs female)	0.95 0.75 to 1.20	0.78* 0.61 to 0.99	0.98 0.78 to 1.25	0.84 0.66 to 1.06
41–50 years (vs 21–40)	0.98 0.72 to 1.34	0.85 0.62 to 1.16	1.06 0.77 to 1.46	0.83 0.61 to 1.15
51–60 years (vs 21–40)	0.92 0.67 to 1.28	0.83 0.60 to 1.15	1.15 0.83 to 1.60	0.8 0.58 to 1.11
61+ years (vs 21–40)	1.07 0.77 to 1.50	1.09 0.77 to 1.54	1.16 0.83 to 1.63	0.84 0.60 to 1.18
Clinical academic credentials (vs academic only)	0.92 0.67 to 1.26	0.75 0.54 to 1.04	0.92 0.66 to 1.28	0.79 0.57 to 1.08
US born (vs not)	0.91 0.67 to 1.23	0.8 0.59 to 1.09	0.84 0.61 to 1.14	1.01 0.74 to 1.38
Reads research daily (vs <daily)	0.95 0.75 to 1.21	0.97 0.76 to 1.23	0.93 0.74 to 1.19	1.1 0.86 to 1.39
CEPH programme (vs school)	1.26 0.98 to 1.62	1.12 0.87 to 1.43	1.11 0.86 to 1.43	1.67*** 1.30 to 2.15
Ivy League institution (vs others)	0.8 0.52 to 1.24	0.71 0.46 to 1.11	0.92 0.59 to 1.43	0.96 0.62 to 1.49
South region (vs Northeast)	0.91 0.65 to 1.29	1.14 0.80 to 1.61	0.93 0.66 to 1.30	1.01 0.72 to 1.43
Midwest region (vs Northeast)	1.09 0.73 to 1.63	1.39 0.93 to 2.07	1.04 0.70 to 1.55	1.23 0.83 to 1.84
West region (vs Northeast)	1.16 0.74 to 1.82	1.2 0.77 to 1.89	0.88 0.56 to 1.39	0.97 0.62 to 1.52
60–<120 s spent reading (vs <60 s)	0.65** 0.50 to 0.84	0.73* 0.55 to 0.96	0.97 0.74 to 1.28	1.31* 1.02 to 1.67
120+ s spent reading (vs <60 s)	0.5*** 0.35 to 0.72	0.61** 0.44 to 0.84	0.66* 0.47 to 0.93	1.55* 1.01 to 2.38
N‡	895	895	895	895

*p<0.05; **p<0.01; ***p<0.001.

†Generalised ordered logit models controlling for all variables in each column.

‡Only survey responses with no missing data included in the multivariate analysis.

Abstract 1, RCT/Service; Abstract 2, Cross-sectional/Service; Abstract 3, RCT/Pharmaceutical; Abstract 4, Cross-sectional/Pharmaceutical.

We were motivated to conduct this study due to a strong expectation that there would be a bias against LIC abstracts, or at least that source would make a difference to how the respondents viewed the strength of evidence in the abstract and whether they would choose to refer the abstract to a peer. Although we found no difference in three of the four abstracts, a high-income source did make a difference to participants' view of the relevance of one of the abstracts. All things being equal, our sample population considered the RCT of the pharmaceutical intervention to be significantly more relevant to their peer group if its source was from the UK rather than from Malawi.

We did take several steps to ensure that if explicit biases existed then we would capture them. We randomised the survey abstracts to control for known and unknown confounders, and this was performed well, as evidenced by the balanced characteristics of the two survey groups. We framed the research as a Speed

Reading survey to encourage respondents to spend the minimum time assessing the abstract and to allow anchoring to specific pieces of information in the abstract to occur, and we made no reference to the hypothesis that we were testing to not influence the responses. We achieved a large sample size to be able to detect small, but meaningful, differences in the distribution of the responses—the completed-survey response rate of nearly 10% is within the range expected for a time-consuming, internet-based survey with no preinvitation recruitment.⁴⁵ The fact that the survey was presented as a Speed Reading test may also have reduced selection bias, in that its stated purpose would not necessarily appeal to one type of researcher, such as those with more global health experience.

However, the result was less dramatic than we expected, occurring in only one of the four abstracts, and it suggests that explicit biases are small and difficult to detect across

a relatively small group of abstracts. The study provides an empirical baseline against which to compare future research into the effect of source on abstract evaluation. Indeed, it could be argued that the implications of this study are encouraging for the population that participated because the two groups of survey respondents treated three of the four abstracts almost identically, irrespective of the source. Public health faculty in the USA seem to be doing what is expected of them. Research is being assessed, by and large, according to its content rather than its origin. For those interested in exploring the barriers to Reverse Innovation, or types of publication bias, this finding may be encouraging.

In our study, we also found that respondents spent on average between 70 and 100 s per abstract. Rapid responders tended to rate abstracts higher, so it is possible that if less time is spent on the abstracts then anchoring to particular triggers might be having a greater effect. We did find that, in Abstract 4, if more time is taken to respond to the abstract then opinion of it improves (for both strength of evidence and referral), however, this is equal between high as well as low-income sources. We also found, as would be expected, that respondents tended to rate the randomised controlled trial abstracts higher for strength of evidence compared to the abstracts that were of a cross-sectional design. As the study was framed as a Speed Reading assessment, survey participants might have felt the need to speed-read the abstracts, which may not mirror normal practice.

We also note that the wide SDs in the outcomes indicate that, despite the large sample size, there is considerable variation in how readers view and consume research. The wide SDs might have reduced our ability to detect differences and further work should be conducted to validate measurement constructs in this context. GRADE⁴⁶ and Jadad⁴⁷ scores are widely used, but usually to assess entire research articles against judgement of research quality, risk of bias, inconsistency, indirectness, imprecision and publication bias.^{48–54} Our study, designed purposefully to be a rapid appraisal only of the research abstract, demonstrated extremely wide variation in the assessment of the limited information provided in the abstracts. This finding may have implications for systematic reviews, meta-analyses and for reviewers of abstracts submitted for conferences.

Considering the volume of abstracts read and consumed on a daily basis from all parts of the globe, if source impacts on one's perception, even by a tiny margin, this might, at scale, be an observable phenomenon. We cannot speculate as to the triggers individuals identify with when reading each individual abstract under relatively rapid, timed conditions, but it is encouraging that overall there were few differences between the two survey groups. As highly trained researchers in public health, we could expect an explicit bias to be extremely small, if present at all. It is possible that in other population groups this survey would present different findings. Policymakers, clinicians, journalists and health service

managers are all important actors in innovation diffusion processes, and may also be involved in peer-review processes for academic publication. Our strategy to include academic public health professionals in this survey is based on a best-case assessment of likely bias. Future research ought to modify the approach we have chosen in accordance with the target population, using other abstracts or developing a research design that allows respondents to serve as their own controls. Although we found only one of the four abstracts eliciting a small (yet statistically significant) difference in rating, it is unclear whether this proportion would hold across the population level in practice. It certainly raises the question of whether abstracts and articles submitted for peer review should be masked to country-of-origin.⁵⁵ The eighth International conference on peer review in biomedical research sets the stage for a more detailed examination of cognitive biases in healthcare evidence interpretation.⁵⁶

Author affiliations

¹Division of Surgery, Department of Surgery and Cancer, Institute of Global Health Innovation, Imperial College London, St Mary's Hospital, London, UK

²Departments of Health Policy and Management and Community Health Sciences, UCLA Fielding School of Public Health, Center for Health Sciences, Los Angeles, California, USA

³Department of Health Services and Outcomes Research, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

⁴Department of Primary Care and Public Health, Imperial College London, London, UK

⁵Manpower Demonstration Research Corporation (MDRC), New York, New York, USA

Acknowledgements The authors gratefully acknowledge the support of New York University during this research. The authors would like to thank Professor Jane Noyes and Dr Martin Emmert, as well as the four external peer reviewers, for helpful comments on earlier drafts of the article.

Contributors MH conceived and designed the research, collected and cleaned the data, helped to analyse the data, wrote the first draft and revised subsequent drafts for important intellectual content. JM analysed the data and helped to design the research, and revised the drafts for important intellectual content. MM conducted an initial pilot of the survey, helped to collect data, contributed to the first draft and revised subsequent drafts for important intellectual content. GJ helped to collect data, design the research and revised subsequent drafts for important intellectual content. CA helped to clean the data and to analyse it, and revised subsequent drafts for important intellectual content.

Funding This study was conducted as part of a Harkness Fellowship awarded to MH from the US Commonwealth Fund (2014–2015). The authors are grateful for the support of the Commonwealth Fund. The article does not represent the views of the Commonwealth Fund.

Competing interests None declared.

Ethics approval New York University—University Committee on Activities Involving Human Subjects.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Kaptchuk T. Effect of interpretive bias on research evidence. *BMJ* 2003;326:1453–5.
2. Ferlie E, Fitzgerald L, Wood M, *et al.* The nonspread of innovations: the mediating role of professionals. *Acad Manage J* 2005;48:117–34.
3. Dearing J, Meyer G, Kazmierczak J. Portraying the new: communication between university innovators and potential users. *Sci Commun* 1994;16:11–42.
4. Fitzgerald L, Ferlie E, Wood M, *et al.* Interlocking interactions, the diffusion of innovations in health care. *Hum Relations* 2002;55:1429–49.
5. Bilkey W, Nes E. Country-of-origin effects on product evaluations. *J Int Bus Stud* 1982;13:89–100.
6. Peterson RA, Jansson S. A meta-analysis of country-of-origin effects. *J Int Bus Stud* 1995;26:883–900.
7. Greenhalgh T, Robert G, Macfarlane F, *et al.* Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q* 2004;82:581–629.
8. Damschroder L, Aron D, Keith R, *et al.* Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci* 2009;4:50.
9. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science* 1981;211:453–8.
10. Bargh J, Chen M, Burrows L. Automaticity of social behaviour: direct effects of trait construct and stereotype activation on action. *J Pers Soc Psychol* 1996;71:230–44.
11. Lord CG, Ross L, Lepper M. Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *J Pers Soc Psychol* 1979;37:2098–109.
12. MacCoun RJ. Biases in the interpretation and use of research results. *Annu Rev Psychol* 1998;49:259–87.
13. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science* 1974;185:1124–31.
14. Harris M. Research papers should omit their authors' affiliations. *BMJ* 2014;349:g6439.
15. Yousefi-Nooraie R, Shakiba B, Mortaz-Hejri S. Country development and manuscript selection bias: a review of published studies. *BMC Med Res Methodol* 2006;6:37.
16. Horton R. North and South: bridging the information gap. *Lancet* 2000;355:2231–6.
17. Patel V, Sumathipala A. International representation in psychiatric literature: survey of six leading journals. *Br J Psychiatry* 2001;178:406–9.
18. Mendis S, Yach D, Bengoa R, *et al.* Research gap in cardiovascular disease in developing countries. *Lancet* 2003;361:2246–7.
19. Yach D, Kenya P. Assessment of epidemiological and HIV/AIDS publications in Africa. *Int J Epidemiol* 1992;21:557–60.
20. Singh D. Publication bias- a reason for the decreased research output in developing countries. *S Afr Psychiatry Rev* 2006;9:153–5.
21. Keiser J, Utzinger J, Tanner M, *et al.* Representation of authors and editors from countries with different human development indexes in the leading literature on tropical medicine: survey of current evidence. *BMJ* 2004;328:1229–32.
22. Ophthofa T, Coronel R, Jansed M. The significance of the peer review process against the background of bias: priority ratings of reviewers and editors and the prediction of citation, the role of geographical bias. *Cardiovasc Res* 2002;56:339–46.
23. Link AM. US and non-US submissions. An analysis of reviewer bias. *JAMA* 1998;280:246–7.
24. Tutarel O. Composition of the editorial boards of leading medical education journals. *BMC Med Res Methodol* 2004;4:3.
25. Winnik S, Speer T, Raptis D, *et al.* The wealth of nations and the dissemination of cardiovascular research. *Int J Cardiol* 2013;169:190–5.
26. Peters DP, Ceci SJ. Peer-review practices of psychological journals: the fate of published articles, submitted again. *Behav Brain Sci* 1982;5:187–255.
27. Kliever M, DeLong D, Freed K, *et al.* Peer review at the American Journal of Roentgenology: how reviewer and manuscript characteristics affected editorial decisions on 196 major papers. *AJR Am J Roentgenol* 2004;183:1545–50.
28. Kirkwood B, Sterne J. *Essentials of medical statistics*. Oxford: Blackwell Science, 2003.
29. Wakefield AJ, Murch SH, Anthony A, *et al.* RETRACTED: ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* 1998;351:637–41.
30. Abeygunasekera A. Learning from low income countries: what are the lessons? Effective surgery can be cheap and innovative. *BMJ* 2004;329:1185.
31. Dogra N, Omigbodun O. Learning from low income countries: what are the lessons? Partnerships in mental health are possible without multidisciplinary teams. *BMJ* 2004;329:1184.
32. McKenzie K, Patel V, Araya R. Learning from low income countries: mental health. *BMJ* 2004;329:1138.
33. Swartz HA, Rollman BL. Managing the global burden of depression: lessons from the developing world. *World Psychiatry* 2003;2:3.
34. Susser E, Collins P, Schanzer B, *et al.* Topics for our times: can we learn from the care of persons with mental illness in developing countries? *Am J Public Health* 1996;86:7.
35. Rosen A. Destigmatizing day-to-day practices: what developed countries can learn from developing countries. *World Psychiatry* 2006;5:1.
36. Haines A, Sanders D, Lehmann U, *et al.* Achieving child survival goals: potential contribution of community health workers. *Lancet* 2007;369:2121–31.
37. Haider R, Ashworth A, Kabir I, *et al.* Effect of community-based peer counsellors on exclusive breastfeeding practices in Dhaka, Bangladesh: a randomised controlled trial. *Lancet* 2000;356:1643–7.
38. Kumar V, Mohanty S, Kumar A, *et al.* Effect of community-based behaviour change management on neonatal mortality in Shivgarh, Uttar Pradesh, India: a cluster-randomised controlled trial. *Lancet* 2008;372:1151–62.
39. Syed S, Dadwal V, Martin G. Reverse innovation in global health systems: towards global innovation flow. *Global Health* 2013;9:36 (30 August 2013).
40. Wright J, Walley J, Philip A, *et al.* Research into practice: 10 years of international public health partnership between the UK and Swaziland. *J Public Health* 2010;32:277–82.
41. Baguley D, Killeen T, Wright J. International health links: an evaluation of partnerships between health-care organizations in the UK and developing countries. *Trop Doct* 2006;36:149.
42. Lam CLK. Knowledge can flow from developing to developed countries. *BMJ* 2000;321:830.
43. Berwick D. Lessons from developing nations on improving health care. *BMJ* 2004;328:1124–9.
44. Crisp N. *Turning the world upside down*. RSM books, 2010.
45. Cook C, Heath F, Thompson R. A meta-analysis of response rates in web- or internet-based surveys. *Educ Psychol Meas* 2000;60:821–36.
46. Malmivaara A. Methodological considerations of the GRADE method. *Ann Med* 2015;47:1–5.
47. Jadad AR, Moore RA, Carroll D, *et al.* Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1–12.
48. Balshem H, Helfand M, Schünemann HJ, *et al.* GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401–6.
49. Guyatt GH, Oxman AD, Vist G, *et al.* GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol* 2011;64:407–15.
50. Guyatt GH, Oxman AD, Kunz R, *et al.* GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol* 2011;64:1294–302.
51. Guyatt GH, Oxman AD, Kunz R, *et al.* GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol* 2011;64:1303–10.
52. Guyatt GH, Oxman AD, Kunz R, *et al.* GRADE guidelines 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol* 2011;64:1283–93.
53. Guyatt GH, Oxman AD, Montori V, *et al.* GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64:1277–82.
54. Guyatt GH, Oxman AD, Sultan S, *et al.* GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 2011;64:1311–16.
55. Harris M, Weisberger E, Silver D, *et al.* 'They hear "Africa" and they think that there can't be any good services'—perceived context in cross-national learning: a qualitative study of the barriers to Reverse Innovation. *Global Health* 2015;11:45.
56. Rennie D, Flanagan A, Godlee F, *et al.* Eighth international congress on peer review in biomedical publication. *BMJ* 2015;350:h2411.

Correction

Harris M, Macinko J, Jimenez G, *et al.* Does a research article's country of origin affect perception of its quality and relevance? A national trial of US public health researchers. *BMJ Open* 2015;5:e008993. The institutional affiliation of the last author of this paper is incorrect. Chloe Anderson's correct affiliation is: MDRC, New York, NY, USA; work supported and completed while at The Commonwealth Fund, New York, NY, USA.

BMJ Open 2016;6:e008993corr1. doi:10.1136/bmjopen-2015-008993corr1



CrossMark