How global is global health research? A large-scale analysis of trends in authorship

Michelle C Dimitris, Matthew Gittings, Nicholas B King

ABSTRACT
Many have called for greater inclusion of researchers from low- and middle-income countries (LMICs) in the conduct of global health research, yet the extent to which this occurs is unclear. Prior studies are journal-, subject-, or region-specific, largely rely on manual review, and yield varying estimates not amenable to broad evaluation of the literature. We conducted a large-scale investigation of the contribution of LMIC-affiliated researchers to published global health research and examined whether this contribution differed over time. We searched titles, abstracts, and keywords for the names of countries ever classified as low-, lower middle-, or upper middle-income by the World Bank, and limited our search to items published from 2000 to 2017 in health science-related journals. Publication metadata were obtained from Elsevier/Scopus and analysed in statistical software. We calculated proportions of publications with any, first, and last authors affiliated with any LMIC as well as the same LMIC(s) identified in the title/abstract/keywords, and stratified analyses by year, country, and countries’ most common income status. We analysed 786 779 publications and found that 86.0% included at least one LMIC-affiliated author, while 77.2% and 71.2% had an LMIC-affiliated first or last author, respectively; however, analogous proportions were only 58.7%, 36.8%, and 29.1% among 100 687 publications about low-income countries. Proportions of publications with LMIC-affiliated authors increased over time, yet this observation was driven by high research activity and representation among upper middle-income countries. Between-country variation in representation was observed, even within income status categories. We invite comment regarding these findings, particularly from voices underrepresented in this field.

INTRODUCTION
Diversity of scientists publishing academic research is increasingly scrutinised, particularly among studies in or about low- and middle-income countries (LMICs). Editors published in BMJ Global Health and the Lancet Global Health have candidly expressed concern about the extent to which global health research is conducted by local experts. Several academic journals promote diversity among contributors, specifically encouraging authors from within region(s) of study, in recent authorship guidelines.

Authorship is a measurable yet imperfect metric of inclusion in research and has been investigated in several small-scale initiatives. Reviews that examined representation of LMIC-affiliated authors of similarly associated research estimated that 21% of orthopaedic, 53% of community health worker, 57% of maternal health, 69% of paediatric health, 73% of surgery, and 75% of palliative care studies, as well as 50% of clinical trials pertaining to HIV, malaria, and tuberculosis, were led by LMIC-affiliated first authors.

What is already known?
► Previous studies of authorship of global health research are small in scale, often rely on manual review, and report widely varying estimates.
► We analysed metadata for over 700 000 publications to examine the extent to which authors affiliated with low- and middle-income countries are represented in global health research, and how this has changed over time.

What are the new findings?
► Proportions of studies with any, first, and last author(s) affiliated with a low- or middle-income country increased over time; however, magnitudes of increases were generally highest for any author and comparatively low for proportions of first and last authors among low-income countries.
► We observed substantial variation in representation by country, even within upper middle-, lower middle-, and low-income categories.

What do the new findings imply?
► We find evidence of uneven progress towards greater representation of low- and middle-income-affiliated authors in health science-related studies about low- and middle-income countries.
► We welcome thorough discussion about the implications of our findings and the barriers and facilitators to diversity in authorship involving several perspectives from researchers worldwide.
Investigations of publications in *The Lancet Global Health* from 2013 to 2017 and *The International Journal of Tuberculosis and Lung Disease* from 2006 to 2008 estimated that between 35% of all and 88% of first authors were affiliated with LMICs, respectively. Reviews of genomic epidemiology and infectious diseases research in Africa/sub-Saharan Africa reported 48%–50% of studies were led by similarly affiliated authors. Recently, a comprehensive investigation of research in sub-Saharan Africa found that 53% of publications were led by an author affiliated with the country of focus; this study examined 7100 publications, and to our knowledge is the largest study of authorship affiliations specifically pertaining to research on/about LMICs.

Unfortunately, all prior research is journal-, subject-, or region-specific and often relies on manual review of authorship affiliations, which limits the scope of research that can be included. Estimates produced by these studies vary widely and thus do not broadly inform representation in the global health research landscape, nor are these methods conducive to monitoring diversity in authorship over time. We are unaware of any large-scale attempts to examine trends over time in authors’ countries of affiliation within global health research. We investigated the extent to which published global health research is authored by researchers affiliated with LMICs, and whether proportions changed over time.

**METHODS**

We searched Scopus, which is the largest database of abstracts/citations of publications, for studies with the name of an LMIC in the title, abstract, or keywords. Scopus includes publications written in many languages, though the title, abstract and keywords must be indexed in English. Importantly, Elsevier maintains structured metadata for all publications indexed in Scopus, including fields for the electronic identifier, title, abstract, and authors’ countries of affiliation.

We identified all countries that were ever classified as upper middle-, lower middle-, or low-income from 1991 to 2017 inclusive, as well as yearly classifications for each country, from the World Bank. We compared this list to countries classified as low or medium on the United Nations Development Programme Human Development Index, and additionally included Kuwait. We identified alternative country names using a search filter proposed by the Cochrane Collaboration, in draft format at the time of data collection for the current study, and the List of Alternate Country Names from Wikipedia. We restricted results to items published in years 2000–2017 inclusive within journals indexed in at least one health science-related field (i.e. dentistry, health professions, medicine, and/or nursing). Our search was conducted on 26 November 2018 and is available in online supplemental material 1.

We extracted structured metadata for all items of all publication types identified by our search from Elsevier/Scopus. Data were downloaded from the Scopus Application Programming Interface between 26 November 2018 to 10 February 2019 via http://api.elsevier.com and http://www.scopus.com. We then excluded items for which data extraction was incorrect/incomplete, items with no author and/or no information on any authors’ countries of affiliation, as well as items captured only due to author names or terms that resembled but were unrelated to LMICs. We identified these terms based on a face validity of at least ten titles/abstracts per country as well as from https://www.macmillandictionary.com (online supplemental material 1).

We used statistical software to parse the author affiliation fields for the names of any/all LMICs included in our search, and to identify instances in which author affiliation matched an LMIC(s) identified in the abstract, title, and/or keywords. We calculated proportions of publications with the name of any LMIC among authors’ countries of affiliation, stratified by year, author order, and countries’ most common income status from 2000 to 2017 (i.e., low/lower middle/upper middle; countries ever classified as an LMIC but most commonly classified as high-income were included in the upper middle-income category). Kuwait, which was identified as low/medium on the Human Development Index, but high-income throughout the study period, was grouped in the upper middle-income category for the purpose of these analyses. We calculated proportions of publications authored by researchers affiliated with the same LMIC(s) identified in the abstract, title, and/or keywords, stratified by country, countries’ most common income status from 2000 to 2017, and author order. Publications that pertained to more than one country were included in all relevant strata; for example, publications that included the names of both lower middle- and low-income countries were included in overall analyses (where we identified whether authors were affiliated with an LMIC), stratified analyses of lower middle-income countries (where we identified whether authors were affiliated with lower middle-income countries), stratified analyses of low-income countries (where we identified whether authors were affiliated with low-income countries), and country-specific analyses (where we identified whether authors were affiliated with each country identified in the title, abstract, and/or keywords).

We used linear regression to estimate average change in proportion of any/first/last author by year, both overall as well as stratified by countries’ most common income status, and $\chi^2$ tests to examine differences in proportions of any/first/last author affiliated with countries’ most common income status by income status categories. Analyses were conducted in Stata V.14.2.

**RESULTS**

**Health-related publications about LMICs increased over time**

Our search yielded 876 084 publications with the name of an LMIC in the title, abstract, and/or keywords. We
excluded 961 items with incorrect/incomplete data abstraction, 257 items published outside of our date range, 9107 with no authors, 63 769 with no information on any authors’ countries of affiliation, and 15 931 items containing only terms that resemble but do not pertain to LMICs. We retained 786 779 publications in total (online supplemental figure S1).

Publications with LMICs in the title, abstract, or keywords increased over time (figure 1). Approximately 86.2% were published in English, with the remaining 13.8% published in 40 different languages, including Afrikaans, Arabic, Azerbaijani, Belarusian, Bosnian, Bulgarian, Catalan, Chinese, Croatian, Czech, Danish, Dutch, Estonian, Finnish, French, German, Greek, Hebrew, Hindi, Hungarian, Icelandic, Italian, Japanese, Korean, Lithuanian, Malay, Norwegian, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovenian, Spanish, Swedish, Thai, Turkish and Ukrainian.

Publications had a median (25th; 75th percentile) of 5 (3; 7) authors, with 8.6% having one and at least 95% having 12 or fewer authors. Publications pertained to a median (25th; 75th percentile) of 1 (1; 1) LMICs, with 88.1% pertaining to one and at least 95% pertaining to two or fewer LMICs.

Publications were most often about upper middle- and lower middle-income countries, while low-income countries were less represented

We identified 45, 53, and 51 countries most commonly classified as low-, lower middle- and upper middle-income, respectively, from 2000 to 2017; we included 25 countries ever classified as upper middle-income, but most commonly classified as high-income from 2000 to 2017, in the upper middle-income category. The largest proportion of publications contained upper middle-income and lower middle-income countries in the title/abstract/authors, while low-income countries were proportionally less represented (figures 2–4). Specifically, 47.3% and 45.1% of publications included upper middle-income and lower middle-income countries, respectively, while only 12.8% of publications included low-income countries.

Proportions of publications with LMIC-affiliated authors increased over time

Overall, 86.0% of publications had at least one LMIC-affiliated author, while 77.2% and 71.2% had an LMIC-affiliated first or last author, respectively. When stratified by countries’ most common income status, 58.7%,
36.8%, and 29.1% of publications with low-income countries in the title/abstract/keywords had any, first, or last author affiliated with a low-income country, respectively; corresponding proportions were 81.5%, 74.1%, and 67.7% when restricted to lower middle-income countries, and 85.7%, 78.2%, and 73.6% when restricted to upper middle-income countries (online supplemental table S1).

**Magnitudes of increases differed by authorship order and countries’ most common income status**

We observed modest increases over time in the proportion of publications authored by researchers affiliated with LMICs (figure 1). Proportions of publications with any, first, or last author affiliated with an LMIC ranged from 74.8% to 89.8%, 66.0% to 80.9%, and 61.2% to 76.6%, respectively, from 2000 to 2017; however, proportions were highest among upper middle-income countries and lowest among low-income countries (figures 2–4). We estimated average yearly increases of 0.72, 0.70, and 0.74 percentage points in any, first, and last authors affiliated with LMICs; analogous increases were 0.88, 0.42, and 0.34 percentage points for low-income, 0.85, 0.90, and 0.91 percentage points for lower middle-income, and 0.32, 0.28, and 0.35 for upper-middle-income countries (online supplemental table S2).

**Considerable between-country variation was observed**

We found variation in both research output and authorship by country when examining publications authored by researchers affiliated with the same LMIC(s) identified in the title, abstract, and/or keywords. Number of publications ranged from 31, 47, and 48 about Tuvalu, Nauru, and Aruba, to 65, 218, 80, 211, and 115, 161 about Brazil, India, and China, respectively. Proportions of any, first, and last author affiliated with the same LMIC ranged from 0 to greater than 90% (figure 5).

By definition, the proportion of publications with first and last authors affiliated with an LMIC of interest increased as the proportion of publications with any author affiliated with the LMIC of interest increased for each country; however, the magnitude of these increases appeared largest for upper middle- and lower middle-income countries, and smallest for low-income countries. For example, lower middle-income countries such as Iran, Tunisia, and China, as well as upper middle-income countries such as Poland, Republic of Korea, and Brazil had high proportions of any, first, and last authors.
affiliated with the same country. Conversely, low-income countries including Kenya, Tanzania, and Uganda exhibited relatively high numbers of publications (9082, 7154, 8216, respectively), and proportions of any author affiliated with the country of interest (68.8%, 68.9%, 67.7%, respectively), yet had relatively low proportions of both first (38.4%, 39.0%, 36.9%, respectively) and last (32.2%, 26.2%, 28.2%, respectively) authors affiliated with the same country. Ethiopia had the highest proportion of any, first, and last authors affiliated with the same country among low-income countries (57.7%, 46.2%, 72.2%, respectively), yet less than half of publications pertaining to Ethiopia had a similarly affiliated last author. Proportions of first and last authors were consistently below 50% among remaining low-income countries.

DISCUSSION

We find suggestive evidence of uneven progress towards greater representation of authors from LMICs in research on LMICs. Although publications on LMICs increased five-fold between 2000 and 2017, most research concerned upper middle-income and lower middle-income countries, while low-income countries were proportionally underrepresented. We find substantial gradients in overall proportions and trends of any, first, and last authors affiliated with LMICs by countries’ most common income status, as well as considerable between-country variation, even within upper middle-, lower middle-, and low-income status categories. While average change in proportion of any authors affiliated with LMICs is high among both low-income and lower middle-income countries, analogous changes in proportions of first and last authors among low-income countries are less than half that observed for lower middle-income countries. Additionally, proportions of any author, first author, and last author are substantially lower for publications on low-income countries.

Strengths and limitations of our approach

We provide a more complete answer to an existing question about the extent to which the global health literature is authored by LMIC-affiliated researchers. Strengths of our approach largely pertain to its scale and use of structured metadata and statistical software rather than manual review. We leveraged detailed metadata maintained by Elsevier/Scopus and parsed structured data fields for the names of LMICs to generate authorship metrics for over 700,000 publications. Our approach is readily scalable for several planned analyses, including by journal/impact factor and topic, and presents potential for efficient monitoring of authorship metrics.

Our approach has many limitations. We did not examine unpublished or unindexed research; while Scopus is the largest repository of abstracts/citations from journals across the world,17 smaller or regional journals are likely underrepresented in this database. Additionally, it is possible that data completeness and quality varied by journal and over time. Although Scopus includes items published in several languages, titles and abstracts must be translated to English for indexing purposes; thus, titles and abstracts that were not translated to English may not be available in this database.

We aimed to improve the sensitivity of our search by including several alternative country names in the English language within our search, but may have missed non-English country names, or titles/abstracts/keywords that include names of only continents, cities, or other regional terms. We did not include terms such as ‘global health’ and ‘LMICs’ within our search, since these are unstandardised and likely to capture reviews/commentaries that were outside the scope of our study. It is possible that authors affiliated with LMICs are underrepresented or overrepresented in this literature in comparison to studies that include the name of an LMIC in the title, abstract, or keywords.

We acknowledge that authorship/order and countries of affiliation are imperfect measures of research activity and regional knowledge, respectively. For example, temporary institutional affiliations assumed during graduate studies may misrepresent students’ regional knowledge. Researchers may be affiliated with several institutions, and only include one or a subset in any given
publication. Scopus metadata captures multiple affiliations, but we cannot rule out affiliations that are either not supplied by authors or unrecognised by journals. Importantly, our approach was largely analogous to that of smaller-scale studies while vastly increasing the scope.

While small-scale studies largely evaluated publications for inclusion using manual review, we identified health science-related publications pertaining to LMICs by searching the Scopus database. Thus, our approach involves a necessary exchange of potentially decreased specificity for greatly increased sensitivity in identifying relevant articles. We were unable to manually review all publications included in our analysis due to its scope; however, we have made our search criteria available, which enhances the transparency and replicability of our analysis (online supplemental material 1). We additionally aimed to improve the specificity of our search by excluding terms similar to LMICs identified by both dictionary sources and face validity of at least 10 titles/abstracts per country (online supplemental material 1); however, identification of such terms may not be comprehensive for countries resembling common names or other geographical locations (for example Chad, Georgia, Jordan). Notably, our findings are not driven by one specific country, as demonstrated by country-specific analyses (figure 5).

Comparing our findings to smaller-scale studies

We estimated that 77.2% of publications about LMICs were led by authors affiliated with LMICs, which is greater than estimated by most previous studies, yet aligned with investigations of global surgery and palliative care research.9 10 Inclusion of more recent publications (i.e. up to/including 2017) may contribute to this discrepancy, since we observed that both number of publications and percent of authors affiliated with LMICs increased over time. Some studies have observed lower LMIC representation among journals with higher impact factors.7 26 Scopus is a large database that includes many journals with a wide range of impact factors; thus, our approach was largely analogous to that of small-scale studies while vastly increasing the scope. Importantly, we find substantial variation in both number of publications and authorship representation by country within all income groups.

Although our hypotheses have been anecdotally supported by researchers across venues at which we have presented preliminary findings, we are mindful of our own limitations in inferring causality from these data. We are three researchers affiliated with a single university in a high-income country, with limited expertise in the contextual factors that shaped our findings. We offer our descriptive findings as evidence for a larger discussion on authorship that includes multiple perspectives. Given the substantial between-country variation observed, we welcome comment and open discussion of our findings from the global health research community, particularly among voices underrepresented in this literature.

CONCLUSION

We find limited evidence for variation in contribution of authors affiliated with LMICs to published global health research over time and substantial variation by country. We recommend future analyses focus on low-income countries, as these demonstrate lowest levels of representation, particularly among first and last authors. We additionally suggest further exploration of between-country differences, even within the same income status, to highlight barriers and facilitators to representation.

Acknowledgements We thank Genevieve Gore, Dr Brian Robinson, and staff at Elsevier/Scopus for their technical contributions. Our project was supported by the Centre for Social and Cultural Data Sciences Seed Grant Programme at McGill University.

Contributors MCD conceived of and planned study, conducted analyses, wrote manuscript and is responsible for content as the guarantor. MG coordinated and performed extraction of all metadata via the Elsevier Application Interface. NBK guided study design, analysis and presentation. All authors read, revised and approved the final manuscript.

Funding Our project was funded by the Centre for Social and Cultural Data Science (CSCDS) Interdisciplinary Seed Grant Programme at McGill University. All authors had full access to data and take responsibility for the integrity of data and accuracy of data analysis.

Disclaimer CSCDS had no role in the study design; in the collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

Competing interests All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Patient and public involvement statement Patients and the public were not involved in the design, conduct, or dissemination of this study. Dissemination to patients/participants is not applicable.

Patient consent for publication Not required.
REFERENCES


