**Methods for analyzing the authorship patterns of EDCTP1-funded publications**

Authors: Julie Chaccour, Joe Brew

Data source: EDCTP publication database, preferably in .csv, .sql or .txt format, otherwise as Excel spreadsheet, extracted from Thomson Reuters Web of Science (exact variables available TBD), including original research articles on epidemiology or clinical trials, total ~700 papers

Groupings:

Each study will be assigned at least one primary country of study (where the study took place), an area of research (HIV, malaria, TB, NTDs) as well as first (primary) and last (senior) authors. First and last authors will be assigned at least one country affiliation each and a gender (see “Gender” below). Authors and study countries can be grouped in the following groups:

* European EDCTP member countries
* West Africa
* East Africa
* Central Africa
* Southern Africa

To the extent that studies can have multiple countries, multiple areas of research, and potentially multiple regions. Analysis will be two-dimensional, analyzing both factors associated with single categorical belonging, as well as factors associated with certain overlaps (see Treatment of Double Affiliations below). Additionally, trends in authorship patterns will be analyzed over time.

Gender:

Gender will be assigned according to likelihood of having a certain gender based on first names. Given that publication databases do not systematically collect and store data on author gender, we will instead estimate the likelihood of a participant being of one of two genders (male or female) based on first name. In order to do this, we use data from the North Atlantic Population Project, and U.S. government (L. Mullen, Gender: Predict gender from names using historical data, 2015. (https://github.com/ropensci/gender).

Treatment of double affiliations:

In an analysis of institutional affiliations of researchers in the biosciences in Germany, UK and Japan, roughly 10% of authors used multiple affiliations in the period of 2008-2014. Judging from the profoundly collaborative nature of global health research, this percentage is likely to be greater and cannot be neglected in the analysis.

In the literature, various methods account for multiple authorships: Chersich et al. (2016) consider an author locally affiliated if one of his affiliations is located in the country of study and hence assign a single affiliation. This is troublesome, however, because journals with fee waivers for LMIC countries allow for using a secondary affiliation as the host affiliation, thereby incentivizing authors to declare LMIC affiliation over what might otherwise be a more dominant declaration. Hence authors with a primary affiliation in Europe could be counted as African authors. Assigning an author to the country with the largest *per-capita* gross domestic product as in Krause et al. (2007) would provide a second estimate towards the other end of the scale.

Lubrano et al (2003) and Faria et al. (2000) consider an author A with affiliations X and Y as two authors AX and AY. Ciaian et al accredit the publications equally to all affiliated institutions (e.g. ½ X and ½ Y). We shall carry out analysis using (I) the declared affiliation of the author, (ii) Krause’s greatest GDP method, and (iii) Lubrano’s weighted method, and will present all results as part of a sensitivity analysis.

Citation data: ScimagoJR (http://www.scimagojr.com/) and Scopus CiteScore (<https://journalmetrics.scopus.com/>) databases will be used to extract relevant information pertaining to journal rankings and citations, respectively. Thomson Reuters Journal Citation Report 2011 will be used for extracting impact factors.

Time from study end date to publication: Study end date will be extracted for each study from the EDCTP database. Online publication data will be extracted from the publication database if available. Otherwise, publication date will be taken from the online publication.