Availability of transmission data of the 2014-2016 West Africa Ebola epidemic: Protocol for a systematic search

Neal Smith1, Ariel Lebenbaum1, Manoj Gambhir2, James Trauer1

1School of Public Health and Preventive Medicine, Monash University, Australia

2IBM Australia Research

Corresponding author: Neal Smith – neal.smith@monash.edu

# Abstract

## Background

The 2014-2016 Ebola outbreak predominately impacting West Africa was unprecedented in its location, duration, magnitude and spatial spread. The difficulties in effective and complete collection of transmission data are well documented. There may also exist a discrepancy between the completeness of privately and publicly available datasets. Independent research teams may encounter difficulties in answering research questions when using incomplete data, particularly when simulating epidemiological models. This review aims to systematically locate publicly available transmission data, and provide insights into potential impacts on these data on research outputs.

## Methods/design

A systematic search of publicly available data repositories will be conducted for datasets that provide details of Ebola cases during the 2014-2016 West Africa Ebola outbreak. Data repositories including Open Access Directory, figshare and re3data.org will be searched, as well as search engines Google and Google Scholar, for relevant material. In addition, a structured MEDLINE search will aim to identify journal articles that used relevant datasets. Screening of identified articles and databases will be done individually by two authors, with the third author to resolve any inconsistencies. Data extraction will be performed in a consistent manner, with results tabulated and analysed qualitatively.

## Discussion

A broad overview of the variation in key dataset properties will be provided. Discussion will be centred around the potential ramifications of the qualities and quantities of data obtained on epidemiologic research, with a focus on mathematical models.

#### Systematic review registration

This protocol was uploaded to monash.figshare (DOI: 10.4225/03/5a2601042602f) on Wednesday, 13 December 2017.

Keywords: Ebola, Transmission, Data availability

# Background

The 2014-2016 West Africa Ebola outbreak was unprecedented in its scale, duration, physical location and the efforts required to control the disease [1]. On declaring the epidemic over on June 9, 2016,the World Health Organization had reported 28 646 cases (confirmed, probable and suspected) and 11 323 deaths [2]. Causes of such an expansive outbreak are complex and multifactorial, but key contributing factors include a lack of early disease detection and diagnosis, uncontrolled movement of cases, under-resourced healthcare and research facilities and a disconnect between cultural practices and interventions to prevent transmission [3].

In an outbreak setting, compiling comprehensive case and transmission data becomes increasingly difficult. These challenged can be exacerbated by factors such as high rates of transmission, delays in diagnosis of initial cases, low resource settings, lack of clarity surrounding relevant data, and absence of centralised data repositories and data collection protocol. In the case of the West Africa Ebola outbreak, each of these components existed.

In responding to an outbreak situation, a clear and comprehensive strategy for eliminating transmission is a key component. Within this, a structured approach to data collection and storage allows research institutions to identify trends in disease spread, locate key areas of interest and test potential interventions before use, to maximise their utility. Providing public access to data allows researchers to compare outputs in a reliable way, so robust conclusions can be drawn on the steps to implement within the wider outbreak response.

It is known that the sharing of data was poor in the early stages of the West Africa Ebola epidemic [4], which increased the difficulty of disease containment. Most data used during the outbreak was unpublished, and much remains so, more than a year beyond all nations being declared free of Ebola. There is evidence that access to timely data, particularly genetic sequencing data, was particularly helpful in generating transmission trees later in the outbreak [4]. Further, researchers developing mathematical models of disease transmission will often use data from completed outbreaks to test model assumptions, calibrate key parameters and examine hypothetical intervention regimes for future situations [5]. Thus, access to complete, detailed transmission data can play a significant role in the success or failure of containing infectious diseases.

The aim of this systematic review is to determine the scope of data collected during the 2014-2016 West Africa Ebola outbreak, and to assess the public availability of this data. Potential uses of each type of dataset will be outlined, and the advantages and disadvantages of data properties discussed. The time between data collection and public availability will be considered to investigate potential implications of the time-lag on generating information on and strategies to combat disease transmission.

Specific dataset properties to be considered include:

* Duration of dataset
* Number of cases in dataset
* Individual or communal representation of cases
* Breadth of data collected on individual cases
* Spatial and temporal resolution of data collected, if any
* Linkages of cases to one another in a transmission tree, or otherwise
* Whether data was collected or used for the purposes of a study before being made publicly available.

# Methods/Design

Eligibility Criteria

### Focus of Dataset

Datasets must provide information on individual cases and/or disease transmission throughout the 2014-2016 West African outbreak. Search results may either be datasets presented in isolation, or published works (academic or otherwise) containing a dataset in a figure, table, or within the body of work.

### Primary data-points of interest:

* The dates of patient diagnosis, infection and appearance of first symptoms
* The location of cases
* Patient outcomes (survival, death, burial status, treatment status)

### Secondary data-points:

* Patient Ebola disease status (either confirmed, susceptible or probable)
* Relationships between patients or patterns of disease transmission
* Genomic data
* Patient travel histories
* Age and gender
* Healthcare worker status
* Disease progression and patient condition upon diagnosis
* Period of infectiousness
* Date of death and/or burial

## Information Sources

### Electronic searches

This search will be conducted in five separate phases: scanning lists of data repositories, searching specific data repositories, locating data on key websites of the Centre for Disease Control (CDC, America) and the World Health Organization (WHO), performing more general Google and Google scholar searches, and searching the journal article database MEDLINE.

Each search will be performed using a relevant and applicable combination of keywords, indexed terms and Boolean operators. Searches will be adjusted for each repository as needed.

### Proposed search strategy

#### Searching for data repositories

The lists of data repositories to be scanned are:

* Open Access Directory
* re3data.org
* Monash University Database

These lists will be manually examined for relevant repositories, with appropriate results searched alongside the following pre-identified sources:

* figshare (available at <https://figshare.com/>)
* Humanitarian Data Exchange (HDX, <https://data.humdata.org/>)
* The GitHub page of Caitlin Rivers’ Ebola data (<https://github.com/cmrivers/ebola>)

#### Proposed data repository strategy

Data repositories will be searched with the term “Ebola”. Additional limits will be applied where relevant and possible. These are currently limited to:

* Date of publication : 2014-present
* Type: dataset

#### Proposed CDC/WHO website strategy

The CDC and WHO websites contain sections dedicated to the 2014-2016 West Africa Ebola outbreak. These sections will be searched for any remaining datasets.

#### OVID Medline Strategy

The journal database Ovid MEDLINE will be searched to locate research articles that used datasets that satisfy our inclusion criteria. The search strategy is as follows, first combining terms within each part, and then combining the three parts:

##### Part one – terms related to Ebola

1. exp1 Ebolavirus/ OR
2. exp Hemorrhagic Fever, Ebola/ OR
3. Ebola\*.mp.

AND

##### Part two – terms specific to the 2014-2016 West Africa Ebola outbreak

1. exp Sierra Leone/ OR
2. Sierra Leon\*.mp. OR
3. exp Liberia/ OR
4. Liberia\*.mp. OR
5. exp Guinea/ OR
6. Guinea\*.mp. OR
7. exp Africa, Western/ OR
8. West Africa\*.mp. OR
9. exp Senegal/ OR
10. Senegal\*.mp. OR
11. exp Mali/ OR
12. Mali\*.mp. OR
13. Exp Nigeria/ OR
14. Nigeri\*.mp. OR
15. Exp Americas/ OR
16. Exp United States/ OR
17. USA\*.mp. OR
18. United State\*.mp. OR
19. America\*.mp. OR
20. Exp Italy/ OR
21. Italy.mp. OR
22. Itali\*.mp. OR
23. Exp United Kingdom/ OR
24. UK.mp. OR
25. United Kingdom.mp. OR
26. Great Britain\*.mp. OR
27. Exp Spain/ OR
28. Spain\*.mp. OR
29. Spani\*.mp. OR
30. Espan\*.mp.

AND

##### Part three – terms related to transmission data

1. suspected case\*.mp. OR
2. probable case\*.mp. OR
3. confirmed case\*.mp. OR
4. outbreak data\*.mp. OR
5. transmission data\*.mp.

The final search was limited to results from 2014 onwards to account for the timing of the outbreak.

Proposed Google/Google Scholar Strategy

Due to the expansive nature of Google searches, the following search terms were used individually, to provide a number of relevant searches:

1. ebola
2. ebola outbreak
3. ebola outbreak statistics
4. ebola statistics
5. ebola WHO
6. ebola CDC
7. ebola data
8. ebola outbreak dataset
9. ebola database
10. ebola clinic
11. ebola transmission

As Google and Google Scholar searches are likely to return several thousand results, searches will be limited to the first 14 pages of results.

## Study Records

### Selection Process

The initial search strategy will be performed by one reviewer. Any necessary changes to the review strategy will be agreed upon by two reviewers. All identified studies will be screened by title and abstract or description, depending on source of results (i.e. data repository, search engine or journal article database). One reviewers will initially undertake this process, with any uncertainties on inclusion being referred to a second reviewer, and over-caution being applied when determining if full-text article should be searched, datasets analysed or websites further assessed.

In the second stage of selection, two reviewers will independently assess search results for potential inclusion. If consensus is not reached on any sources, a third independent reviewer will be consulted.

Where similar or identical datasets are located by different searches, their inclusion as a single or multiple dataset will be considered on a case-by-case basis.

For transparency, a flow diagram adapted from the PRISMA statement [6] will be generated to show the number of works present throughout the review process.

### Data Collection Process

For included works, data extraction will be performed independently by two reviewers, with inconsistencies resolved by consensus between researchers. If necessary, a third researcher will be consulted. Relevant data will be extracted from tables, figures and within the main text. Data will be collated in tabular form, with data points listed below.

Comparison of data collection will be piloted on a subset of accepted works to ensure reviewers are being consistent in approach, and to modify data items or extraction processes if necessary.

### Data Items

Due to the potentially varied nature of datasets, it is not known what data will be present for extraction. Subject to change, the information we will collect on each dataset includes:

* Name of dataset
* Source of dataset
* Date of latest issue of dataset
* Date dataset became publicly available
* URL
* Number of data entries
* Number of cases of data
* List of available data

Where appropriate, a brief description of certain data will be written (for example, if data is of individual patients or clusters, or the spatial resolution of case data). Any other important notes about datasets will be documented as required.

### Data Synthesis

Primary data synthesis will be a table summarising each dataset, including the information listed above. Due to the nature of the data, quantitative analysis will not be conducted.

Qualitative analysis will primarily be an overview of the properties of available datasets. Discussions will concern key strengths and weaknesses of available data, clear differences between the quality or quantity of particular data, and any trends in data available from particular resources. Relationships between dataset size and completeness will be explored, in particularly whether datasets that are smaller, or from the early or late stages of the epidemic, is generally more comprehensive than larger datasets. Through collecting this information, we will discuss the implications of data availability on epidemiological research, with a focus on mathematical modelling.

# Discussion

One of the difficulties in the initial stages of the West Africa Ebola outbreak was the lack of thorough collection and open dissemination of transmission and case data. It is well known that the open sharing of data provides increased opportunities for collaboration, comparison of experimental and modelling results, and robustness of conclusions from these studies [7]. This review aims to provide a detailed list of publicly available datasets of the 2014-2016 epidemic, and examine the impacts of the timing of public access to these data on outbreak spread.

Whilst there is a great deal of information on techniques to performing systematic reviews of the medical literature [6], the same does not exist for methods to performing a comprehensive search of public databases. On advisement from research librarians, this four-part review strategy has been devised to encompass a wide range of potential data storage locations. Searching lists of data repositories will add to the well-known repositories already selected for this search. Google and Google Scholar searches, as well as direct searching of key organisation websites (such as the CDC and WHO), will provide the most commonly accessed public datasets. The MEDLINE search will give insights into the data accessed and/or collected by other researchers throughout the outbreak. This covers a wide range of potential data and is expected to return a comprehensive number of relevant datasets.

Competing interests: The authors declare that they have no competing interests.

Authors’ contributions: N.R.S., M.G and J.M.T. conceived the study. A.L. and N.S. wrote the first draft, and N.R.S., M.G. and J.M.T. revised the protocol. All authors read and approved the final manuscript.

Acknowledgements: We would like to acknowledge Associate Professor Jonathan Keith for his assistance in study conception and design.

Funding sources: N.R.S is a recipient of an Australian Government Research Training Program (RTP) Scholarship. J.M.T. is a recipient of a Monash Bridging Fellowship. A.L is a recipient of a Monash University School of Public Health and Preventative Medicine Summer Vacation Scholarship.

Role of funding sources: no funding sources contributed to the protocol or manuscript.

# References

1. Piot P, Coltart CEM, Atkins KE. 2017 Preface: “The 2013-2016 West African Ebola epidemic: data, decision-making and disease control”. *Philos Trans R Soc Lond B Biol Sci* **372**(1721):20170020. (doi:10.1098/rstb.2017.0020)

2. World Health Organization. 2016 [cited 2017 Dec 5] Situation Report: Ebola virus disease 10 June 2016. Available from: http://apps.who.int/iris/bitstream/10665/208883/1/ebolasitrep\_10Jun2016\_eng.pdf?ua=1

3. Coltart CEM, Lindsey B, Ghinai I, Johnson AM, Heymann DL. 2017 The Ebola outbreak, 2013–2016: old lessons for new epidemics. *Philos Trans R Soc B Biol Sci* **372**(1721):20160297. (doi:10.1098/rstb.2016.0297)

4. Whitty CJM. 2017 The contribution of biological, mathematical, clinical, engineering and social sciences to combatting the West African Ebola epidemic. *Philos Trans R Soc B-Biological Sci* **372**(1721):4. (doi:10.1098/rstb.2016.0293)

5. Lekone PE, Finkenstädt BF. 2006 Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics* **62**(4):1170–7. (doi:10.1111/j.1541-0420.2006.00609.x)

6. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. 2009 The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Med* **6**(7). (doi:10.1371/journal.pmed.1000100)

7. Chretien JP, Riley S, George DB. 2015 Mathematical modeling of the West Africa ebola epidemic. *Elife* **4**(DECEMBER2015):1–15. (doi:10.7554/eLife.09186)

# Additional Files

Additional file 1: **PRISMA-P 2015 checklist: recommended items to include in a systematic review protocol.** Completed PRISMA-P 2015 checklist for systematic review protocol: availability of transmission data of the 2014-2016 West Africa Ebola epidemic.

|  |  |  |  |
| --- | --- | --- | --- |
| **Section and topic** | **Item Number** | **Checklist item** | **Page number(s)** |
| **ADMINISTRATIVE INFORMATION** | | | |
| Title: | | | |
| Identification | 1a | Identify the report as a protocol of a systematic review | 1 |
| Update | 1b | If the protocol is for an update of a previous systematic review, identify as such | N/A |
| Registration | 2 | If registered, provide the name of the register (such as PROSPERO) and registration number | 1 |
| Authors: | | | |
| Contact | 3a | Provide name, institutional affiliation, email address of all protocol authors; provide physical mailing address of corresponding author | 1 |
| Contributions | 3b | Describe contributions of protocol authors and identify the guarantor of the review | 7 |
| Amendments | 4 | If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments | N/A |
| Support: | | | |
| Sources | 5a | Indicate sources of financial or other support for the review | 7 |
| Sponsor | 5b | Provide name for the review funder and/or sponsor | N/A |
| Role of sponsor or funder | 5c | Describe role of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol | 7 |
| **INTRODUCTION** | | | |
| Rationale | 6 | Describe the rationale for the review in the context of what is already known | 2 |
| Objectives | 7 | Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO) | 2-3 |
| **METHODS** | | | |
| Eligibility criteria | 8 | Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review | 3 |
| Information sources | 9 | Describe all intended information sources (such as electronic databases, contact with study authors, trials registers or other grey literature sources) with planned dates of coverage | 3 |
| Search strategy | 10 | Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated. | 4-5 |
| Study records: | | | |
| Data management | 11a | Describe the mechanism(s) that will be used to manage records and data throughout the review | 7 |
| Selection process | 11b | State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis | 6 |
| Data collection process | 11c | Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators | 6 |
| Data items | 12 | List and define all variables for which data will be sought (such as PICO items, funding source), any pre-planned data assumptions and simplifications | 6 |
| Outcomes and prioritization | 13 | List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale | N/A |
| Risk of bias in individual studies | 14 | Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level., or both; state how this information will be used in data synthesis | N/A |
| Data synthesis | 15a | Describe criteria under which study data will be quantitatively synthesised | N/A |
|  | 15b | If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I2, Kendall’s τ) | N/A |
|  | 15c | Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression) | N/A |
|  | 15d | If quantitative synthesis is not appropriate, describe the type of summary planned | 7 |
| Meta-bias(es) | 16 | Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies) | N/A |
| Confidence in cumulative evidence | 17 | Describe how the strength of the body of evidence will be assesses (such as GRADE) | N/A |