DATA NOTE

Extracting novel antimicrobial emergence events from scientific literature and medical reports [version 1; peer review: 2 approved with reservations]

Emma Mendelsohn ¹, Noam Ross ¹, Allison M. White^{1,2}, Karissa Whiting^{1,3}, Cale Basaraba^{1,4}, Brooke Watson Madubonwu^{1,5}, Erica Johnson^{1,6}, Mushtaq Dualeh^{1,7}, Zach Matson^{1,7}, Sonia Dattaray^{1,8}, Nchedochukwu Ezeokoli ^{1,9}, Melanie Kirshenbaum Lieberman^{1,10}, Jacob Kotcher^{1,11}, Samantha Maher ¹, Carlos Zambrana-Torrel ¹, Peter Daszak¹

¹EcoHealth Alliance, New York, NY, 10018, USA

²Current Address: Office of U.S. Foreign Disaster Assistance, USAID, District of Columbia, USA

³Current Address: Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

⁴Current Address: Mailman School of Public Health, Columbia University, New York, NY, USA

⁵Current Address: American Civil Liberties Union, New York, NY, USA

⁶Current Address: Biology Department, Graduate Center of the City University of New York, New York, NY, USA

⁷Current Address: Oak Ridge Institute for Science and Education Fellow, Centers for Disease Control and Prevention, Atlanta, GA, USA

⁸Current Address: Health Union, Philadelphia, PA, USA

⁹Current Address: Guidehouse, Chicago, IL, USA

¹⁰Current Address: School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, USA

¹¹Current Address: Department of Geoscience, Hobart and William Smith Colleges, Geneva, NY, USA

V1 First published: 12 Nov 2020, 9:1320
<https://doi.org/10.12688/f1000research.26870.1>
 Latest published: 21 Jun 2021, 9:1320
<https://doi.org/10.12688/f1000research.26870.2>

Abstract

Despite considerable global surveillance of antimicrobial resistance (AMR), data on the global emergence of new resistance genotypes in bacteria has not been systematically compiled. We conducted a study of English-language scientific literature (2006-2017) and disease surveillance reports (1994-2017) to identify global events of novel AMR emergence (first clinical reports of unique drug-bacteria resistance combinations). We screened 24,966 abstracts and reports, ultimately identifying 1,773 novel AMR emergence events from 294 articles. Events were reported in 66 countries, with most events in the United States (152), India (129), and China (128). The most common bacteria demonstrating new resistance were *Klebsiella pneumoniae* (352) and *Escherichia coli* (218). Resistance was most common against antibiotic drugs imipenem (89 events), ciprofloxacin (85) and ceftazidime (82). We provide an open-access database of emergence events with

Open Peer Review

Approval Status

	1	2	3
version 2			
(revision)
21 Jun 2021
version 1			
12 Nov 2020

- Barbara Tornimbene**, World Health Organization, Geneva, Switzerland
- Sergey Eremin** , World Health Organization, Geneva, Switzerland

standardized fields for bacterial species, drugs, location, and date, and we discuss guidelines and caveats for data analysis. This database may be broadly useful for understanding rates and patterns of AMR evolution, identifying global drivers and correlates, and targeting surveillance and interventions.

Keywords

Antimicrobial resistance, global health, open-access data

This article is included in the **Pathogens** gateway.

This article is included in the **Antimicrobial Resistance** collection.

2. **Ellen Stobberingh**, Maastricht University Medical Centre (MUMC), Maastricht, The Netherlands

3. **Jorge Pinto Ferreira** , World Organisation for Animal Health, Paris, France

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding authors: Emma Mendelsohn (mendelsohn@ecohealthalliance.org), Peter Daszak (daszak@ecohealthalliance.org)

Author roles: **Mendelsohn E:** Conceptualization, Data Curation, Methodology, Project Administration, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Ross N:** Conceptualization, Data Curation, Methodology, Project Administration, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **White AM:** Conceptualization, Data Curation, Methodology, Project Administration, Supervision, Validation, Writing – Original Draft Preparation; **Whiting K:** Conceptualization, Data Curation, Methodology, Software, Validation, Writing – Original Draft Preparation; **Basaraba C:** Conceptualization, Data Curation, Methodology, Software, Validation; **Watson Madubunwu B:** Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation; **Johnson E:** Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation; **Dualeh M:** Data Curation; **Matson Z:** Data Curation; **Dattaray S:** Data Curation, Project Administration; **Ezeokoli N:** Data Curation; **Kirshenbaum Lieberman M:** Data Curation; **Kotcher J:** Data Curation; **Maher S:** Data Curation; **Zambrana-Torrel C:** Conceptualization, Methodology; **Daszak P:** Conceptualization, Methodology

Competing interests: No competing interests were disclosed.

Grant information: United States Agency for International Development (USAID) Emerging Pandemic Threats PREDICT (Cooperative Agreement No. AID-OAA-A-14-00102).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Mendelsohn E *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Mendelsohn E, Ross N, White AM *et al.* **Extracting novel antimicrobial emergence events from scientific literature and medical reports [version 1; peer review: 2 approved with reservations]** F1000Research 2020, 9:1320 <https://doi.org/10.12688/f1000research.26870.1>

First published: 12 Nov 2020, 9:1320 <https://doi.org/10.12688/f1000research.26870.1>

Introduction

Antimicrobial resistance (AMR) is a global health crisis that has compromised the effective treatment and prevention of a multitude of infections. The rise in AMR has been associated with increased mortality, longer hospitalizations, complications with medical procedures such as surgery and chemotherapy, and higher healthcare costs^{1–3}. Resistance to antibiotics is a global public health issue and a particular concern in low- and middle-income countries, where many high-burden diseases such as malaria, respiratory infections, and tuberculosis can no longer be treated by common antimicrobial drugs^{1,4}. Combating AMR requires a multidimensional global response to optimize antimicrobial drug use, improve awareness, increase traceability and usage reporting, and promote research^{1,5–7}.

AMR surveillance by researchers, hospital networks, and state governments is key to characterizing and responding to the crisis. Current global-scale datasets primarily focus on the presence or prevalence of known resistance genotypes and phenotypes in bacterial populations. In 2014, the World Health Organization (WHO) published surveillance data obtained from 129 member states on nine bacterial pathogen-antibacterial drug combinations of public health importance, finding high rates of resistance reported across the globe³. The [ResistanceMap database](#), maintained by the Center of Disease Dynamics, Economics & Policy, provides nationally-aggregated data from 46 countries on the prevalence of resistance in 12 bacterial species against 17 classes of antibiotics⁸.

To our knowledge, however, there is no publicly available dataset that specifically identifies the spatial and temporal patterns of the emergence of novel bacterial pathogen resistance to antibiotic drugs in humans. Such data are critical to understanding macroecological patterns and drivers of AMR emergence and identifying geographic and phenotypic targets for surveillance, research, and interventions. Further, this database may support on-going agreements and programs developed by intergovernmental

organizations such as the United Nations, World Health Organization, Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Convention on Biological Diversity, and development agencies such as the World Bank, the Global Environment Facility, and other regional banks.

We conducted a systematic review to identify novel AMR emergence events reported in English-language scientific literature from 2006–2017 and disease surveillance reports from 1994–2017. We focused specifically on human clinical cases and included any reported resistance of a bacterium to an antimicrobial drug (i.e., we did not limit the search to a subset of bacteria or drugs). We screened 24,966 abstracts and reports, identifying 1,791 articles potentially reporting novel emergence events, and ultimately, 294 relevant studies reporting 1,773 total novel events. We present an open-access database of these events with cleaned and standardized fields for location of emergence, bacterial species, antimicrobial drug, emergence date, and data source. We provide usage notes on systematic biases and ambiguities in the data. In addition, we provide data on all screened and processed articles from which data were harvested.

Methods

Development of the AMR emergence database was a multi-step process consisting of a systematic literature search, abstract and report screening, article review and coding, and data cleaning and standardization ([Figure 1](#)).

Literature search

We drew from PubMed and Embase scientific literature databases and ProMED-mail to develop our database of events. PubMed and Embase collectively encompass a large fraction of English-language biomedical scientific literature, while ProMED-mail consists of clinical reports and news alerts that may not ultimately be published in the scientific press and have a faster reporting speed. We selected a recent ten-year period for

Figure 1. Overview of database building process.

scientific literature (2006–2017). ProMED-mail reports were drawn from a longer time span (1994–2017), as we found scientific articles published 2006–2017 frequently described events occurring many years before.

We searched for peer-reviewed manuscripts in PubMed and Embase published using the following terms:

'antibiotic resistance'/exp OR ('antibiotic' AND ('resistant' OR 'resistance')) OR 'antimicrobial resistance'/exp OR ('antimicrobial' AND ('resistant' OR 'resistance')) AND (first OR novel OR new OR emerging OR emergent) AND (case OR patient) AND [humans]/lim AND [2006–2017]/py

This search, completed August 2018, yielded 23,770 results.

Because ProMED-mail searches use partial word stem-based matching, we used only the search terms *'antibiotic resistance'* and *'antimicrobial resistance'*, which yielded 1,196 results. A total of 24,966 articles were compiled for screening.

Abstract and report screening

We used the following inclusion criteria to screen the results of our literature search: the article must have described at least one **clinical case** (a) of infectious disease in a **patient** (b) caused by a **novel** (c) **resistance** (d) to a particular **antimicrobial drug or drug combination** (e) in **bacteria** (f). In this definition:

a) A “clinical case” is an individual who presents with symptoms to a medical professional and is determined by a medical professional to have been infected with the bacterium in question. Antimicrobial resistance in an asymptomatic individual’s commensal bacteria (as determined by a screening study, a “challenge” study, or a laboratory trial) does not meet the requirements of a clinical case.

b) The “patient” from which the bacterium of interest is identified must be human and may be of any age or gender. More than one patient can be included in an emergence event if all patients fell ill and were confirmed to be infected with a resistant bacterium at the same time, for example in the early stages of an outbreak.

c) “Novel” indicates that resistance to a given antimicrobial treatment in the bacterial species in question has not previously been detected or described in the country in question. In this definition, new mechanisms of resistance of a given bacteria to a given antimicrobial combination do not count as novel resistance (e.g. a novel plasmid carrying beta-lactamase in a bacterial species with previously described beta-lactam resistance).

d) “Resistance” is the ability of the bacteria in question to survive standard antibacterial treatment against it. Survival is measured by the bacteria’s ability to continue to cause disease in its host or to spread to others for longer than the standard period following treatment. Standard treatments are determined by the WHO and by the countries in which the articles in the systematic review take place.

e) An “antimicrobial drug or drug combination” is a drug, drug class, or specific combination of drugs used to treat bacterial infections in humans. We focused exclusively on antibiotic resistance rather than resistance to other antimicrobials due to better data availability and based on the hypothesis that drivers for other antimicrobials might be different than for antibiotics.

f) “Bacteria” is a species or strain of pathogen in the domain Bacteria. This study does not assess the impact of resistance in viruses, fungi, or other pathogen types.

For results from the PubMed and Embase search, we programmatically downloaded abstracts and metadata. Article abstracts were manually reviewed by a team of screeners (all authors, see *Author contributions*, below) to determine whether they likely contained a report of a novel emergence event, according to the above criteria. To ensure uniformity in abstract evaluation, all reviewers received training on the inclusion criteria and were required to achieve 90% agreement with a practice set of 100 previously-screened abstracts. Abstracts were screened separately by two individuals. Reviewers classified articles as “yes”, “maybe”, or “no” for inclusion. If both reviewers classified an article as “yes” or “maybe”, the article was downloaded as full-text for further review and to be coded for the database. The inter-scorer agreement rate was 82%. In cases when an article was marked as “no” by one of the reviewers, and “yes” by the other reviewer, a third reviewer was assigned to determine if the article should be included. 1,583 articles from PubMed and Embase passed review criteria and were downloaded for further review. The R package *metagear*⁹ was used to screen articles and manage screening data.

Full texts were downloaded for ProMED-mail search results, as ProMED-mail reports do not have abstracts. The first lines of each report were screened separately by two reviewers. If either reviewer classified a text as “yes” or “maybe”, it was selected for further review and to be coded for the database. 208 ProMED-mail texts passed this screening, contributing to a total of 1,791 total articles from PubMed, Embase, and ProMED-mail selected for review. Further details on the reproducible screening workflow are available in the data repository (see *Data availability*)¹⁰.

Article coding

Articles from PubMed, Embase and ProMED-mail that passed screening were selected for full-text analysis, each by one reviewer, unless quality assurance checks required follow-up review (see *Data cleaning*). Reviewers read full-text articles to determine whether they fully met the case criteria, above, and to extract data by coding the text. Articles were excluded at this stage if it was found that they referred to or were duplicate reports of a previous emergence of the same drug-bacteria resistance within the country, if they reported on a non-bacterial pathogen, or if they did not identify the drug or bacterial species. A total of 294 articles were retained after full-text screening.

Articles were coded for four required fields: study country, drug name, bacteria, and event date. Drugs were coded to the lowest available taxonomic rank (i.e., specific drug rather than drug class, when available). Similarly, bacteria were coded at the species level when available. Where articles did not include an emergence location or date, location was inferred from the location of study authors' institutions (often hospitals), and publication year was used for event date.

In addition to the required fields, articles were coded for the following secondary fields when available: patient attributes (age, gender, country of residence, recent travel locations, symptoms, comorbidities, and outcome), bacterial strains and markers, drug minimum inhibitory concentration (MIC) values, and hospital location (city, state/province). Screeners used MAXQDA¹¹ software for article coding. To our knowledge, open-source equivalents to MAXQDA are not available; however, open-source software such as qcoder¹² could be used to replicate MAXQDA if combined with suitable PDF pre-processing steps.

Data cleaning

We matched free-form values coded in article text to standardized values and ontologies. All locations were matched to Google Place names and geocoded. Country names were maintained according to article reporting and were not standardized to official country recognitions (e.g., United Nations member states). Drug names and bacteria were matched and standardized against the Medical Subject Headings (MeSH)¹³ and National Center for Biotechnology Information (NCBI) Organismal Classification¹⁴ ontologies, respectively, as provided by the Biportal platform¹⁵. Where reported names did not exactly match ontologies or had ambiguous matches, we manually reviewed and corrected names to match the ontologies, in some cases requiring review of the original study to confirm accuracy. Dates were converted to ISO 8601 format. In cases when studies reported a range of dates, only the start date was included in the database. Other, optional fields (patient demographics, etc.) were not standardized in the current database release.

Data cleaning and standardization was performed in R version 3.6.1¹⁶, using the tidyverse framework¹⁷ for data manipulation. Geocoding used the Google Geocoding API via the ggmap R package¹⁸. Study dates were standardized using the lubridate R package¹⁹.

We implemented quality assurance checks throughout the data processing pipeline. We checked for errors in the MAXQDA article coding by confirming that all values were labeled and that links between values were properly assigned (e.g., links between drug names and MIC values). We checked for any studies missing study location, study date, drug, or bacteria, and manually revisited these articles to confirm missing fields. We also investigated any study reporting more than one location and/or date to confirm whether the study described multiple emergence events.

An earlier version of this article can be found on medRxiv (doi: <https://doi.org/10.1101/2020.08.13.20165852>).

Results/Discussion

We present a database of 1,773 records of first clinical reports of unique bacterial-drug AMR detections by country, ranging from 1998 to 2017, drawn from 294 peer-reviewed articles and reports. (While the ProMED-mail search extended to 1994, the first reported event that met our study criteria occurred in 1998.) This database, serving as a complement to existing databases that track resistance and spread, will allow researchers to target efforts for surveillance and interventions and to analyze factors that contribute to AMR emergence.

Database materials are available at DOI [10.5281/zenodo.3964895](https://doi.org/10.5281/zenodo.3964895) (see *Data availability*)¹⁰. Field names and descriptions from the database (filename `events-db.csv`) are detailed in [Table 1](#). The file `data-processed/articles-db.csv` contains metadata about each article in the database, including citation information and full abstracts. This file can be joined with `events-db.csv` by the `study\_id` field.

In addition to the database, we present intermediate data used in the process of database development. All abstracts and ProMED-mail reports that were screened are in the `screening` directory. Raw exports from coded full-text articles are in .xlsx form in `data-raw/coded-segments`. (The coded full text articles themselves are not available in the data repository.) A pre-filtered, pre-transformed database that contains all fields (primary and secondary) and events (including some non-emergent events) is in `data-processed/segments.csv`. Further details on directory structure and usage are provided in the data repository documentation.

We identified AMR emergence events in 66 countries, with the most reported events in the United States (152), India (129), and China (128) ([Figure 2](#)). Events were reported from 1998 through 2017, with the most events occurring in 2011 ([Figure 3](#)). See *Database usage notes* for discussion on the effects of reporting bias on the spatial and temporal coverage of the database.

We found that *Klebsiella pneumoniae* and *Escherichia coli* were the most common bacteria in emergence events ([Figure 4](#)), supporting results from other databases that have found high rates of AMR prevalence for these species^{3,8}. Of concern, our database indicates that both bacteria species had the greatest number of reports of novel resistance to imipenem and meropenem ([Figure 4](#)), which are carbapenem antibiotics often considered last resort treatment options for infections acquired in health care settings^{3,20,21}. Also concerning were 23 cases of emergent resistance to colistin in 23 countries and 14 distinct bacteria, including *K. pneumoniae* and *E. coli*. Colistin is critically important for treating infections when no other options are available^{21,22}.

Database usage notes

The database is developed using only English-language scientific literature and medical reporting events. It therefore represents events reported in this limited subset rather than truly representative clinical cases and strongly reflects reporting effort and practices. It is likely that AMR emergence events are

Table 1. Database fields and descriptions.

Field	Description
`study_id`	Unique study identification number that can be joined with `articles_db` for study metadata
`study_country`	Name of country where event occurred. Note that there are some studies that report on events in multiple countries.
`study_iso3c`	Three letter International Organization for Standardization (ISO) code
`study_location`	Full study location (including hospital, city, and state if available)
`study_location_basis`	Spatial basis of study location (e.g., "hospital, city, state_province_district, country")
`residence_location`	Location of patient residence
`travel_location`	Patient travel locations, if any reported. Multiple locations are separated by `;`.
`drug`	Antimicrobial drug, standardized to the Medical Subject Headings (MeSH) ontology ¹³ . Drug combinations are concatenated by `+`.
`drug_rank`	Taxonomic classification of drug (i.e., drug name or group)
`segment_drug_combo`	TRUE/FALSE resistance is to a combination of drugs
`drug_parent_name`	Name of the taxonomic parent of antimicrobial drug, standardized to the Medical Subject Headings (MeSH) ontology ¹³ .
`bacteria`	Name of resistant bacteria, standardized to NCBI Organismal Classification ontology ¹⁴ .
`bacteria_rank`	Taxonomic classification of bacteria name (e.g., "species", "genus")
`bacteria_parent_name`	Name of the taxonomic parent of bacteria, standardized to NCBI Organismal Classification ontology ¹⁴ .
`bacteria_parent_rank`	Taxonomic classification of bacteria parent name (e.g., "species", "genus")
`start_date`	Date that emergence was reported in format of <i>yyyy-mm-dd</i>
`end_date`	Date that emergence was resolved, if reported, in format of <i>yyyy-mm-dd</i>
`start_date_rank`	Specificity of the start date (i.e., year, month, day)

Figure 2. Global antimicrobial resistance (AMR) emergence events. Points represent locations. Countries are shaded by event count.

Figure 3. Count of global antimicrobial resistance (AMR) emergence events in the database by year.

Figure 4. Global antimicrobial resistance (AMR) emergence events for top 12 drugs (a) and bacteria (b), and for their combinations (c). Counts within bars represent distinct number of resistant bacteria (a) and drugs (b).

systematically missing from the database for countries that are not English speaking and/or have less health care monitoring and reporting capacity, including those with lower GDP. Therefore, it is imperative that any analysis of the data account for the effects of reporting bias¹⁰.

Temporal patterns of AMR emergence events in the database reflect potentially incomplete coverage prior to 2006, as only ProMED-mail (not PubMed or Embase) was searched for this period. Further, the decrease in emergence events from 2011–2017 reflects lags between event occurrence and reporting.

There are several sources of ambiguity in the dataset due to imprecise reporting in literature (Table 2). While most studies

reported the exact city or hospital of the emergence event (n = 200), others reported only state/province (n = 21) or study country (n = 73). Differences in geographic scales would need to be considered if using this dataset for spatial analysis. In addition, there are inconsistencies in studies' reporting of drug and bacteria names. While most studies reported specific drug names (n = 87), others reported broad classes of drugs (n = 35). Study authors that reported classes of drugs may have administered multiple drugs to the patient(s), but because the specific drugs were not reported, we were only able to count the drug as part of a single event. Similarly, bacteria were primarily reported to the species level (n = 106), but some were reported only to the genus or family level (n = 5). In these cases, it is possible that the species was novel or not identifiable. Many

Table 2. Classification specificity of four primary database fields.

	Classification	Count
Drug	drug name	84
	drug group	35
	drug group/name combo	3
Bacteria	species	106
	genus	4
	family	1
Location	hospital	123
	city	77
	state/province/district	21
	country	73
Date	day	47
	month	122
	year	134

studies that reported bacterial species also reported strains, and it is possible that there are differences in resistance by strain. However, due to inconsistent reporting, we have not standardized reported strains.

Finally, the database presents first AMR emergence events by country. Emergence events may be due to either mutation events or transport from other geographies. Analyses examining determinants of first emergence at a country level should consider import as a possible mechanism. Some studies reported locations of residence and recent travel of patients, and where available we have included this information, which may support such analysis. Analyses only examining first global emergence events should remove records of all but the earliest dates of each unique bacterial-drug combination.

Data availability

Zenodo: Extracting novel antimicrobial emergence events from scientific literature and medical reports. <https://doi.org/10.5281/zenodo.3964895>¹⁰.

This project contains the following underlying data:

- `events-db.csv` (AMR emergence events database)
- `data-processed/articles-db.csv` (Metadata about each article in the database, including citation information

and full abstracts. This file can be joined with `events-db.csv` by the `study\_id` field.)

- `screening/` directory (All abstracts and ProMED-mail reports that were screened for potential inclusion in the database)
- `data-raw/coded-segments` (raw exports from coded full-text articles)
- `data-processed/segments.csv` (A pre-filtered, pre-transformed database that contains all fields and events)

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver \(CC0 1.0 Public domain dedication\)](#).

Code availability

Source code available from: <https://github.com/ecohealthalliance/amr-db>

Archived source code at time of publication: <https://doi.org/10.5281/zenodo.3964895>¹⁰

License: [MIT](#)

References

- World Health Organization: **Global action plan on antimicrobial resistance**. 2015.
[Reference Source](#)
- Cosgrove SE: **The relationship between antimicrobial resistance and patient outcomes: mortality, length of hospital stay, and health care costs**. *Clin Infect Dis*. 2006; **42 Suppl 2**: S82–89.
[PubMed Abstract](#) | [Publisher Full Text](#)
- World Health Organization: **Antimicrobial resistance: global report on surveillance 2014**. 2014.
[Reference Source](#)
- Byarugaba DK: **A view on antimicrobial resistance in developing countries and responsible risk factors**. *Int J Antimicrob Agents*. 2004; **24**: 105–110.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Morgan DJ, Okeke IN, Laxminarayan R, *et al.*: **Non-prescription antimicrobial use worldwide: a systematic review**. *Lancet Infect Dis*. 2011; **11**(9): 692–701.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Collignon P: **The importance of a One Health approach to preventing the development and spread of antibiotic resistance**. *Curr Top Microbiol Immunol*. 2013; **366**: 19–36.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Interagency Coordination Group on Antimicrobial Resistance: **No Time to Wait: Securing the Future from Drug-Resistant Infections**. 2019.
[Reference Source](#)
- ResistanceMap: **Center for Disease Dynamics, Economics & Policy**. 2017.
[Reference Source](#)
- Lajeunesse MJ: **Facilitating systematic reviews, data extraction and meta-analysis with the metagear package for R**. *Methods in Ecology and Evolution*. 2016; **7**: 323–330.
[Publisher Full Text](#)
- Konno K, *et al.*: **Ignoring non-English-language studies may bias ecological meta-analyses**. *Ecol Evol*. 2020; **10**(13): 6373–6384.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- VERBI Software, MAXQDA 2018. 2017.
[Reference Source](#)
- Duckles B, Sholler D, Draper J, *et al.*: **qcoder: Lightweight Qualitative Coding**. *R package version 0.1.0*. 2020.
[Reference Source](#)
- National Library of Medicine: **Medical Subject Headings**. 2020.
[Reference Source](#)
- National Library of Medicine: **National Center for Biotechnology Information (NCBI) Organismal Classification**. 2012.
[Reference Source](#)
- Whetzel PL, *et al.*: **BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications**. *Nucleic Acids Res*. 2011; **39**(Web Server issue): W541–545.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- R Core Team: **R Foundation for Statistical Computing, Vienna, Austria**. 2019.
[Reference Source](#)
- Wickham HAM, Bryan J, Chang W, *et al.*: **Welcome to the tidyverse**. *J Open Source Softw*. 2019; **4**: 1686.
[Publisher Full Text](#)
- Kahle D, Wickham H: **ggmap: Spatial Visualization with ggplot2**. *The R Journal*. 2013; **5**: 144–161.
[Reference Source](#)
- Grolemund G, Wickham H: **Dates and Times Made Easy with {lubridate}**. *J Stat Softw*. 2011; **40**: 1–25.
[Reference Source](#)
- Codjoe FS, Donkor ES: **Carbapenem Resistance: A Review**. *Med Sci (Basel)*. 2017; **6**(1): 1.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- World Health Organization: **Critically Important Antimicrobials for Human Medicine**. 2016.
[Reference Source](#)
- Paterson DL, Harris PNA: **Colistin resistance: a major breach in our last line of defence**. *Lancet Infect Dis*. 2016; **16**(2): 132–133.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 15 February 2021

<https://doi.org/10.5256/f1000research.29674.r78874>

© 2021 Stobberingh E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ellen Stobberingh

Care And Public Health Research Institute (CAPHRI), Maastricht University Medical Centre (MUMC), Maastricht, The Netherlands

The authors analysed the prevalence of novel AMR in humans in the country. Several questions need to be clarified:

- Inclusion criteria:

- AMR in asymptomatic commensals were not included as these do not meet the inclusion criteria of clinical cases. However, resistance in commensals is frequently the precursor of resistance in clinical isolates. Excluding the commensals might contribute to lower prevalence and later report of resistance.
- Focus on humans only did not address the One health approach which is especially relevant in AMR because of the spread from animals to humans. This should be added in the discussion.
- Novel: was defined as novel for that country. But the resistance might be described already in other countries. Transfer from these countries for instance via travel might contribute to spread of these AMR to other countries and is not a novel resistance. Pick up of AMR resistance during travel starts mostly as a commensal and later on possible as infection. Please comment.
- Resistance: defined as to survive AMR treatment, i.e. the bacteria *is able to continue to cause disease in the host or to spread to others for a longer than the standard period as determined by the WHO or the national guidelines*. What to do in case of discrepancy between these two?
- Interscore 82% in case of discrepancy the decision of a third reviewer was decisive?
- What is the possible explanation for the highest prevalence in 2011?
- Resistance to carbapenems as last resort, colistin when no other options are available. This suggests a discrepancy, please rephrase.

Is the rationale for creating the dataset(s) clearly described?

Partly

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Antimicrobial resistance prevalence and spread, Bacteriology.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 10 Jun 2021

Emma Mendelsohn, EcoHealth Alliance, New York, USA

Thank you for your thorough review and comments (especially appreciate your thoughts on the One Health context and potential spread mechanisms). We address your comments below.

AMR in asymptomatic commensals were not included as these do not meet the inclusion criteria of clinical cases. However, resistance in commensals is frequently the precursor of resistance in clinical isolates. Excluding the commensals might contribute to lower prevalence and later report of resistance.

We agree that asymptomatic commensals are important to identifying resistance before clinical emergence. However, we did not include commensals in the database as we found there is high variability in study design among these studies (e.g., drugs tested). Clinical reports, while still subject to variability and bias, tend to be more consistent in design (reporting only on resistances observed with illness). We address this concern in the *Abstract and report screening* section of the manuscript.

Focus on humans only did not address the One health approach which is especially relevant in AMR because of the spread from animals to humans. This should be added in the discussion.

This is a good point and we have added text about potential risk factors (including livestock) for emergence and the relevance to One Health in the discussion section.

Novel: was defined as novel for that country. But the resistance might be described already in other countries. Transfer from these countries for instance via travel might contribute to spread of these AMR to other countries and is not a novel resistance. Pick up of AMR resistance during travel starts mostly as a commensal and later on possible as infection. Please comment.

We agree and note that emergence events within a country may be due to mutation or transport from other geographies. For first global emergence events, users of the database should filter to the first occurrence of each bacteria-drug combination. We have added discussion about travel and migration being potential risk factors for emergence in countries.

Resistance: defined as to survive AMR treatment, i.e. the bacteria is able to continue to cause disease in the host or to spread to others for a longer than the standard period as determined by the WHO or the national guidelines. What to do in case of discrepancy between these two?

As most authors did not report which guidelines they used, we deferred to the authors selection of protocols and did not evaluate any potential discrepancies. We have added text to clarify this.

Interscore 82% in case of discrepancy the decision of a third reviewer was decisive?

Yes; we have rephrased to clarify this.

What is the possible explanation for the highest prevalence in 2011?

We have rephrased the text in *Data usage notes* to clarify that the peak in 2011 is likely an effect of reporting lag for more recent events.

Resistance to carbapenems as last resort, colistin when no other options are available. This suggests a discrepancy, please rephrase.

Good catch. We have rephrased.

Competing Interests: No competing interests were disclosed.

Reviewer Report 07 December 2020

<https://doi.org/10.5256/f1000research.29674.r75116>

© 2020 Eremin S et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Barbara Tornimbene

World Health Organization, Geneva, Switzerland

Sergey Eremin

World Health Organization, Geneva, Switzerland

The paper represents an impressive effort to fill in the gaps in our understanding of emergence and spread of antimicrobial resistance in bacterial pathogens and provides results of a systematic review of English-language scientific literature and surveillance reports to identify the spatial and temporal patterns of the emerging AMR.

Is the rationale for creating the dataset(s) clearly described?

Although the author rationale for generating a publicly available source of temporal and spatial data for emerging resistance is clear, it could be better explained in the title and the introduction that the creating of a database is the objective of the study. The authors could also go into more detail on how the database will support on-going agreements and programs, particularly by defining on-going agreements. It would be also important to clarify what will be the role of the government of countries that appear in the dataset, in terms of acknowledging and validating published data.

Are the protocols appropriate and is the work technically sound?

Merging of results originating from peer-reviewed literature, which undergoes a strict scientific scrutiny, and data generated by ProMED-mail, which does not follow the same rigour, can be exposed, particularly when reporting novel events, to high level of inconsistency. No distinction between study designs screened in the literature was applied and results adjusted accordingly. The exclusion criteria could be expanded. For example, no assessment of the quality and reliability of these studies was done before including them in the final pool of articles.

The search strategy may benefit from inclusion of terms related to “drug resistance”, the term used in quite a number of reports instead of “antibiotic resistance” or “antimicrobial resistance”, especially when reporting infections caused by MDR or XDR or PDR pathogens.

Finally, the authors mentioned in the “Database usage notes” section that the reported events represent a limited subset of information and strongly reflect reporting effort and practices. Other limitations are also listed in this section. This should probably be part of the discussion and better reflected on the abstract when listing the most common bacteria demonstrating new resistance. Moreover, an in-depth discussion on the impact of these bias should be added, particularly the effect that variability in reporting effort and practices, and quality and reliability of the studies, have on obtained results.

Are sufficient details of methods and materials provided to allow replication by others?

Some definitions in the inclusion criteria are not sufficiently clear.

A clinical case is defined as “an individual who presents with symptoms to a medical professional and is determined by a medical professional to have been infected with the bacterium in question”. This sounds rather vague and we believe that the authors could be more consistent when defining “bacterium in question” or “bacterium of interest”. This is important to understand how cases were included.

The definition of novel resistance seems to be uncertain. It has to be clearly explained in the paper why only phenotypic test results were chosen and why only specific drug-bug combinations are considered and not emergence of certain types of multidrug resistance. The authors might also want to clarify why detection of a novel gene responsible, using example given in the paper, for production of a new type of beta-lactamase or, specifically, carbapenemase, is not considered novel and should be ignored. Such an event has to be detected and followed as it may have a significant impact on treatment options and other public health implications. Again, more details should be given on how the “novelty” was identified when reviewing the papers. For example, the paper coded 15378 in the database describes both phenotypic and molecular resistance testing results obtained from 12 clinical isolates from 12 patients with infections caused by *Enterococcus faecium* and 11 supposedly novel resistance events were included in the database. But the only event that was defined by the authors of the report as novel was the first detection in the country of the clonal complex 17 VREF. Resistance to e.g. clindamycin, gentamicin, and several other antibiotics in all reported isolates could hardly be considered novel or emerging.

The authors should better explain the added value of including the “drug_parent_name” variable with the names of the taxonomic parent of antimicrobial drug, standardized to the MeSH ontology. While it may be reflecting the methodology of the database creation, using a classification more suited to the field of study, such as e.g. the Anatomical Therapeutic Chemical (ATC) classification system may be considered.

Finally, as the main goal of the database is to present temporal and spatial data, a better definition of temporal data should be given. In the paper the authors mention “event date”, “start date”, “study date” and “end date”. Aside the lack of consistency it should be clearer if the authors are considering the date the patients started experiencing symptoms, the date the patients was diagnosed, the date the diagnosis results were obtained, or the date the study started. This is key to allow replication by others.

Are the datasets clearly presented in a useable and accessible format?

The detailed dataset is easily accessible and could be downloaded but is not sufficiently user-friendly. Using it would require certain data management skills from the end-users and therefore might limit their number.

Is the rationale for creating the dataset(s) clearly described?

Partly

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Surveillance of antimicrobial resistance.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 10 Jun 2021

Emma Mendelsohn, EcoHealth Alliance, New York, USA

Thank you for your thorough review. We appreciate your attention to detail and thoughtful comments, which we address below.

Is the rationale for creating the dataset(s) clearly described?

Although the author rationale for generating a publicly available source of temporal and spatial data for emerging resistance is clear, it could be better explained in the title and the introduction that the creating of a database is the objective of the study. The authors could also go into more detail on how the database will support on-going agreements and programs, particularly by defining on-going agreements. It would be also important to clarify what will be the role of the government of countries that appear in the dataset, in terms of acknowledging and validating published data.

We agree and have emphasized the database as a study objective by changing the title to “A global repository of novel antimicrobial emergence events”. We also added details on the rationale and main goals of the paper in the introduction, highlighting the relevance of this dataset in supporting on-going international agreements and programs.

Are the protocols appropriate and is the work technically sound?

Merging of results originating from peer-reviewed literature, which undergoes a strict scientific scrutiny, and data generated by ProMED-mail, which does not follow the same rigour, can be exposed, particularly when reporting novel events, to high level of inconsistency. No distinction between study designs screened in the literature was applied and results adjusted accordingly. The exclusion criteria could be expanded. For example, no assessment of the quality and reliability of these studies was done before including them in the final pool of articles.

Thank you for the relevant comment. This is an important point to be considered by the end-users. We have clarified in the manuscript that we did not conduct an assessment of study/report quality and have described this as a source of uncertainty. In addition, we have added a field to the database indicating whether an event is from a published study or from ProMED-mail, and we include in the text summarized counts of events by source type.

The search strategy may benefit from inclusion of terms related to “drug resistance”, the term used in quite a number of reports instead of “antibiotic resistance” or “antimicrobial resistance”, especially when reporting infections caused by MDR or XDR or PDR pathogens.

We agree that “drug resistance” may yield relevant results. Unfortunately we are unable to

redo the existing literature search at this time due to the high level of effort required and additional human resources not available at this time. We will include this term in future iterations of the project.

Finally, the authors mentioned in the “Database usage notes” section that the reported events represent a limited subset of information and strongly reflect reporting effort and practices. Other limitations are also listed in this section. This should probably be part of the discussion and better reflected on the abstract when listing the most common bacteria demonstrating new resistance. Moreover, an in-depth discussion on the impact of these bias should be added, particularly the effect that variability in reporting effort and practices, and quality and reliability of the studies, have on obtained results.

We thank the reviewer for the suggestion. Bias in reporting effort is indeed a recurrent problem in surveillance. We have added mention of reporting biases to the abstract and discussion section (referring readers to Database usage notes for further detail). We also now address quality of studies as a source of uncertainty in the database.

Are sufficient details of methods and materials provided to allow replication by others?

Some definitions in the inclusion criteria are not sufficiently clear.

A clinical case is defined as “an individual who presents with symptoms to a medical professional and is determined by a medical professional to have been infected with the bacterium in question”. This sounds rather vague and we believe that the authors could be more consistent when defining “bacterium in question” or “bacterium of interest”. This is important to understand how cases were included.

We adjusted the wording to clarify that bacteria is any species of pathogen in the domain Bacteria, and have replaced the term “bacterium in question” with “bacterium species reported”.

Due to inconsistencies in reporting, we were unable to consistently identify pathogens by strain. However, this data is available as an unstandardized field in the database and we added .

The definition of novel resistance seems to be uncertain. It has to be clearly explained in the paper why only phenotypic test results were chosen and why only specific drug-bug combinations are considered and not emergence of certain types of multidrug resistance. The authors might also want to clarify why detection of a novel gene responsible, using example given in the paper, for production of a new type of beta-lactamase or, specifically, carbapenemase, is not considered novel and should be ignored. Such an event has to be detected and followed as it may have a significant impact on treatment options and other public health implications. Again, more details should be given on how the “novelty” was identified when reviewing the papers. For example, the paper coded 15378 in the database describes both phenotypic and molecular resistance testing results obtained from 12 clinical isolates from 12 patients

with infections caused by *Enterococcus faecium* and 11 supposedly novel resistance events were included in the database. But the only event that was defined by the authors of the report as novel was the first detection in the country of the clonal complex 17 VREF. Resistance to e.g. clindamycin, gentamicin, and several other antibiotics in all reported isolates could hardly be considered novel or emerging.

We agree this is a source of uncertainty in the database. In response to your comment, we have revisited the drug name standardization and disaggregated all drug combinations, as we had not been consistent in identifying combinations. We are now treating all drugs as being separately administered. Unfortunately, due to inconsistencies in reporting, we are not able to effectively characterize which drugs were administered independently and which were administered as part of a complex. We address this uncertainty in the Database usage notes.

The authors should better explain the added value of including the “drug_parent_name” variable with the names of the taxonomic parent of antimicrobial drug, standardized to the MeSH ontology. While it may be reflecting the methodology of the database creation, using a classification more suited to the field of study, such as e.g. the Anatomical Therapeutic Chemical (ATC) classification system may be considered.

Based on your comments we explored standardizing drug names to ATC. We found that results were very similar to MeSH (i.e., counts changed only slightly) . We discuss this in the manuscript and have added an alternate version of the database with ATC standardization to the project repository.

Finally, as the main goal of the database is to present temporal and spatial data, a better definition of temporal data should be given. In the paper the authors mention “event date”, “start date”, “study date” and “end date”. Aside the lack of consistency it should be clearer if the authors are considering the date the patients started experiencing symptoms, the date the patients was diagnosed, the date the diagnosis results were obtained, or the date the study started. This is key to allow replication by others.

We switched to consistent usage of “start date” and clarified that it refers to the date the patient presented to the hospital or clinic.

**Are the datasets clearly presented in a useable and accessible format?
The detailed dataset is easily accessible and could be downloaded but is not sufficiently user-friendly. Using it would require certain data management skills from the end-users and therefore might limit their number.**

We added a sentence to clarify that the database is available for download as a single csv file (`events-db.csv`)

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research