**ABSTRACT**

The Emerging Infectious Disease Repository (EIDR) combines a curated and transparent database on the occurrence, characteristics, and drivers of emerging infectious disease (EID) reports between 1940 and 2013 with an interactive web application built to communicate this information. The EIDR database contains 350 EID events and is available at <https://eidr.ecohealthalliance.org/> . The composition of the EIDR database builds existing EID databases and also reveals the limitations of compiled EID case studies. The challenge associated with collecting and standardizing comprehensive information on complex historical events is considerable, including precisely defining and discerning key aspects of EIDs.

**INTRODUCTION**

The World Health Organization estimated that in 2012 infectious and parasitic diseases accounted for 15.8% of all disability-adjusted life-years (DALYs) (WHO 2014). Although this is a 4% decrease from an estimate of 19.5% of all DALYs in 2000, the rate of pathogen emergence is increasing(Jones et al. 2008; Pike 2014) and the opportunity for pandemic emergence remains high (World Health Organization 2014; Woolhouse ME 2008). Globalization has produced an efficient mechanism to spatially distribute EIDs. A mere 20 years after HIV’s discovery, the virus was the fourth highest cause of death worldwide (Smolinski et al. 2003). More recently, the vulnerability of global health security was revealed in the ongoing and largest-ever Ebola Virus Disease epidemic in West Africa. This epidemic has caused well over 8,000 deaths in one year, and incurred economic damages of between $3.8 and $32.6 billion internationally (WHO 2014). To combat these EID threats the emerging infectious disease community must understand the driving factors underlying disease emergence so that informed and effective prevention, preparation, and response strategies can be developed.

Studying the origins of EIDs is complex and many types of methods (qualitative and quantitative) have been used (WHO 2014; Smolinski et al. 2003; Grace et al;Taylor et al. 2001; Weiss and McMichael 2004; Woolhouse et al. 2012; Morens et al. 2004; Funk et al. 2013). One method has focused solely on identifying and studying EID reports to identify a disease’s earliest known emergence event (Pike et al. 2014; Grace et al. 2012). This approach allowed analysis of the geographic and temporal trends in EIDs and yielded the first map of EID hotspots (Jones et al. 2008). Despite the apparent face validity of combining multiple independent EID events (case studies and reports) for EID spatial prediction, this method is limited by the complexity surrounding historic emergence events and difficulties in finding and validating underlying data (Funk et al. 2013). The Emerging Infectious Disease Repository (EIDR) was developed to deal with these limitations. It was derived largely from the database used in (Jones et al. 2008), with the goal of exploring EID events in greater detail by creating an expanded, highly curated, database of EID events and to help correct some of these apparent methodological issues and to better communicate EID case studies to the EID research community and general public.

**METHODS**

***EID Event Collection***

For the purpose of EIDR, an EID event is defined as the original case or cluster of cases representing the emergence of an infectious disease in human populations (Jones et al. 2008)*.* Emergence is defined as the development of any of the following with respect to a given microorganism: (a) earliest instance of natural human infection; (b) reappearance after control or elimination; (c) new or increasing drug resistance; (d) new or expanding geographic region; (e) increasing incidence; or, (f) increasing virulence. Potential EID events were discarded if no clinical significance or relevance could be attributed to the pathogen in question. All potential EID events were evaluated based the EID definition above and were evaluated by infectious disease experts at EcoHealth Alliance. See Supplementary Materials for detailed descriptions of these emergence categories.

The events in EIDR date back to 1940, a cut-off chosen by Jones et al. (2008), and informed by the Institute of Medicine’s (IOM) resources on EIDs (Smolinski, 2003). Most of the infectious diseases classified as emerging by the IOM likely emerged after 1940. Potential EID events were collected from a review of meta-analyses on disease emergence and via literature review. Some of the events between 1940 and 2004 were derived from the Jones et al. study (2008) that expanded on a previously published EID list (Taylor et al. 2001). Events between 2004 and 2013 derive from a recent effort to map emerging zoonoses (Grace et al. 2012), a review of trends in viral discovery (Rosenberg et al. 2013), or were compiled through a systematic literature review. The sample size of events is provided in the results section.

***Data Collection and Review***

For each EID event, data were collected on a set of variables identified as important by a team of EID subject matter experts. These variables are designed to capture critical spatial, temporal, clinical, epidemiologic, economic, pathogen, and host information. Driver categories were based on those published by the IOM (2003), Lederberg et al. (1992), and Jones et al. (2008). Additionally, data were collected on potential drivers associated with each EID event and published on EIDR (Jones et al. 2008*,* Smolinski et al. 2003; Lederberg et al.1992).

The following EID drivers are used in EIDR; (a) international travel and commerce; (b) breakdown of public health measures; (c) climate and weather; (d) war and famine; (e) human susceptibility to infection; (f) antimicrobial agent use; (g) ecosystem changes; (h) medical industry changes; (i) human behavior; (j) proximity to wildlife; and, (k) agricultural industry changes (Jones et al. 2008; Smolinski et al. 2003; Lederberg et al. 1992). Table 1 provides a list of all EIDR variables. Definitions of all variables, and their sub-variables, are located in Supplementary Materials. EIDR contains 350 EID events; 296 (out of a total of 335) from Jones et al. (2008), 38 from Grace et al. (2012), 6 from Rosenberg et al. (2013), and 10 from this study’s internal literature review. For a list of the 39 events from the Jones et al. (2008) database that were excluded from the EIDR database see Supplementary Materials.

Infectious disease emergence locations were resolved to the most specific spatial information available and this was frequently a point representing the smallest administrative region associated with an event. Rarely, multiple potential locations are provided for a single emergence event due to insufficient spatial temporal information within the available literature. For example, a report describing an event that includes simultaneous confirmed cases of a given disease from two adjacent towns would include both locations in EIDR.

Short abstracts were written for all events. When possible, direct language from text was captured to justify values for subjective variables. If no information could be found on a particular variable the absence of data was captured. General contextual information for each event was acquired from various sources, some of which may be unrelated to EID events. For example, taxonomic information is from the National Center for Biotechnology Information and economic information is from the World Bank (NCBI 2015; WBG 2015). EID subject matter experts individually reviewed each EID event contained in EIDR twice at a minimum.

***Statistics***

Chi-square tests were used to compare the distribution of categorical variables within the EIDR database and Jones et al. database (2008). Data collection methods and criteria for EID event inclusion differed between studies, so the datasets are predominantly independent and appropriate for the chi-square test.

**RESULTS**

***EID Events***

The EIDR database is dominated by EID events caused by bacteria (50.0%) and viruses (31.7%). Vector-borne diseases are associated with 22.0% of EID events in EIDR. EID events occurred primarily in North America (31.7%), Europe (24.3%) and Asia (18.3%), although no adjustment has been made to offset potential bias (e.g., information, surveillance, reporting).

Events representing the earliest instance of natural human infection by a microorganism are the most numerous (55.7%), followed by events representing new or expanding drug resistance (20.9%), increasing incidence (6.9%), new or expanding region (6.3%), increased virulence (5.7%), and reemergence after control or elimination (5.1%). The most commonly identified cause of EID events is the use of antimicrobial agents (20.3%). Other significant causes include human susceptibility to infection (18.9%), proximity to wildlife (11.1%), human behavior (11.1%), and ecosystem change (10.6%). Notably, no cause could be identified in 28.9% of EID events.

The majority of EID events involve known zoonotic pathogens (63.1%). Just over half of these zoonotic EID events involve a specifically documented instance of transmission of a microorganism from animals to humans during the event (52.0% of zoonotic EID events, 32.9% of all EID events). In EID events involving zoonotic pathogens but without a specifically documented instance of animal-to-human transmission, the transmission pathway is primarily unknown (63%), although nosocomial transmission occurred in a notable number (13.9%).

***EIDR Web Application***

An interactive web-application displays the information stored in EIDR (http://eidr.ecohealthalliance.org). Through EIDR website, EID events can be explored in a variety of ways. The “Emergence Events” view displays EID events in a table. Users can choose which EIDR variables they would like to view (Fig. 1), and perform specific searches using a filter feature that allows users to search for events with a common variable, like a specific host, or pathogen. A map of all EID events is offered through the “Event Map” view (Fig. 2). Additional methods and variable definitions are available on the “About”, and “Variable Definitions” pages.

Users can explore individual EID events in greater detail through individual event pages. Clicking on an event in the “Emergence Events” table can access event pages. Each EID event page contains a detailed report on the event, including a narrative abstract, a map showing the location of the event, tables of additional data, and a discussion board that allows users to comment on the event (Fig. 3). In some cases data are displayed with supporting textual evidence. References for each event are available in the event pages.

**Statistical Testing**

A Chi-sqaure test was used to compare the Jones et al. database to the EIDR database. The Jones et al. database contain comparable percentages of EID events associated with zoonotic diseases (63.1%, 60.3%, p = 0.444), vector-borne diseases (22.4%, 22.8%, p = 0.900), bacteria (50.0%, 54.3%, p = 0.257), and viruses (31.7%, 25.4%, p = 0.066).

**DISCUSSION**

EIDR is the combination of an expansive, highly curated, and transparent database of EID events with a user-friendly, engaging, and interactive design. The composition of the EIDR database largely replicates the findings of Jones et al. (2008) albeit that this manuscript doe not contain temporal or geographical trends. Chi-square tests demonstrate that EIDR and the Jones et al. database contain comparable percentages of EID events associated with zoonotic diseases (63.1%, 60.3%, p = 0.44), vector-borne diseases (22.4%, 22.8%, p = 0.90), bacteria (50.0%, 54.3%, p = 0.25), and viruses (31.7%, 25.4%, p = 0.06). This study and Jones et al. (2008) identified antimicrobial agent use as the most common cause of EID events. The large number of resistant microbes and the large number of cases of multiple resistance not identified in EIDR could easily swamp out other classes of EID events, given the rapid and ongoing discovery of resistant strains in patients.

Despite the verification process used to construct EIDR, it uncovers some of the limitations of using of using individual EID reports to study trends in disease emergence. EID events are often complex and EID case studies have been subjected to varying levels of scrutiny. Elucidating the geographic or temporal origin and drivers of even well-known EIDs (e.g., HIV-1 or Nipah virus) can take many years, and often involves a combination of microbiological analyses, multi-year ecological studies, and public health investigations (Faria et al. 2014; Pulliam et al. 2012). Identifying the root causes of events is critical to understanding disease emergence, but is often difficult or impossible to discern root causes. This is particularly true for zoonotic disease transmission events that may require multiple studies to determine host involvement. In EIDR, for most events involving zoonotic diseases no transmission route could be identified. Additionally, substantial labor is required to create reliable EID event databases, with EIDR itself taking several years to complete. Although future studies can use the EIDR database to reduce EID investigation time, maintaining the EIDR database so that it reflects the most current definition of emergence, and contains the most recent EID events is a formidable task. A specific limitation of EIDR is that the database only contains events gathered from published scientific literature. EIDR does not contain EID events that were not published in peer-reviewed literature. Verifying and validating potential EID events that have been noted in previous studies is sometimes difficult. Extensive effort was made to search for these potential EID events, often with limited success. Lastly, for many EID events and in particular those that are classed as emerging due to increasing incidence or geographic expansion, the identification of when or where they emerged can be very difficult. This is particularly so for those without large caseload datasets and a more quantifiable definition for EIDs seems to be a critical goal the classification of events as EID events is shrouded in ambiguity. This studies finding’s support Funk et al.’s (2013) argument that a more quantifiable and objective EID definition is needed for EIDs.

The transparency and accessibility of the EIDR database is made possible by the EIDR web-application and should spur constructive conversations about the most effective methods to study and define disease emergence objectively. Future research should be conducted to determine and create more objective and comparable definitions for emerging infectious diseases so the scientific community can make effective comparisons between diseases. Computational methods of gathering information on EIDs will make it easier to quantify emergence and should be fervently explored.

**References**

Grace D, Mutua F, Ochungo P, Kruska R, Jones K, Brierley L, Lapar ML, Said M, Herrero M, Phuc PM, Thao NB, Akuku I, Ogutu F. 2012. "Mapping of poverty and likely zoonoses hotspots." In. Nairobi Kenya: ILRI.

Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. The early spread and epidemic ignition of HIV-1 in human populations. Science. 2014;346(6205): 56–61. doi:[10.1126/science.1256739](http://dx.doi.org/10.1126/science.1256739" \t "pmc_ext)

Funk S, Bogich TL, Jones KE, Kilpatrick AM, Daszak P. 2013. Quantifying trends in disease impact to produce a consistent and reproducible definition of an emerging infectious disease, PLoS One 8: e69951.

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451: 990–993. doi: [10.1038/nature06536](http://dx.doi.org/10.1038/nature06536" \t "pmc_ext)

Lederberg J, Shope RE, Oaks SC. 1992. Emerging Infections. Microbial Threats to Health in the United States. Washington, DC: Natl. Acad. Press. 294 pp.

Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases.Nature. 2004;430:242–9 and. 10.1038/nature02759

NCBI. 2015. Home - Taxonomy - NCBI. http://www.ncbi.nlm.nih.gov/taxonomy

World Bank Group. 2015. <http://www.worldbank.org/en/topic/health/brief/world-bank-group-ebola-fact-sheet>

Pike J, Bogich T, Elwood S, Finnoff DC, Daszak P. Economic optimization of a global strategy to address the pandemic threat. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(52):18519-18523. doi:10.1073/pnas.1412661112.

Pulliam JR, Epstein JH, Dushoff J, Rahman SA, Bunning M, Jamaluddin AA, et al. Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis. J R Soc Interface. 2012;9:89–101. 10.1098/rsif.2011.0223

Rosenberg R, Johansson MA, Powers AM, Miller BR. 2013. Search strategy has influenced the discovery rate of human viruses. Proc Natl Acad Sci U S A 110: 13961–13964.

Smolinski MS, Hamburg MA, Lederberg J, Institute of Medicine (U.S.). Committee on Emerging Microbial Threats to Health in the 21st Century Microbial threats to health: emergence, detection, and response. Washington, DC: National Academies Press; 2003.

Taylor L, Latham S, Woolhouse M. Risk factors for human disease emergence. Philos Trans R Soc B.2001;356:983–9. doi: 10.1098/rstb.2001.0888.

Weiss RA, McMichael AJ. Social and environmental risk factors in the emergence of infectious diseases. Nat Med. 2004;10:S70–S76.

Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. 2012. Human viruses: discovery and emergence. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367:2864–2871.

Woolhouse ME, Howey R, Gaunt E, Reilly L, Chase-Topping M, Savill N. Temporal trends in the discovery of human viruses. *Proceedings of the Royal Society B: Biological Sciences*. 2008;275(1647):2111-2115. doi:10.1098/rspb.2008.0294.

World Health Organization. 2014. 'WHO Global Health Estimates', *WHO*.

|  |  |  |
| --- | --- | --- |
| **Emergence Driver Variable** | **EIDR** | **Jones et al. (2008)** |
| Average age of death | X |  |
| Average age of infected | X |  |
| Disease | X | X |
| Driver (e.g., ecosystem changes, international travel, war and famine) | X | X |
| Drug resistance | X | X |
| Duration of event | X |  |
| EID category (e.g., earliest instance of natural human infection, reappearance after control or elimination) | X |  |
| End date | X |  |
| End date description | X |  |
| Event transmission | X |  |
| General transmission | X |  |
| Host age | X |  |
| Host use | X |  |
| Initially reported name | X |  |
| Life expectancy in the first year of the event | X |  |
| Location | X | X |
| Number infected | X |  |
| Number of deaths | X |  |
| Occupation | X |  |
| Pathogen host(s) | X |  |
| Pathogen type | X | X |
| Per capita national GDP in the first year of the event | X |  |
| Reported symptoms | X |  | |
| Specific host(s) involved in the event | X |  | |
| Start date | X | X | |
| Start date description | X |  | |
| Taxonomic information | X |  | |
| Testing method | X |  | |
| Transmission of the microorganism from animals to people | X |  | |
| Vector-borne | X | X | |
| Zoonosis (Not event specific) | X | X | |

Table 1. A list of all EIDR variables and whether they were present in the Jones et al. (2008) data.