**ABSTRACT**

Data associated with emerging infectious disease events underpin many geographic and temporal analyses of emerging infectious disease. The Emerging Infectious Disease Repository (EIDR) aims to create an accurate database on the occurrence, characteristics, and drivers of emerging infectious disease events and to communicate the data through an interactive web application. Potential emergence events were collected from a review of meta-analyses on disease emergence and via an exhaustive literature review (e.g., PubMed, ProMED-mail, WHO Weekly Epidemiological Record, CDC Morbidity and Mortality Weekly Report). All EIDR data were validated through published sources and evaluated by experts. EIDR contains 350 events between 1940 and 2013. The most commonly identified drivers of emergence were the use of antimicrobial agents, followed by human susceptibility to infection, proximity to wildlife, human behavior and ecosystem change. The majority of emergence events involve zoonotic pathogens. Critical gaps were identified. The specific transmission pathway is unknown for nearly 1/3rd of the zoonotic pathogen emergence events. No drivers were identified in approximately 30% of all events. EIDR advances the ability to confront challenges associated with collecting and accessing comprehensive information on complex historical events by including many additional event variables and facilitating future efforts through an accessible platform.

**INTRODUCTION**

In 2012, infectious and parasitic diseases accounted for 15.8% of all disability-adjusted life-years and the opportunity for pandemic emergence remains high (Jones et al., 2008; WHO, 2014). Globalization has produced an efficient mechanism to spatially distribute emerging infectious diseases (Smith et al., 2007). The vulnerability of global health security has recently been revealed through the emergence of Zika Virus in the Americas and by the largest-ever Ebola Virus Disease epidemic in West Africa (Heymann et al., 2015; PAHO/WHO, 2016). To combat these 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 emerging infectious diseases is complex and many types of methods (qualitative and quantitative) have been used (Taylor et al. 2001; Smolinski et al., 2003; Morens et al., 2004; Weiss and McMichael, 2004; Grace et al., 2012; Woolhouse et al., 2012; Funk et al., 2013; WHO, 2014). One method has focused on studying case reports to identify a disease’s earliest known emergence event (Grace et al., 2012; Pike et al., 2014). This approach allowed analysis of the geographic and temporal trends in emerging infectious diseases and yielded the first map of emerging infectious disease hotspots (Jones et al., 2008). This method is limited by the complexity surrounding historic disease 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 infectious disease emergence events in greater detail. The repository created an expanded, curated, database of these events that now provides the research community with a platform to track the sources of all relevant data.

**METHODS**

***Definition and Collection of Emerging Infectious Disease Events***

For the purpose of EIDR, an emerging infectious disease 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)*.* Within EIDR, 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. To be included within EIDR an event had to meet one of the above definitions and required a specific start date, location, an identified pathogen and an associated human infection. 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), which was informed by the Institute of Medicine’s examples which provided data that many emerging infectious diseases have occurred since 1940 (Smolinski et al., 2003). Potential emergence events were not evaluated if they occurred after 2013. Potential emergence events were collected from a review of meta-analyses on disease emergence (Taylor et al., 2001; Jones et al., 2008; Grace et al., 2012; Rosenberg et al., 2013) and via an exhaustive literature review (e.g., Web of Science, PubMed, ProMED-mail, WHO Weekly Epidemiological Record, CDC Morbidity, and Mortality Weekly Report). Meta-analyses not specific to disease emergence were not included. Some of the events between 1940 and 2004 were derived from the Jones et al. study (2008) that expanded on a previously published list of emerging infectious disease events by 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 systematic literature review. EIDR contains 350 emerging infectious disease 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 literature review. 39 events were excluded from the Jones et al. (2008) database EIDR database and are available in the supplementary materials.

***Data Collection and Review***

For each emergence event, data were collected on a set of 31 variables identified as important by a team of emerging infectious disease subject matter experts. These variables are designed to capture critical spatial, temporal, clinical, epidemiologic, economic, pathogen, and host information. Driver categories mostly derive directly from those published by the The Institute of Medicine (2003), Lederberg et al. (1992), and Jones et al. (2008), with the exception of medical industry changes, human behavior, proximity to wildlife, and agricultural industry changes, which are more specific derivatives of the drivers listed in Smolinski et al. (2003) and Lederberg et al. (1992). Table 1 provides a list of all EIDR variables. A complete list of drivers and their definitions is provided in Table 2. Data collection for many variables, such as "Driver” and “Type of Emergence”, was contingent on qualitative information from sources (e.g., a description of infected wildlife living near the event location merited the "Driver” value “Proximity to Wildlife”). Definitions of all variables, and their sub-variables, are located in supplementary materials.

Infectious disease emergence locations were resolved to the most specific spatial information available. 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 coded during data collection. All variables missing data for a given event were displayed within an event’s “Event Page” of the EIDR website. General contextual information for each event was acquired from various sources, some of which may be unrelated to emergence 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). Emerging infectious disease subject matter experts individually reviewed each emergence event contained in EIDR twice at a minimum. Experts used the original sources to verify that all data for each event was complete and correctly classified into the appropriate variables. More than a month of expert time was devoted to the review process.

***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 emergence event inclusion differed between the studies. Therefore, the data within each dataset are independent despite containing some overlapping data. This makes the datasets appropriate for the chi-square test.

***EIDR Web Application***

An interactive web-application displays the information stored in EIDR (http://eidr.ecohealthalliance.org). Through the EIDR website, emergence events can be explored in a variety of ways. The “Emergence Events” view displays 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 (e.g., a specific host or pathogen). A map of all emergence 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 emergence events in greater detail through individual event pages. Clicking on an event in the “Emergence Events” table can access event pages. Each 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.

**RESULTS**

***Emerging Infectious Disease Events***

The EIDR database is dominated by emerging infectious disease events caused by bacteria (50.0%) and viruses (31.7%). Vector-borne diseases are associated with 22.0% of emergence events in EIDR. 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 emergence 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 events.

The majority of emergence events involve known zoonotic pathogens (63.1%). Just over half of these zoonotic emergence events involve a specifically documented instance of transmission of a microorganism from animals to humans during the event (52.0% of zoonotic emergence events, 32.9% of all emergence events). In emergence 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%). Distilling these results, the transmission pathway is unknown for roughly 1/3 of all events involving zoonotic pathogens.

**DISCUSSION**

EIDR advances the ability to confront challenges associated with collecting and accessing comprehensive information on complex historical events by including many additional event variables and facilitating future efforts through an open access Internet application. The composition of the EIDR database largely replicates the findings of Jones et al. (2008), but EIDR contains temporal and geographical trends. The highly vetted EIDR database does not significantly differ (p < 0.05) from the summative results of the widely cited database of Jones et al. (2008). EIDR and Jones et al. (2008) have comparable percentages of emergence events associated with zoonotic diseases (63.1%, 60.3%, p < 0.44), vector-borne diseases (22.4%, 22.8%, p < 0.90) and bacteria (50.0%, 54.3%, p < 0.28). Some significant differences (p < 0.10) may exist with emergence events related to viruses (31.7%, 25.4%, p < 0.07). This study and Jones et al. (2008) identified antimicrobial agent use as the most common cause of infectious disease emergence. The large number of resistant microbes and the large number of cases of multiple resistance not identified in EIDR could easily overwhelm other classes of emerging infectious disease events, given the rapid and ongoing discovery of resistant strains in patients. Based on these findings, future research should focus on identifying the drivers of antimicrobial resistance.

EIDR uncovers some of the limitations of using individual emerging infectious disease reports to study trends in disease emergence. Emergence events are often complex and emerging infectious disease case studies have been subjected to varying levels of scrutiny. Elucidating the geographic or temporal origin and drivers of even well-known emerging infectious diseases (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 (Pulliam et al., 2012; Faria et al., 2014;). Even when potential hosts are known, it is often difficult to discern the transmission route for zoonotic disease events. Additionally, substantial labor is required to create reliable emerging infectious disease event databases, with EIDR itself taking several years to complete. Although future studies can use the EIDR database to reduce emerging infectious disease investigation time, maintaining the EIDR database so that it reflects the most current definition of emergence, and contains the most recent infectious disease emergence events is a formidable task. A specific limitation of EIDR is that the database only contains events gathered from published scientific literature. Verifying and validating potential emerging infectious disease events that have been noted in previous studies is sometimes difficult. Extensive effort was made to search for these potential emerging infectious disease events, often with limited success. Lastly, for many emergence 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 study’s finding’s support Funk et al’s (2013) argument that a more quantifiable and objective definition is needed for emerging infectious disease.

The 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 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**

Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. (2014) The early spread and epidemic ignition of HIV-1 in human populations. Science 346(6205): 56–61; DOI:[10.1126/science.1256739](http://dx.doi.org/10.1126/science.1256739) [Online October 2, 2014]

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; DOI:10.1371/journal.pone.0069951 [Online August 14, 2013]

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. Report to Department for International Development, UK Zoonoses Project 4. Available:https://cgspace.cgiar.org/bitstream/handle/10568/21161/ZooMap\_July2012\_final.pdf

Heymann, David L., et al. (2015) Global health security: the wider lessons from the west

African Ebola virus disease epidemic. The Lancet 385(9980):1884-1901; DOI:<http://dx.doi.org/10.1016/S0140-6736(15)60858-3> [Online May 9, 2015]

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. (2008) Global trends in emerging infectious diseases. Nature 451:990–993; DOI: [10.1038/nature06536](http://dx.doi.org/10.1038/nature06536) [Online February 21, 2008]

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

Morens DM, Folkers GK, Fauci AS. (2004) The challenge of emerging and re-emerging infectious diseases. Nature 430:242–910; DOI:1038/nature02759 [Online July 8, 2004]

NCBI. (2015) Home - Taxonomy - NCBI. <http://www.ncbi.nlm.nih.gov/taxonomy>

Pan American Health Organization / World Health Organization (2016) Zika

Epidemiological Update – 21 April 2016. Available:http://www.paho.org/hq/index.php?option=com\_content&view=article&id=11599&Itemid=41691&lang=en [Accessed April 26, 2016]

Pike J, Bogich T, Elwood S, Finnoff DC, Daszak P. (2014) Economic optimization of a global strategy to address the pandemic threat. Proceedings of the National Academy of Sciences of the United States of America111(52):18519-18523; DOI:10.1073/pnas.1412661112 [Online November 17, 2014]

Pulliam JR, Epstein JH, Dushoff J, Rahman SA, Bunning M, Jamaluddin AA, et al. (2012) Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis. J R Soc Interface 9:89–101; DOI:10.1098/rsif.2011.0223 [Online June 1, 2011]

Rosenberg R, Johansson MA, Powers AM, Miller BR. (2013) Search strategy has

influenced the discovery rate of human viruses. Proc Natl Acad Sci USA 110: 13961–13964; DOI:10.1073/pnas.1307243110 [Online June 27, 2013]

Smith, KF., Sax DF., Gaines SD., Guernier V., Guégan J. (2007) Globalization of

Human Infectious Disease. Ecology88(8):1903-1910; DOI:10.1890/06-1052.1 [Online August 1, 2007]

Smolinski MS, Hamburg MA, Lederberg J. (2003) 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.

Taylor L, Latham S, Woolhouse M. (2001) Risk factors for human disease emergence. Philos Trans R Soc B. 356:983–9; DOI:10.1098/rstb.2001.0888 [July 29, 2001]

Weiss RA, McMichael AJ. (2004) Social and environmental risk factors in the emergence of infectious diseases. Nat Med. 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; DOI:10.1098/rstb.2011.0354 [Online October 19, 2012]

Woolhouse ME, Howey R, Gaunt E, Reilly L, Chase-Topping M, Savill N. (2008) Temporal trends in the discovery of human viruses. Proceedings of the Royal Society B: Biological Sciences 275(1647):2111-2115; DOI:10.1098/rspb.2008.0294 [Online September 22, 2008]

World Health Organization (2014) WHO Global Health Estimates Available:

<http://www.who.int/healthinfo/global_burden_disease/en/> [accessed May 20, 2015]

|  |  |  |
| --- | --- | --- |
| **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 |  |
| Emerging infectious disease 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.

|  |  |
| --- | --- |
| **Driver** | **Definition** |
| International travel and commerce | Human travel, goods trade, invasive species from trade or travel |
| Breakdown of public health measures | Issues of vaccine production, vaccine, vaccine enforcement, sanitation, failure to quarantine, toxic chemical exposure, air quality, hazardous waste management |
| Climate and weather | Climate change, deforestation, drought, rainy season, tsunami, hurricane, typhoon, flood, earthquake, stagnant pools, heat wave |
| War and famine | War, conflict, political unrest, malnutrition, famine, substandard living conditions due to conflict, refugee camp conditions, intent to harm. |
| Human susceptibility to infection | Any immune deficiency from an underlying disease, condition, age, or scenario |
| Antimicrobial agent use | The overuse, or misuse of antimicrobial agents |
| Ecosystem changes | Changes to the environment, like expansion of agricultural lands, or clear-cutting that dramatically alters an area's ecosystem |
| Medical industry changes | New vaccination practices, changes in medical protocol, new techniques, new equipment |
| Human behavior | Living conditions, population density, urbanization, migration, food conumption, drug use, human gatherings, daily life routines, recreation |
| Proximity to wildlife | Living near wildlife, suburbanization, being in the wild |
| Agricultural industry changes | Intensification of agriculture practices, use of antibiotics or pesticides, new agricultural practices, new feeds |
| Bushmeat | Consuming, hunting, or trading bushmeat |

Table 2. A list of the drivers and definitions used for EIDR.