**Response to Reviewers’ Comments**

**Reviewer 1:**

Title:

* the title current needs work as you have just included the database

telling the reader nothing about the article. Readers want to be captured by

the title so I would recommend either building on this to include why this topic

is important, a summary of the finding or the key message. I would also delete

the acronym from the title – it is distracting.

*We agree with the reviewer.* The title has been modified to better capture the value of EIDR.

The Emerging Infectious Disease Repository: A novel resource for investigating the emergence of infectious disease.

Keywords:

* I would attempt to limit they keywords to 10 words. This change

should be relatively easy as the terms seem to be overlapping.

*We agree with the reviewer.* We removed “zoonosis” and “infectious disease map”, because the other keywords largely overlap with these.

Abstract: The abstract needs to be reworked I have provided suggestions in

steps of how to reframe it below:

* You need to introduce the topic (yes you have explained the repository

but you need to mention why this is important and what the aims and

objectives of this paper are. I would also highly recommend deleting

the URL link from the abstract).

*We agree with the reviewer.* We have added a sentence at the beginning of the abstract that highlights the value of these disease occurrence databases. We have also reworked the second sentence to include the goal of the study. We have removed the URL link.

“Data associated with emerging infectious disease events have been instrumental in geographic and temporal analyses of emerging infectious disease. The goal of the Emerging Infectious Disease Repository (EIDR) is to create a curated database on the occurrence, characteristics, and drivers of emerging infectious disease events within an interactive web application built to communicate this information.”

* You need to succinctly explain the methods, design, approaches and

sample applied to this research.

*We agree with the reviewer.* We have added the following to the abstract to address this recommendation:

“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). The data contained in the EIDR databases cover everything currently reported by existing emerging infectious disease databases that could be validated through published sources.”

* It seems your results of this paper only identify the limitations and

challenges of the database – to me the most interesting findings of the

paper is the EID events results (why was this not included?).

*We agree with the reviewer.* We have added the following sentences to bring key results into the abstract.

“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 known zoonotic pathogens. Critical gaps were identified. For example, the specific transmission pathway is unknown for nearly 1/3rds of the emergence events involving zoonotic pathogens. Additionally, no drivers were identified in approximately 30% of all events.”

* You need to add a so what section? Saying why this information adds

value to the field and provide recommendation to improve current

limitations.

We have altered the final sentences of the abstract to address these recommendations and assure that the values of the repository to overcome challenges and existing limitations are clearly identified.

“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:

* You paint a picture of the issue well in the opening sentence so the

reader knows this is a critical issue. However, for me the opening

paragraph of an article needs to be sharp and punchy. Currently your

opening paragraph is a list of examples of the issue. I would tighten

these 6 sentences possibly into 4, for example sentence 1-2 could be

merged and shortened as they overlap.

*We agree with the reviewer.* We have combined the first two sentences, eliminated the sentence that quantifies the impact of the Ebola epidemic, eliminated the sentence pertaining to HIV and added a reference to Zika Virus.

“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 widespread, and ongoing, 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.”

* Because you are explaining current issues the Smolinski paper from

2003 is outdated. I think this sentence should be replaced with a

current issue in this space, e.g. Zika virus.

*We agree with the reviewer.* We have removed the sentence concerning HIV and added a reference to the current Zika Virus. See response provided above.

* You need some information on the aims and questions this paper

seeks to respond to as I am left attempting to work out the purpose of

this paper.

*We agree with the reviewer.* We have reworked the second paragraph of the introduction, focusing on making clear the purpose of this project.

“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 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, highly curated, database of these events that now provides the research community with a platform to track the sources of all relevant data.”

* I think the EID Event collection that currently sits in the methods

probably works better as background information but needs to be

refined so a person not from this discipline can pick up this paper and understand what this database is and why these EID papers were the focus for EcoHealth Alliance specialists.

*We agree with the reviewer.* We retitled the section, and added sentences to clarify to readers that this paragraph is part of the methods. It provides the definition of emerging infectious disease, and the event inclusion criteria used to conduct the study.

***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.”

Methods:

* Shouldn’t the sample be explained in the methods section.

*We agree with the reviewer.* We removed that sentence, because the sample is explained in the methods section

* Page 4 when explaining the EID drivers is there a way of making this

information easier to follow (possibly in a table or figure) or possibly

combining them with the variable table?

*We agree with the reviewer.* We have added a table (Table 2) to more succinctly display the drivers used in EIDR. We have also replaced the list of drivers with the following sentence to clarify the origins of each driver:

“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 drivers listed in Smolinski et al. (2003) and Lederberg et al. (1992). A complete list of drivers and their definitions is provided in Table 2.”

* It would be good to know the exclusion and inclusion criteria for this

project and possibly include it in a figure of how this was done.

*We agree with the reviewer.* We have added the following sentence to describe our inclusion criteria.

“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 case.”

* Also how long was the review process undertaken of these papers. It

would be helpful for the reader to understand how these paper was

reviewed.

*We agree with the reviewer.* We have added the following sentences to describe the review process in greater detail.

“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.”

* The last 4 paragraphs of the methods left me asking a number of

questions about this research. I am a social scientist by training so this

is not my field but when I have undertaken systematic reviews I have

had to clearly explain the steps for reviewing papers and how they

were done at the same time explaining the theories and methodologies

applied. I am left not knowing what these are for this paper – I guess

what I am saying is you need to explain the process much more clearly

for people not from this discipline who may not know why these

approaches were selected. What I look for in a methods is that anyone could pick up your paper and replicate it – currently I could not do this.

*We agree with the reviewer.* We used this suggestion to guide us as we added additional sentences to the paper, including the two additions to the methods that are mentioned directly above this comment and additions to the introduction. We also reworked a sentence in the methods to read:

“Potential emergence events were collected from a review of meta-analyses on disease emergence (2,7,8,13) and via an exhaustive literature review (e.g. Web of Science, PubMed, ProMED-mail, WHO Weekly Epidemiological Record, CDC Morbidity and Mortality Weekly Report). The data contained in the EIDR databases cover everything currently reported by existing emerging infectious disease databases that could be validated through published sources.”

Results:

* EID events is your most critical results in this paper to me. It would be

great if you could elaborate on what these findings mean.

In response to this comment and other related comments from this reviewer we revised many parts of the manuscript to emphasize the importance of this work to the field. Changes in the abstract, introduction and discussion were made to address this comment.

* EIDR Web Application section does not read like results and more like

background information.

To address this concern by the reviewer, we have moved this section to the end of the methods since it is part of the approach.

* Statistical testing:

a. Firstly, the Chi-square test has been mentioned in the methods

and doesn’t not need to be mentioned again.

*We agree with the reviewer.* We have removed the sentence, “A Chi-square test was used to compare the Jones et al. database to the EIDR database”, from the results section.

b. The authors need to explain what this information means as

currently this data looks tacked on.

The intent of the Chi Square analysis was to examine if the EIDR results argue against the use of Jones et al. results. The Jones et al. paper has been widely cited (in excess of 2000 times), so if the results from EIDR were markedly different it would be important to bring this to the attention of researchers and practitioners. Thus, it was important to note that the results are comparable between the highly vetted EIDR and the Jones et al. paper.

We have moved the section regarding the comparisons to Jones et al., into the discussion section and added the following additional sentences to provide context:

“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 differences (p< 0.10) may occur with regard to emergence events related to viruses (31.7%, 25.4%, p < 0.07).”

Discussion:

* The authors need to consider extending the discussion beyond

summarizing results. You need to provide more than the 2 last sentences

reviewing how this research contributes to future practice and provide

recommendations.

In response to the reviewer’s comments, we added new information on the importance of the work to the introduction and also have added the following sentence to the end of the 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 accessible platform.”

* I am also left after reading this paper knowing little more than when I started

reading this paper. With this in mind you need to emphasis why this paper is

critical to the research field.

In response to this comment and other related comments from this reviewer we revised many parts of the manuscript to emphasize the importance of this work to the field. Changes in the abstract, introduction and discussion were made to address this comment.

General information

* Please check the references are consistent and abide by EcoHealth

guidelines.

*We agree with the reviewer.* We have corrected the references to make sure they are consistent with EcoHealth guidelines.

* I find the overuse of acronyms distracting and should be reduced.

*We agree with the reviewer.* We have removed all acronyms besides EIDR from the manuscript.

**Reviewer 2:**

Comments to the Author

This manuscript presents a new resource, the Emerging Infectious Disease Repository, which represents the culmination of many years of effort directed at compiling existing EID event datasets and digging deeper to add data on various additional parameters of interest. I think the resource represents a profound amount of work, but unfortunately the written presentation here leaves much to be desired. I am left with the impression that the EIDR was a monumental effort, and the paper, unfortunately, gives the impression that the authors have lost steam.

The most notable writing improvements fall into three categories:

a) explaining what is meant by key, broad statements

b) making discussions more precise

c) correcting numerous grammatical errors. I list several here but eventually stopped.

I’ve also requested more detail about particular definitions. At the risk of seeming pedantic, these details are critical for inferences made following any analyses relying on these data. These inferences will be limited by the quality of the data, which is hard to judge if definitions are vague.

Abstract:

* The sentence beginning “The composition of the EIDR database builds existing EID databases…” is unclear. Do authors mean that the data contained in the EIDR databases cover everything currently reported by existing EID databases (such as those compiled by Jones, Smith, and others) while also revealing the limitations of those databases for analysis and inference?

*We agree with the reviewer.* We have expanded and sharpened this sentence to capture the suggestion:

“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). The data contained in the EIDR databases cover everything currently reported by existing emerging infectious disease databases that could be validated through published sources.”

* The last sentence is disjointed and lacks conclusion – do the authors mean to imply that this new EIDR achieves these elements (precisely defining and discerning key aspects of EIDs) that are lacking in existing databases? If so, a statement to this effect should be made explicitly.

*We agree with the reviewer.* We have reworked the final sentence based on the reviewers’ suggestion:

“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:

* the Globalization of infectious diseases paper by Kate Smith should be cited somewhere. Line 41 (“Globalization has produced…”) is a natural place to do this.

*We agree with the reviewer.* Smith et al., 2007 is now cited in the following sentence:

“Globalization has produced an efficient mechanism to spatially distribute emerging infectious diseases (Smith et al., 2007).”

* The 2015 paper by Heymann et al. in the Lancet should be cited: Global health security: the wider lessons from the west African Ebola virus disease epidemic. Line 51.

*We agree with the reviewer.* Heymann et al., 2015 is now cited in the following sentence:

“The vulnerability of global health security has recently been revealed through the widespread, and ongoing, 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).”

* If available, please update the numbers cited for the damages and deaths from the Ebola outbreak with more recent WHO 2015 figures.

*We agree with the reviewer.* We removed the damages and deaths information in response to Reviewer 1’s suggestion that we streamline this section of the introduction.

* MAJOR: “Despite the apparent face validity of combining multiple independent EID events…for spatial prediction, this method is limited by the complexity surrounding historic emergence events and difficulties in finding and validating underlying data”. This statement needs to be unpacked and fleshed out to give proper context for the EIDR. What value does this new database really add beyond those that have already been reported and analyzed? Please provide at least 1-2 examples of specific predictions that have been made that are limited by complexity. What complexity limits the predictions? Preferably, the authors will discuss 1-2 limitations that the new EIDR has successfully removed or reduced.

In response to this comment, we have reworked this particular sentence –deleting the phrase mentioned by the reviewer.

“This method is limited by the complexity surrounding historic emergence events and difficulties in finding and validating underlying data (Funk et al., 2013).”

We added explicit information on what limitations the EIDR overcomes – specifically within the methods we added the following sentence:

“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.”

Also, the abstract and discussion now include examples of the types of complexities associated with historic events, including that the specific transmission pathway is unknown for many zoonotic pathogen emergence events and the lack of identified drivers for many events.

* Related to the above: how exactly does the EIDR find and validate underlying data better than existing databases? Presumably the authors will answer this later, but a preface statement here would improve the Intro.

In response to this comment we clarified the methods used to find and validate the data. We made a number of changes to address this comment. Below we provide two specific examples of changes:

From paragraph 2 of the methods: “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).”

From last sentence of the methods section entitled Data Collection Review: “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.”

* The last sentence has some grammar issues.

*We agree with the reviewer.* The sentence has been split into two sentences that provide a clear explanation of the goal and value of the project.

“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, highly curated, database of these events that now provides the research community with a transparent platform to track the sources of all relevant data.”

Methods

* Were data only collected for events ranging between 1940-2013? If so, are there plans to keep data updated or adding events that are not included in this version of the database?

As noted in the paper, EIDR can be updated to include additional events, or modify events currently in EIDR. At the moment, the authors do not have plans to update the resource. EIDR only contains events spanning 1940-2013.

In determining the lower temporal cut-off, we followed Jones et al. (2008). In their Nature article they state:

“We chose the 1940 cut-off based on the Institute of Medicine’s examples of a currently or very recently emerging disease, all of which had their likely temporal origins within this time period.”

To make our decision clear we added a phrase to the following sentence:

“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 show that currently or very recently emerging infectious diseases have occurred since 1940 (Smolinski et al., 2003).”

To make it clear that EIDR contains no events that occurred after 2013 we added the following sentence.

“Potential emergence events were not evaluated if they occurred after 2013”.

* Perhaps related to this point: I notice that the public dataset on human infectious disease events reported by Smith et al. 2014 (PRS-Interface) is not mentioned or included in this paper. It will be important to set the boundaries on what data are included so that readers will not be left wondering why major sources (like the Smith resource) are not included here. (data available here: <http://ramachandran-data.brown.edu/datarepo/request.php?request=explorePublicStudy&StudyID=6&instit=BROWN)>

*We agree with the reviewer.* The data in EIDR pertain to emergence events only. Therefore, only meta-analyses specifically concerned with disease emergence were included. To make this clear, we have added the following sentence:

“Meta-analyses not specific to disease emergence were not included.”

Smith et al’s. (2014) provides a powerful database of over 12000 infectious disease outbreaks, but does not contain data specifically pertaining to emergence events, so was not included.

* There are no units assigned to any of the data fields? Some of these are self-explanatory (e.g., sexually transmitted is presumably binary?) but others are not clear – for example, proximity to wildlife – there are three categories (near wildlife, in the wild, suburban), so is this variable categorical or based on some distance threshold? On the first pass, the “Zoonosis” field seems pretty straightforward (yes/no), but if the “Host Use” field is “Unknown”, will “Zoonosis” always show up as “No”, or is “N/A” an option for this field?

The Supplemental Information contains all the information on how data were coded. Most information is binary. In some cases, variables are coded using specified values. “Proximity to wildlife”, is one of 12 potential values for the variable “driver”. “Near wildlife”, “in the wild”, and “suburban” are all listed within the description of the specific driver value, “proximity to wildlife”. They are not coded individually, so that level of information is not captured. Collection of data for many variables, (e.g. "Driver”, “Type of Emergence”) was contingent on qualitative information from sources (e.g. descriptions of wildlife activity near the event location, or direct involvement of wildlife with event transmission). Further, “N/A” is an option for all data fields.

To make clear that many variables are qualitative the following sentence was added:

“Collection of data for many variables, (e.g. "Driver”, “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”)”

* Page 5 states that if no data were found on a particular variable, the absence of data was capture. How exactly is data absence recorded in EIDR?

*We agree with the reviewer.* We have reworked the existing sentence pertaining to capturing the absence of data and added an additional sentence describing how this absence is displayed to users.

“If no information could be found on a particular variable the absence of data was coded for during data collection. All variables missing data for a given event were displayed within an event’s “Event Page” of the EIDR website.”

* Please be more precise about how virulence is measured. I see a vague definition but no actual metrics of virulence stated in the main text or in the Supplementary materials. What metrics enable distinction between a highly virulent and a non-virulent pathogen?

*We agree with the reviewer.* “Increased Virulence” is a value for the qualitative variable, “Type of Emergence”. In order to be categorized as an emergence event attributed to “Increased Virulence” factors that indicated increased virulence had to be identified in the sources concerning the event (e.g. novel disease associated with the pathogen)

The following sentence was added to clarify that “Type of Emergence” was a qualitative variable:

“Collection of data for many variables, (e.g. "Driver”, “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”)”

* Please be more precise about “no clinical significance”. Is this a frequency metric (the number of people known to have developed disease: none? One? Less than 100?)?

*We agree with the reviewer.* The term has been removed and its meaning incorporated into the exclusion criteria provided by the following sentence:

“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."

* Page 3 line 15, “based on…”

*We agree with the reviewer.* We have reworked the line and expanded the description of what EIDR “Drivers” are based on.

“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 drivers listed in Smolinski et al. (2003) and Lederberg et al. (1992). A complete list of drivers and their definitions is provided in Table 2.”

* Why would most diseases emerge after 1940? (lines 27-29 page 3)

*We agree with the reviewer.* We have made it clear that this conclusion derives from Smolinski et al. (2003).

“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).”

* Where are the methods for the systematic literature review?

*We agree with the reviewer.* We have added a sentence including the resources most frequently used in the literature review.

“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).”

* The rationale provided for the chi-squared statistics seem contradictory. The authors state that data on several critical variables and EID drivers were collected from the Lederberg and Jones papers. Also the majority of the events in EIDR come from the Jones paper (with the exception of 39 that were excluded). Authors then state that the datasets (EIDR and Jones) are predominantly independent of each other based on differing criteria for event inclusion. Please reconcile these statements. And what exactly where the chi-squared tests used to compare?

To elucidate the purpose of the chi-squared statistics the following sentence has been added to the conclusion:

“The composition of the EIDR database largely replicates the findings of Jones et al. (2008), but does not contain temporal or 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 differences (p< 0.10) may occur with regard to emergence events related to viruses (31.7%, 25.4%, p < 0.07).”

As pointed out of the paper, regardless of where the information originally came from, no information made it into EIDR unless it was collected independently by the author. EIDR required a stringent set of vetting criteria in order to be included. Therefore, although the databases contain many of the same events, the data extracted from these events was collected using different methods and can be analyzed as independent samples.

The second sentence of the Chi-Squared section of the methods was expanded to better explain the appropriateness of the chi-square analysis.

“Data collection methods and criteria for emergence event inclusion differed between the studies. The data within each dataset are, therefore, independent, despite deriving from some of the same events. This makes the datasets are appropriate for the chi-square test.”

Results:

* Are web comments reviewed before they post? If so, how are they reviewed? It would be useful to know how transparent this process is.

Currently, web comments are not reviewed before they are posted, however, EIDR has the capacity to limit who can make comments, by controlling who can sign in to EIDR. The author’s do not have plans to actively maintain the site at this time, but this capability is built into to the platform.

* Statistical testing results were confusing – is the first percentage reported for EIDR (63.1%) or is it the second number (60.3%)? Same goes for the remaining sets of percentages. Also, there are no degrees of freedom, sample sizes, or chi-squared statistics reported anywhere?

*We agree with the reviewer.* The following sentences were added to the conclusion to clarify the statistical findings.

“The composition of the EIDR database largely replicates the findings of Jones et al. (2008), but does not contain temporal or 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 differences (p< 0.10) may occur with regard to emergence events related to viruses (31.7%, 25.4%, p < 0.07).”

Discussion:

* Several typos/grammar errors in the second sentence.

*We agree with the reviewer.* The sentence has been reworked and is provided below:

“The composition of the EIDR database largely replicates the findings of Jones et al. (2008), but does not contain temporal or geographical trends.”

* Remove the percentages from the Discussion – redundant with results section

*We agree with the reviewer.* The chi-square analysis has been removed from the results and is now only present in the Discussion. The sentence has been modified and expanded to several sentences to remove confusion.

“The composition of the EIDR database largely replicates the findings of Jones et al. (2008), but does not contain temporal or 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 differences (p< 0.10) may occur with regard to emergence events related to viruses (31.7%, 25.4%, p < 0.07).”

* First sentence, second paragraph, typo: “of using”

*We agree with the reviewer.* The second “of using” has been removed.

* Page 8, line 13: “but it is often difficult”

*We agree with the reviewer.* The sentence has been reworked to clarify the message.

“Even when potential hosts are known, it is often difficult to discern the transmission route for zoonotic disease events.”

* There is something wrong with the sentence in line 48, page 8; authors meaning is unclear.

*We agree with the reviewer.* The sentence has been removed.

* Page 9, “study’s” not “studies”; “findings” not “finding’s”.

*We agree with the reviewer.* “studies” has been corrected to “study’s”.

General comments:

* The text of Figures 1-2 was too blurry to be deciphered or reviewed. I tried both the HTML and the PDF versions with no luck.