%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Please note that whilst this template provides a

% preview of the typeset manuscript for submission, it

% will not necessarily be the final publication layout.

%

% letterpaper/a4paper: US/UK paper size toggle

% num-refs/alpha-refs: numeric/author-year citation and bibliography toggle

\documentclass[a4paper,alpha-refs,gigabyte]{oup-contemporary}

%%% Journal toggle; only specific options recognised.

%%% (Only "gigabyte", "gigascience" and "general" are implemented now. Support for other journals is planned.)

\journal{gigabyte}

\usepackage{graphicx}

\usepackage{siunitx}

\usepackage{tabularray}

%%% Flushend: You can add this package to automatically balance the final page, but if things go awry (e.g. section contents appearing out-of-order or entire blocks or paragraphs are coloured), remove it!

% \usepackage{flushend}

\title{Data from: Rodent trapping studies as an overlooked information source for understanding endemic and novel zoonotic spillover}

%%% Use the \authfn to add symbols for additional footnotes, if any. 1 is reserved for correspondence emails; then continuing with 2 etc for contributions.

\author[1,2,3,\authfn{1}]{David Simons}

\author[3,4]{Lauren A. Attfield}

\author[3]{Kate E. Jones}

\author[3,5]{Deborah Watson-Jones}

\author[1]{Richard Kock}

\affil[1]{The Royal Veterinary College, London, United Kingdom}

\affil[2]{University College London, London, United Kingdom}

\affil[3]{Imperial College London, London, United Kingdom}

\affil[4]{The London School of Hygiene and Tropical Medicine, London, United Kingdom}

\affil[5]{Mwanza Intervention Trials Unit, National Institute for Medical Research, Mwanza, Tanzania}

\authnote{\authfn{1}dsimons19@rvc.ac.uk}

%%% Paper category

\papercat{Data Release}

%%% "Short" author for running page header

\runningauthor{Simons et al.}

%%% Should only be set by an editor

\jvolume{00}

\jnumber{0}

\jyear{2020}

\begin{document}

\begin{frontmatter}

\maketitle

\begin{abstract}

Rodents are a diverse, globally distributed and ecologically important order of mammals that are known and predicted hosts of zoonotic pathogens. The sampling of rodents and their pathogens are taxonomically and spatially biased, limiting inference of the hazard of spillover of zoonotic pathogens into human populations from these rodent vectors. Data on the distribution and occurrence of rodent hosts are typically derived from consolidated databases (e.g., IUCN, GBIF) which suffer from these biases. Here, we synthesise data from 127 rodent trapping studies, published between 1964–2022 conducted in West Africa to provide an additional source of information that can supplement consolidated databases to characterise the range and occurrence of rodent species. We combine these occurrence data with results from pathogen screening to produce a dataset containing detection/non-detection data for 65,628 individual small mammals identified to species level from at least 1,611 trap sites across 14 countries of West Africa in addition to 32 microorganisms identified to species and genus level that are known or potential pathogens. The produced dataset is expected to mitigate some of the spatial and taxonomic biases in current databases to improve analyses of rodent-borne zoonotic pathogen spillover hazard across West Africa.

\end{abstract}

\begin{classifications}

Rodents; Zoonosis; Vector-Borne Diseases; West Africa

\end{classifications}

\end{frontmatter}

\section{Data Description}\label{sec:background}

\subsection{Context}

Rodents are a diverse, globally distributed and ecologically important order of mammals. Along with bats, these two orders are proposed to contain the greatest number of host species of known and predicted of zoonotic pathogens. Of 2,220 extant rodent species, 244 (10.7\%) are described as hosts of 85 zoonoses \citep{han\_2015}. Rodent borne endemic zoonoses are a significant public health threat across much of West Africa and include bacterial, viral and protozoan pathogens. Important endemic rodent-borne zoonoses in West Africa include Lassa fever (caused by \textit{Lassa mammarenavirus}), Leptospirosis (caused by \textit{Leptosira sp.}) and Toxoplasmosis (caused by \textit{Toxoplasma gondii}) \citep{garry\_2022, allan\_2015, brouat\_2018}. It is likely that many other potential zoonoses are circulating in rodent populations in this region that have not been described or identified as causing human infections \citep{suu-ire\_2022}. Described zoonotic pathogens are generally associated with multiple rodent species and therefore understanding the structure of rodent communities, their spatial distribution and their pathogens are vital to understand the hazard of endemic zoonotic disease spillover and novel zoonotic pathogen emergence \citep{han\_2016}.

Studies assessing the risk of outbreaks of endemic zoonoses and novel pathogen emergence often use consolidated datasets such as the Global Biodiversity Information Facility (GBIF) and International Union for Conservation of Nature (IUCN) Redlist to model host occurrence \citep{basinski\_2021, mylne\_2015, pigott\_2016}. Despite the importance of understanding the true distribution of rodent hosts and their pathogens curated biodiversity databases such as the are affected by taxonomic and geographical sampling biases \citep{boakes\_2010, bowler\_2022}. These biases can subsequently limit inference from produced species distribution models that are used to quantify the hazard of zoonotic disease spillover into human populations and guide public health interventions \citep{beck\_2014}. Rodent trapping studies are also taxonomically and spatially biased, despite this, combining data from these studies with data from GBIF and IUCN has the potential to increase the sampled area for commonly occurring species by up to 160\% and mitigate some of the effects of these biases \citep{simons\_2023}.

The current dataset, a synthesis of 127 rodent trapping studies conducted across West Africa can aid the development of models based on rodent reservoir occurrence to estimate the potential for pathogen spillover into human populations by providing additional locations of sampling. For example, a recent article developed a model of the risk of Lassa fever spillover based on both \textit{Mastomys natalensis} occurrence and pathogen prevalence in West Africa \citep{basinski\_2021}. This study included 167 locations of \textit{M. natalensis} detections which could be potentially expanded to 337 detections and 320 non-detections using the current dataset.

\section{Methods}

\subsection{Search strategy}

This dataset contains information on small mammal detections and non-detections obtained from rodent trapping studies conducted in West Africa between 1964 and 2022. Data have been extracted from published articles, biodiversity surveys and impact assessments. Studies were identified through a search conducted in Ovid MEDLINE, Web of Science (Core collection and Zoological Record), JSTOR, BioOne, African Journals Online, Global Health and the pre-print servers, BioRxiv and EcoEvoRxiv using the following terms as exploded keywords:

\begin{itemize}

\item[1] Rodent OR Rodent trap\* AND

\item[2] West Africa

\end{itemize}

Similar searches were conducted in additional resources, including the UN Official Documents System, Open Grey, AGRIS FAO and Google Scholar. Searches were run on 2022-05-01.

We included studies for further analysis if they met all of the following inclusion criteria;

\begin{itemize}

\item[1] Reported findings from trapping studies where the target was a small mammal

\item[2] Described the type of trap used or the length of trapping activity or the location of the trapping activity

\item[3] Included trapping activity from at least one West African country

\item[4] Recorded the genus or species of trapped individuals

\item[5] Were published in a peer-reviewed journal or as a pre-print on a digital platform or as a report by a credible organisation

\end{itemize}

We excluded studies if they met any of the following exclusion criteria:

\begin{itemize}

\item[1] Reported data that were duplicated from a previously included study

\item[2] No full text available

\item[3] Not available in English

\end{itemize}

One author (DS) screened titles, abstracts and full texts against the inclusion and exclusion criteria. At each stage; title screening, abstract screening and full text review, a random subset (10\%) was reviewed by a second author (LAA). Table \ref{tab:studies} contains the year of publication, name of the first author, title of the study, publication and unique identifier of the included study.

Table 1:

\subsection{Data extraction}

Data were extracted from eligible studies using a standardised tool that was piloted on five randomly selected studies. Table \ref{tab:extraction} contains the variable names and descriptors that were abstracted into three sheets. The first sheet, “Study data”, contained information on the included study, the purpose of the study, methodology of rodent sampling and species identification. The second sheet, “Rodent data”, contained information on the number of individuals of each species detected at a trapping location, alongside geographic coordinates of the sampling location and habitat type. Data for this section were expanded by adding non-detections if the rodent species was detected at other sampling sites within the study. Finally, the third sheet, “Pathogen data”, contained information on the testing of the individual rodent species for known and suspected zoonotic pathogens. Unprocessed data is archived in a Zenodo repository within the \verb|data\_raw| folder \citep{simons\_2022}.

Table 2:

\subsection{Data validation and quality control}

Species identification was assumed to be accurate in included studies, for studies reporting genus level or multiple possible species names for trapped individuals’ data were extracted as presented in the study. Species names were mapped to GBIF taxonomy to resolve changes in taxonomic classification using the `taxize` package (version 0.9.98) in the R statistical programming language (version 4.1.2) \citep{chamberlain\_2020, R\_core\_2021}. Geographic locations of trapping studies were extracted using GPS locations for the most precise location presented. Missing locations were found using the National Geospatial-Intelligence Agency GEOnet Names Server based on placenames and maps presented in the study \citep{ngia\_2021}. All locations were converted to decimal degrees in the EPSG:4326 coordinate reference system.

For included studies with available data we extracted information on all microorganisms and known zoonotic pathogens tested and the method used (e.g., molecular or serological diagnosis). Where assays were able to identify the microorganism to species level this was recorded, for non-specific assays higher order attribution was used (e.g., to family level). For studies reporting summary results all testing data were extracted, this may introduce double counting of individual rodents, for example, if a single rodent was tested using both molecular and serological assays. Where studies reported indeterminate results, these were also recorded.

\subsection{Data processing and exploration}

R code to process the raw extracted data into the Darwin Core format, for rodent occurrence data and associated pathogen detection data along with metadata has been archived as a Zenodo repository \citep{ref\_zenodo}.

An RShiny web application has been produced to visualise the data contained in this release. The web based application is available at \href{https://diddrog11.shinyapps.io/scoping\_review\_app/}{this webpage}. This supports exploration of the location of sampling sites for both rodents and their pathogens within included studies and the sampling effort reported by the study.

\subsection{Reuse potential}

This dataset of harmonised rodent species detections obtained from rodent trapping surveys conducted across West Africa will contribute to understanding rodent biodiversity across the region. It is envisaged that this dataset will be of particular interest to researchers investigating the risk of rodent borne zoonotic pathogen outbreaks and emergence in this region and beyond. This data will expand the geographical coverage of occurrence data within GBIF for most of the rodent species detected in the included rodent trapping studies with additional data on non-detections of these species across the region. Where possible, dates of rodent sampling have been included which may be of benefit to researchers investigating how occurrence patterns of rodent species may vary over time, which will be important to understanding changes in the context of climate, landuse and population changes.

\subsection{Data availability}

As part of this series of Data Release articles this dataset is available under a CC0 license from GBIF.

\section{Declarations}

\subsection{List of abbreviations}

\begin{itemize}

\item GBIF

\item IUCN

\end{itemize}

\subsection{Competing Interests}

The authors declare that they have no competing interests.

\subsection{Funding}

All sources of funding for the research reported should be declared. The role of the funding body in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript should be declared. Please use \href{http://www.crossref.org/fundingdata/}{FundRef} to report funding sources and include the award/grant number, and the name of the Principal Investigator of the grant.

\subsection{Author's Contributions}

The individual contributions of authors to the manuscript should be specified in this section. Guidance and criteria for authorship can be found in our \href{https://doi.org/10.1093/gigascience/giy122}{editorial}. We would recommend you follow some kind of standardised taxonomy like the \href{http://docs.casrai.org/CRediT}{CASRAI CRediT} (Contributor Roles Taxonomy).

\section{Acknowledgements}

Please acknowledge anyone who contributed towards the article who does not meet the criteria for authorship including anyone who provided professional writing services or materials.

Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements section. If you do not have anyone to acknowledge, please write ``Not applicable'' in this section.

See our \href{https://doi.org/10.1093/gigascience/giy122}{editorial} for a more explanation of acknowledgements and authorship criteria.

%% Specify your .bib file name here, without the extension

\bibliography{paper-refs}

\end{document}