



# Institute for Health Metrics and Evaluation

## Data Release Information Sheet

### ***Data Summary***

Dataset name: Rift Valley Fever Environmental Suitability and Spillover Potential Geospatial Estimates

Date of release: March 16, 2020

Summary:

This dataset includes predictions for the environmental suitability of Rift Valley Fever (RVF) transmission at the monthly level, as well as calculations of spillover potential, which combines suitability predictions with human and livestock population data. It also includes occurrence data extracted from a literature review combined with that downloaded in October 2018 from the Food and Agriculture Organization of the United Nations' (FAO) EMPRES-i database of RVF occurrences in mammals.

The dataset includes the following:

- GeoTIFF raster files for pixel-level mean environmental suitability predictions for each of the 12 calendar months and average months of suitability per year for 1995-2016
- CSV files of each administrative level 2 units' average spillover quintile for each month and average months per year in the top quintile of spillover values
- Extracted occurrence data
- Code files and custom polygons used to generate the estimates

Relevant publications and visualizations:

- Hardcastle AN, Osborne JCP, Ramshaw RE, Hulland EN, Morgan JD, Miller-Petrie MK, et al, Informing Rift Valley Fever preparedness by mapping seasonally varying environmental suitability. *Forthcoming*.

### **Acknowledgements**

Contributing organizations:

- Institute for Health Metrics and Evaluation (IHME)

Funders:

- Bill and Melinda Gates Foundation (BMGF)

Suggested Citation:

Institute for Health Metrics and Evaluation (IHME). Rift Valley Fever Environmental Suitability and Spillover Potential Geospatial Estimates. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2020.

## ***Data Files Information***

### **CSV files of monthly mean spillover quintiles**

Stored in files named IHME\_RVF\_AVG\_QTILE\_<MONTH>\_Y2020M02D28.CSV

(Example: AVG\_QTILE\_01 gives values for January, AVG\_QTILE\_02 gives values February, etc.)

- **Measure:** Average quintile of spillover values across all non-zero spillover values from each administrative unit, month, and year
- **Level of aggregation:** Admin2, corresponding to second administrative level areas as defined in the 2018 Database of Global Administrative Areas (GADM) shapefiles

Variable	Variable Label	Variable Definition
ADM0_CODE	Admin 0 GADM code	Zero level administrative unit unique identifying code in GADM 2018 shapefile
ADM0_NAME	Admin 0 Name	Zero level administrative unit name
ADM1_CODE	Admin 1 GADM code	First level administrative unit unique identifying code in GADM 2018 shapefile
ADM1_NAME	Admin 1 Name	First level administrative unit name
ADM2_CODE	Admin 2 GADM code	Second level administrative unit unique identifying code in GADM 2018 shapefile
ADM2_NAME	Admin 2 Name	Second level administrative unit name
avg	Average spillover quintile	Average spillover quintile (1-5) compared to all non-zero values for second administrative units in each month and year

### **CSV file of average months per year in top spillover quintile**

Stored in file named IHME\_RVF\_MONTHS\_PER\_YEAR\_SRY\_Y2020M02D28.CSV

- **Measure:** On average, for each year from 1995-2016, the number of months that each district was ranked in the top quintile of all non-zero spillover values across all second administrative units, months, and years
- **Level of aggregation:** Admin2, corresponding to second administrative level areas as defined in the 2018 Database of Global Administrative Areas (GADM) shapefiles

Variable	Variable Label	Variable Definition
ADM0_CODE	Admin 0 GADM code	Zero level administrative unit unique identifying code in GADM 2018 shapefile

Variable	Variable Label	Variable Definition
ADM0_NAME	Admin 0 Name	Zero level administrative unit name
ADM1_CODE	Admin 1 GADM code	First level administrative unit unique identifying code in GADM 2018 shapefile
ADM1_NAME	Admin 1 Name	First level administrative unit name
ADM2_CODE	Admin 2 GADM code	Second level administrative unit unique identifying code in GADM 2018 shapefile
ADM2_NAME	Admin 2 Name	Second level administrative unit name
months_per_year	Months per year in top spillover quintile	On average, for each year from 1995-2016, the number of months that each district was ranked in the top quintile of all non-zero spillover values across all second administrative units, months, and years

**GeoTIFF raster files for pixel-level estimates of environmental suitability**

Stored in files named: IHME\_RVF\_MEAN\_<MONTH>\_Y2020M02D28.TIF

(Example: MEAN\_01 gives values for January, MEAN\_02 gives values for February, etc.)

- **Indicator:** For a given month, the mean environmental suitability prediction across years 1995-2016
- **Measure:** Environmental Suitability (0-1)
- **Stat:** Mean values of the modelled relationships projected onto each pixel averaged across bootstrap model results and over all years for a given month.
- **Year:** Representative of 1995-2016

Note that rasters mask (i.e., have NA values) for lakes.

**GeoTIFF raster file for pixel-level estimates of average suitable months per year**

Stored in file named: IHME\_RVF\_SUITABLE\_MONTHS\_YEAR\_Y2020M02D28.TIF

- **Indicator:** On average, the number of months per year classified as suitable for RVF transmission based on an optimized threshold for that month
- **Measure:** Months per year of suitability (0-12)
- **Stat:** Mean values of months of suitability in a given pixel across years 1995-2016, based on binary classification maps, which are based on modelled relationships projected onto each pixel averaged across bootstrap model results and over all years and months
- **Year:** Representative of 1995-2016

Note that rasters mask (i.e., have NA values) for lakes.

**Data Input Sources – Occurrence Data**

Stored in file named: IHME\_RVF\_OCCURRENCE\_Y2020M02D28.CSV

This CSV file contains extracted RVF occurrence data and relevant metadata about the input sources as suggested in the [Guidelines for Accurate and Transparent Health Estimates Reporting \(GATHER\)](#), a statement that promotes best practices in reporting health estimates.

<b>Variable</b>	<b>Variable Label</b>	<b>Variable Definition</b>
nid	NID	A unique identifier in the Global Health Data Exchange (GHDx) assigned to each publication that was extracted
title	Title	Title of the publication
author	Author	Article's author(s)
doi	DOI	Article's DOI
abstract	Abstract	Article's abstract, if available
source_title	Source Title (Journal)	Journal in which the article was published
year	Year	Article's publication year
source	Source (Database)	Database where article was found (PubMed, Web of Science, or Scopus)
pmid_if_applicable	PMID (if applicable)	PMID if the article is from PubMed
full_text_link_if_included	Full text link (if included)	Link to the full text, if available
file_id	File ID	Reference to pdf in format FirstAuthor_Year (e.g., Smith_2017)
occ_id	Occurrence ID	Unique identifier assigned to each occurrence of RVF. A single pdf may represent more than one occurrence. Each row will have its own occ_id, starting at 1 and numbered consecutively.
organism_type	Organism type	What type of organism tested positive for RVF (human, mammal, or vector)
organism_specific	Organism specific	Specifies the exact organism that tested positive for RVF.
pathogen	Pathogen	Name the pathogen identified
pathogen_note	Pathogen note	Miscellaneous notes regarding pathogen

Variable	Variable Label	Variable Definition
patient_type	Patient type	<p><u>Unspecified</u>: Cases that lacked sufficient epidemiological evidence to classify them as any other status (e.g. serosurvey studies)</p> <p><u>NA</u>: Non-applicable field; case was not a patient (e.g. mammal)</p> <p><u>Import</u>: Cases that were brought into a non-endemic country after transmission occurred elsewhere</p>
transmission_route	Transmission route	<u>Unspecified</u> : Lacked sufficient epidemiological evidence to classify as any other status
clinical	Clinical	Describes whether the RVF case was symptomatic, asymptomatic, or suspected
diagnostic	Diagnostic	Describes the class of diagnostic method that was used. PCR, serology, or reported
diagnostic_note	Diagnostic note	More detailed information related to the specific test used (e.g., rk39, IgG, or IgM serology)
serosurvey	Serosurvey	<p>Describes the context if serological testing was used</p> <p><u>Diagnostic</u>: testing of symptomatic patients</p> <p><u>Exploratory</u>: historic exposure determined among healthy asymptomatic individuals</p>
country	Country	ISO3 code for country in which the case occurred
origin	Origin	Open-ended field to provide more details on the specific in-country location of the RVF case
problem_geography	Problem Geography	This field was utilized if the RVF case was reported in a location that could cause uncertainty when determining exact geographic occurrence (e.g., hospital, abattoir).
lat	Latitude	Latitude measured in decimal degrees
long	Longitude	Longitude measured in decimal degrees

<b>Variable</b>	<b>Variable Label</b>	<b>Variable Definition</b>
latlong_source	Lat/long source	The source from which latitude and longitude were derived  All data from the EMPRES-i database have a value of “empres” in this column.
loc_confidence	Location confidence	States the level of confidence that researchers had when assigning a geographic location to the RVF case (good or bad). An answer of ‘good’ meant the article stated clearly that the case occurred in a specific geographic location and no assumptions were required on part of the researcher. An answer of ‘bad’ meant the article did not clearly state the specific geographic location of the MERS-CoV case, but the researcher was able to infer the location of occurrence. The field SITE_NOTES was utilized to detail the logic behind researchers’ decisions when inference was required.
shape_type	Shape type	The geographic shape type assigned to the RVF occurrence (point or polygon)
poly_type	Polygon type	If the RVF occurrence was assigned a shape_type of polygon, was it admin (GAUL/GADM), custom, or buffer?
buffer_radius	Buffer radius	If a MERS-CoV occurrence was assigned a buffer, what is the radius in km?
poly_reference	GAUL/GADM year or custom shapefile	Identifier used to reach the necessary shape file in ArcGIS
poly_id	Polygon ID	A standardized and unique identifier assigned to each GAUL/GADM shapefile
poly_field	Polygon field	Which type of polygon was used to geo-position the occurrence? (e.g., if admin1 polygon was used, enter ADM1_CODE)
site_notes	Site notes	Miscellaneous notes regarding the site of occurrence

Variable	Variable Label	Variable Definition
month_start	Month start	<p>Month that the occurrence(s) began. If the article provided a specific month of illness onset, the month was assigned a number from 1-12 (1=January, 2=February, etc.). If the article did not provide a specific month of illness onset, then researchers assigned a value of 'NA'. If sampling occurred over multiple time periods, the beginnings of these periods were recorded here with semicolons in between them. For example, if a paper reported that sampling was done from December 2011 to February 2012, and then again from August 2012 to September 2012, this field would read (12;8).</p>
month_end	Month end	<p>Month that the occurrence(s) ended, defined as the date a patient tested negative for MERS-CoV. If the article provided a specific month for recovery, the month was assigned a number from 1-12 (1=January, 2=February, etc.). If the article did not provide a specific month of symptom onset, then researchers assigned a value of 'NA'. If sampling occurred over multiple time periods, the ends of these periods were recorded here with semicolons in between them. For example, if a paper reported that sampling was done from December 2011 to February 2012, and then again from August 2012 to September 2012, this field would read (2;9).</p>

Variable	Variable Label	Variable Definition
year_start	Year start	<p>Year that the occurrence(s) began. If the year of illness onset was not provided in the article, the IHME standard was used:</p> <p>(year_start = publication year – 3)</p> <p>If sampling occurred over multiple time periods, the beginnings of these periods were recorded here with semicolons in between them. For example, if a paper reported that sampling was done from December 2011 to February 2012, and then again from August 2012 to September 2012, this field would read (2011;2012).</p>
year_end	Year end	<p>Year that the occurrence(s) ended. If the article did not provide a specific year for recovery, the IHME standard was used:</p> <p>(year_end = publication year – 1)</p> <p>If sampling occurred over multiple time periods, the ends of these periods were recorded here with semicolons in between them. For example, if a paper reported that sampling was done from December 2011 to February 2012, and then again from August 2012 to September 2012, this field would read (2012;2012).</p>
year_accuracy	Year accuracy	<p>If years were reported, this field was assigned a value of '0'. If assumptions were required, this field was assigned a value of '1'.</p>

### Custom Polygons

These are derived from the data extracted from the literature review. They are documented in the poly\_reference field of the IHME\_RVF\_OCCURRENCE\_Y2020M02D28.CSV file (where the type in the poly\_type field is "custom").



## ***Additional Information***

### **Terms and Conditions**

<http://www.healthdata.org/about/terms-and-conditions>

### **Contact information**

To request further information about this dataset, please contact IHME:

Institute for Health Metrics and Evaluation

2301 Fifth Ave., Suite 600

Seattle, WA 98121

USA

Telephone: +1-206-897-2800

Fax: +1-206-897-2899

Email: [data@healthdata.org](mailto:data@healthdata.org)

[www.healthdata.org](http://www.healthdata.org)

These files may be updated periodically, so we appreciate hearing feedback or additional information about how these data are being used.