

Training Report

RISK ANALYSIS AND RISK MAPPING FOR TRANSBOUNDARY ANIMAL DISEASES

12-16th December 2016

**Bernard Bett
International Livestock Research Institute**

Summary

We conducted a 5-day training on risk analysis, statistical modelling and risk mapping in Entebbe for 37 veterinary epidemiologists from all the IGAD countries including Tanzania between 12 and 16th December 2016. The training aimed to sensitize the participants on the methods they could use to analyze epidemiological data collected from their cross-sectional surveys supported by the *Standard Methods and Procedures in Animal Health (SMP-AH)* and *Improving Animal Disease Surveillance In Support of Trade in IGAD Member States (STSD)* projects led by African Union Interafrican Bureau for Animal Resources (AU IBAR). Participants were selected based on their involvement in the implementation of these projects and high levels of expertise on veterinary epidemiology. They had also completed multiple trainings on the subject and so this training provided a review on the specific methods that were required for the analysis of the data collected. The training was administered using power point presentations, plenary demonstrations and group exercises. At the end of the training, the project coordinators formulated a plan for that would be used to support the country teams to complete the analyses and subsequently develop technical reports and publications.

1 Introduction

The SMP-AH and STSD projects have generated a lot of data/reports that needed to be analyzed/validated to enable the development of reliable project reports and policy documents that could be used to guide prevention and control of trans-boundary animal diseases (TADs). The SMP-AH project, for instance, had supported value chain studies to estimate the risk of release of foot and mouth disease virus (FMDv) and *Mycoplasma spp.*, the causative agent for contagious bovine pleuropneumonia (CBPP), via live animals and meat exported from Borena-Adama beef supply chain. The STSD project on the other hand, had funded, cross sectional surveys on Rift Valley fever (RVF), FMD, CBPP and *peste des petits ruminants* (PPR) in some of the IGAD countries; each country involved in the project was therefore expected to analyze their data and generate publishable results and risk maps where possible. The projects coordinators convened this training workshop in Entebbe to allow all the countries that participated to share experiences on the cross sectional surveys and revise some of the methods they could use for this task.

The participants were identified through their respective Chief Veterinary Officers (CVOs) based on two key criteria: (i) high involvement in the implementation of one or both of the two projects, and (ii) good expertise on veterinary epidemiology. The training was therefore designed with an assumption that the participants had had a good exposure to quantitative methods required for the analysis of their data and the key focus was to harmonize methods required for the work across countries. The training used R although STSD project had purchased STATA for some countries. R was prioritized over STATA because all the participants could download it and participate fully in group exercises and it could be used to run multiple tasks, including disease mapping.

Extensive statistical and mathematical procedures have been developed for implementing risk analysis and risk mapping and hence it is good to highlight that this training only covered basic procedures under each topic. It is clear therefore that much more work will be needed in the course of data analysis given that field data often require more technical analytical procedures than those covered here.

The report is structured into three sections. Section 1 covers Risk Analysis, section 2 Risk Factor Analysis and section 3 Risk Mapping.

2 Training sessions

2.1 Risk analysis (led by Sam Okuthe and Bernard Bett)

Principles of qualitative and quantitative risk analysis were covered in the first two days of the training. On the first day, methods for **qualitative risk analysis** based on the World Organization for Animal Health (OIE) framework were reviewed. Sam Okuthe led initial discussions which covered the following topics:

- Definition of risk analysis
- Principles of risk analysis in trade focusing more on equivalence
- Components of risk analysis – hazard identification, risk assessment, risk management and risk communication
- Components of risk assessment – release, exposure and consequence assessment as well as risk estimation
- Parameters estimates, their uncertainty and variability estimates and how to combine them during risk estimation

This presentation was followed by a review of an analysis that had been done by ILRI under the SMP-AH project to assess the risk of release of FMDv and *Mycoplasma mycoides mycoides* SC via live animals and beef exported from the Borena-Adama value chain in Ethiopia. Bernard gave a presentation outlining the steps used to complete the analysis including:

- Review of the beef value chain targeted for the assessment
- Development of the risk pathways in stakeholder meetings
- Collation of secondary data and information and how they were used to estimate risk parameters
- Risk estimation based on the combination matrix developed by Zepeda (1998).

After this discussion, the participants were distributed into four groups and asked to review a draft report from the assessment. They were expected to identify: (i) areas that required more data/information and (ii) parameters whose estimates needed to be reviewed. This exercise covered most of the afternoon. Feedback on the group work was presented on the second day and were later used to refine the risk assessment report developed by ILRI.

For the rest of the time on second day, Bernard reviewed methods for **quantitative risk analysis**. A distinction was made between methods used to estimate qualitative and quantitative parameters while emphasizing that variability (and to some extent uncertainty) of the latter type of parameters are usually estimated using statistical distributions. Bernard illustrated statistical distributions that are often used including Poisson and Gamma (for Poisson processes e.g. rate/incidence problems), Beta (for discrete events e.g. proportions), and uniform distributions.

Exercises on quantitative analyses aimed to reproduce the results published by Woube et al (2015)¹ which estimated the risk of release of *Mycoplasma mycoides mycoides* SC through the exportation of 280,000 from a disease free zone in northwestern Ethiopia. Woube et al. (2015) used @Risk software but in this training, R was used instead given that @Risk software was not available to the participants. The participants were therefore guided to build risk assessment models using R as illustrated below.

```
#Woube et al. 2015 CBPP risk analysis

#No. of exports per year
no_exported<-runif(1000, min=255001, max=279989)
mean(no_exported); quantile(no_exported, 0.05); quantile(no_exported, 0.95)

#individual probabilities
p1<-runif(10000, min=0.006, max=0.48)
mean(p1); quantile(p1,0.05); quantile(p1,0.95)

p2<-runif(1000, min=0.1, max=0.2)
mean(p2); quantile(p2,0.05); quantile(p2,0.95)

p3<-runif(1000, min=0.022, max=0.08)
mean(p3); quantile(p3,0.05); quantile(p3,0.95)

p4<-runif(1000, min=0.0000037, max=0.67)
mean(p4); quantile(p4,0.05); quantile(p4,0.95)

p5<-1-runif(1000, min=0.8, max=0.9)
mean(p5); quantile(p5,0.05); quantile(p5,0.95)

p6<-runif(1000, min=0.022, max=0.08)
mean(p6); quantile(p6,0.05); quantile(p6,0.95)

#Overall probability
p=p1*p2*p3*p4*p5*p6

mean(p); quantile(p, 0.05); quantile(p, 0.95)

#expected number of infected animals exported
ninfe<-p*no_exported
mean(ninfe); quantile(ninfe, 0.05); quantile(ninfe, 0.95)

#Probability of at least one exported animal infected
```

¹ Woube, Y.A., Dibaba, A.B., Tameru, B., Fite, R., Nganwa, D., Robnett, V., Demisse, A., Habtemariam, T., 2015. Quantitative risk assessment of entry of contagious bovine pleuropneumonia through live cattle imported from northwestern Ethiopia. *Prev. Vet. Med.* 122, 61–69.

```
expinf<-1-((1-p)^no_exported); mean(expinf); quantile(expinf, 0.05); quantile(expinf, 0.95)

#Number of animals exported to get the first infected case

nn<-rnbinom(no_exported, size=1, prob=p); mean(nn)

#Years until the first infected animal is exported
years<-1/ninfe
mean(years)
quantile(years, 0.05)
quantile(years, 0.95)
```

Session 2: Risk factor analysis (led by Bernard Bett)

Descriptive and analytical approaches for analyzing epidemiological data were reviewed during this session (offered in the third and fourth day of the training). This provided a background for the multivariable modelling required for risk factor analysis as well as risk mapping. Data from a cross sectional survey on rotavirus in piglets in Busia and Teso counties, western Kenya published by Amimo et al. (2017)², were used for demonstrations and group exercises. The structure of the data is illustrated in Table 1. All the analyses were ran in R – trainees were therefore guided through the procedures required to implement each task.

² Amimo, J.O., Otieno, T.F., Okoth, E., Onono, J.O., Bett, B., 2017. Risk factors for rotavirus infection in pigs in Busia and Teso subcounties, Western Kenya. *Trop. Anim. Health Prod.*49(1): 105 – 112.

Table 1. Structure of the data used for demonstrations and exercises used in Session 2

ID	pigpop	humanpop	altitude	to_road	to_river	to_town	longitude	latitude	sex	age	breed	house	herd	RV
1	30.99521	4274	1147	2113.562	695.1723	2.510676	34.00583	0.2095	1	0	2	1	1	1
2	30.99521	4274	1146	2050.506	823.2817	2.480699	34.00687	0.20875	NA	1	2	0	1	0
3	30.99521	4274	1147	1999.728	837.6055	2.434765	34.00732	0.20888	NA	1	2	0	2	0
4	30.99521	4274	1147	1999.728	837.6055	2.434765	34.00732	0.20888	NA	1	2	0	2	0
5	30.99521	4274	1163	1997.181	1880.576	3.07154	34.01005	0.19972	1	0	2	1	5	0
6	30.99521	4274	1163	1997.181	1880.576	3.07154	34.01005	0.19972	1	0	2	1	5	0
..														
206	79.83804	3043	1361	701.5457	4554.388	5.873693	34.36407	0.74837	0	0	2	0	1	0

Description of the variables:

Variable	Description
ID	Subject's unique identification number
pigpop	Estimated population density of pigs by sublocation
humanpop	Estimated human population density by division
altitude	Altitude -- m above sea level – at the sampling site from digital elevation model
to_road	Euclidean distance from a major road to a sampling site in m
to_river	Euclidean distance from a major river to a sampling site in m
to_town	Euclidean distance from a town to a sampling site in km
longitude	Longitude in decimal degrees
latitude	Latitude in decimal degrees
sex	Sex of an animal: 1 – Male, 0 – Female, NA – Missing data
age	Approximate age of an animal: ≤4 months, 1 - > 4 months
breed	Breed of the animal: 1 – cross breed, 2 – local
house	Husbandry: 0 – free range with tethering/housing, 1 – full-time housing/tethering
herds	Herd identification number
RV	Rotavirus infection status: 1 – infected, 0 – uninfected

Importing and exploring data

The first demonstration involved setting up the working directory, importing data to R and checking names of the variables, number of records, etc. The original data file was saved as a CSV file with the name: Rotavirus_day2³. The commands used for these tasks are:

```
setwd("D:/Entebbe")           #function sets a working directory
data<-read.csv("Rotavirus_day2.csv", header=TRUE) # importing data from the working directory
names(data)                   # generating names of the variables
head(data)                    # viewing the first 6 rows of the data
tail(data)                    # viewing the last 6 rows of the data
```

The participants were notified that it was a useful practice to check the structure of the data before commencing any analyses. They were also sensitized that it was critical to develop an analytical framework such as a causal web diagram that would guide the analysis in terms of identifying relationships and variables to analyze.

Descriptive analyses

Descriptive analyses were done to determine the distribution of the independent variables and their association with the outcome of interest, i.e., rotavirus infection. Similar analyses including histograms, box plots, mean, medians, quartile ranges, variance and T-test were implemented for all continuous variables -- pig population, human population, altitude, distances to major road, river and towns to expose the participants to a range of tests that could be used. The commands used for pig population (pigpop variable) are given below. The participants were asked to replicate these analyses using the other continuous variables by replacing "pigpop" with the name of the variable being analyzed.

```
#pig population
hist(data$pigpop)           #draw a histogram
boxplot(data$pigpop)        #draw a box plot
boxplot(data$pigpop~data$RV) #box plot of pig population over RV
median(data$pigpop)         #obtain a median
quantile(data$pigpop, 0.5)  #another command for obtaining the median
quantile(data$pigpop, 0.25) #obtain the first quartile
mean(data$pigpop)           #obtain the mean
var(data$pigpop)            #variance
t.test(data$pigpop~data$RV) #implement T test
```

Analyses involving categorical variables were implemented by generating frequency tables. Categorical variables included sex, breed, house, and age. First, individual tables for the outcome – RV – and each of the categorical variables were generated to determine their frequencies. After this joint tables were generated and used to estimate Chi Square tests. The commands used for these analyses with sex as the

³ Data can also be imported as a text file (txt), excel, etc. and there are instructions on how to achieve this on-line.

independent variable are given below. The participants were guided on how to interpret the results of each of the commands and asked to replicate the analyses for the other variables.

```
# sex
table(data$RV)           #Generates a frequency table for RV
table(data$sex)          # Generates a frequency table for sex
ts<-table(data$RV, data$sex) # Generates a 2x2 table for RV and sex
ts                       # an object used to store results of the command above
chisq.test(ts)           # Implements a Chi square test using the results stored in ts
```

Variable labels

Participants were guided on how to assign labels to categorical variables that are entered in the database as numerical codes or dummies. An example used was on assigning the labels “male” and “female” to the sex variable that was entered as 1 and 0 for male and female, respectively. This was done in three stages. First, the participants were asked to generate a frequency table for sex. This was followed by a step to assign the labels to this variable and finally, a step to regenerate the frequency table to visualize the changes made. The commands used were:

```
table(data$sex)           #generates a simple table
data$sex<-factor(data$sex, levels=c(0,1), labels=c("female", "male")) #assigns labels to sex
table(data$sex)           #generates a new table with labels
```

Changing a continuous variable to factor/categorical variable

The participants were shown procedures for changing a continuous variable into a factor variable. The first step would involve defining the number of classes one would like to generate from a continuous variable and minimum and maximum values that will be used to define a class. There are good guidelines on how to create classes from a continuous variable depending on the distribution of the variable being studied and the objectives of the analysis. Similarly, there are a number of commands in R for implementing this procedure, including *cut()*. A mechanistic approach that required multiple commands was used in this training for the participants to follow the logic behind variable categorization process.

The variable used was pig population (*pigpop*); its frequency distribution is demonstrated in Figure 1. Classes designed (arbitrarily for demonstrations only) were: <25, 25 - <30, 30 - <54 and \geq 54. The first step involved generating a new variable *pigpop2* to be the factor variable. The variable was then indexed to allow the definition of categories as given below.

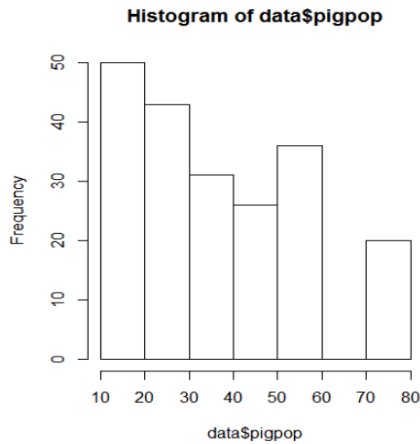


Figure 1. Histogram of pig population in Busia County

```

data$pigpop2<-rep(0, nrow(data))
data$pigpop2[data$pigpop<25]=1
data$pigpop2[data$pigpop>=25 & data$pigpop<30]=2
data$pigpop2[data$pigpop>=30 & data$pigpop<54]=3
data$pigpop2[data$pigpop>=54]=4
#generates a new variable pigpop2
#assign code 1 to pigpop <25
#assign code 2 to pigpop 25 - <30
#Assign code 3 to pigpop 30 - <54
#assign code 4 to pigpop ≥54

```

Building a regression model

Most of the participants had conducted regression analyses before and so a plenary discussion was done to review the principles of developing and evaluating a regression model and key assumptions of a linear and logistic regression models. Demonstrations done and group exercises focused on linear and logistic regression models. Topics covered in these discussions were:

- implementing backward and forward techniques for variable selection and elimination
- evaluating linearity assumption using polynomials,
- checking for goodness of fit of a model, and
- Interpreting the results

i. Building a model

For logistic regression modeling, the outcome/dependent variable used was RV (positive [1] or negative [0]). Procedures for implementing backward and forward selection of independent variables – using age, sex, breed, house, pigpop, humanpop, altitude, etc. were demonstrated. The first step involved setting all the categorical independent variables as factors to ensure that the software treated as so. Logistic regression modelling was implemented using the *glm()* function.

Functions used to set categorical variables -- age, sex, breed, house as factor variables were:

```
as.factor(age)
as.factor(sex)
as.factor(breed)
as.factor(house)
```

A full model comprising categorical and continuous variables was then formulated using:

```
model1<-glm(formula=data$RV~1+data$age+data$sex+data$breed+data$house, family=binomial())
summary(model) # to show the results stored in the object model
```

The number [1] is added to the regression formula to specify that an intercept should be included in the model.

The participants were asked to work in groups to build a parsimonious model using backward and forward variable selection procedures. Wald test was used to determine variables that were significant in the model; those not significant were removed. The final model that was used for discussion in the trainings (based on the data described in Table 1) is presented in the text box below.

```
Call:
glm(formula = data$RV ~ 1 + data$age + data$house, family = binomial())

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0288  -0.8177  -0.5819   1.3337   1.9285

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.3601     0.2421  -1.488   0.1369
data$age     -0.7664     0.3392  -2.259   0.0239 *
data$house   -0.5638     0.3235  -1.743   0.0814 .
---
Signif. Codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 243.00  on 205  degrees of freedom
Residual deviance: 232.75  on 203  degrees of freedom
AIC: 238.75
```

The interpretation discussed in the plenary was as follows:

Age and housing were kept in the model even though the later was not significant with an alpha = 0.05. The results show that compared to a piglet of ≤ 4 months, that of >4 months has lower log odds of RV infection by 0.77 units (the reference class used in the model is ≤ 4 months). Similarly, an animal raised under full-time housing conditions or tethered had lower log odds of RV infection compared to that raised under free-range system of management (by 0.56 units). The reference used for the latter case was free-range system. The model did not show any evidence for over-dispersion as the ratio between the residual deviance and degrees of freedom

(232.75/203) was close to 1. In this case, no further analyses were done. There are other goodness of fit test that could be used for this case such as Hosmer and Lemeshow test.

Procedures for building a linear model were demonstrated by using pigpop as an outcome and distance to road (to_road) as the only predictor. The function used was:

```
model2<-glm(formula=pigpop~1+to_road)
summary(model)
```

The main differences between the linear and logistic model presented above were also discussed. Once a final model was obtained, procedures for conducting residual analysis were demonstrated. Commands that can be used to generate Pearson residuals, fitted values and to plot these residuals with fitted values are:

```
pearson.res<-residuals(model1, "pearson") #generate Pearson residuals
fitted<-fitted(model1) #generate fitted values
plot(fitted, pearson.res, xlab="Fitted values", ylab="Pearson residuals")
```

Session 3: Risk mapping (led by Bernard Bett)

The fifth day of the training focused on risk mapping. A conceptual framework for conducting spatial analysis described by Pfeiffer et al. (2008)⁴ was presented. The framework is illustrated in Figure 3.

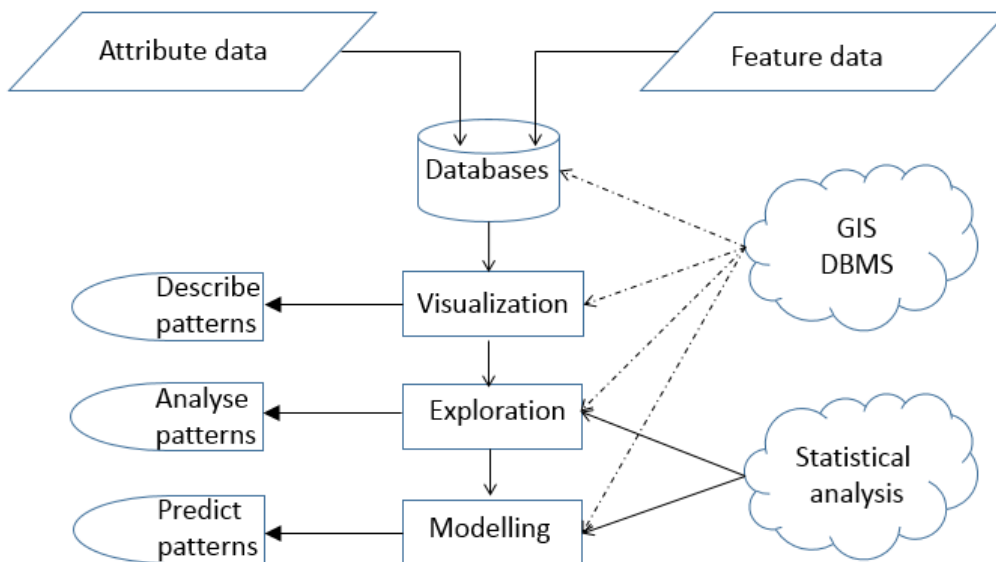


Figure 3. A conceptual framework for spatial analysis (Source: Pfeiffer et al., 2008)

⁴ Pfeiffer, D.U., Robinson, T.P., Stevenson, M., Stevens, K.B., Rogers, D.J., Clements, A.C.A., 2008. Spatial analysis in epidemiology. Oxford University Press Inc. New York.

The framework illustrates the different stages/analyses that spatial data could be taken through. It was emphasized that a good risk map should be generated from a good model – such as those implemented in Session 2—and not just plotting out raw data as a map. This is because a map of raw data e.g. disease prevalence will only represent disease burden in areas sampled. A model-based risk map provides predicted risk estimates based on spatial data e.g. land use patterns and so one will be able to gauge disease burden across a given spatial domain (e.g. a country) rather than just focusing on the sampled areas.

In the framework, disease data are combined together with attribute data in a database (e.g. GIS database). The participants were notified that most of the disease maps generated in the region have undergone all the procedures given in the Framework.

Point verses areal data

Types of spatial data sets – point verses areal -- were reviewed. Point data are represent as plot of cases using Cartesian coordinates. They allow for the representation of multiple cases in the same location and attributes of each case can be readily described. Areal or aggregated data involves summarizing group of individual data into single values as means, variance, etc., and such a value is assigned to a specific location e.g. district or county.

Spatial verses edge effect

A distinction was made between spatial and edge effect.

Spatial effect is realized when values measured at locations that are close together have similar values. If this effect was constant on x-y plane, then spatial effect is considered to be stationary (isotropic), otherwise it is non-stationary (anisotropic). If spatial effect (dependence) is present, data from geographically close locations contribute less additional information to the overall database than if they were further apart. For statistical modelling, this effect is often accounted for by specifying a neighborhood structure using a graph or a mesh for areal or point data, respectively. This structure will allow for the specification of spatial random effects during modelling.

Edge effect is realized when data along some of the edges of a spatial domain are either incomplete, unavailable or no-existent e.g. near a sea.

Installation and reading of R packages for spatial analysis

Before commencing demonstrations and group work, the participants were guided on how to install some of the packages required for spatial analysis in R. These were:

```
install.packages("maptools") #required for reading and manipulating geographic data, e.g. ESRI shapefiles
install.packages("sp")      #provides classes and methods for spatial data
install.packages("rgdal")   #useful for multiple operations including map projections
install.packages("rgeos")   #useful for topological operations on geometries
install.packages("ggplot2") #for elegant R graphics
```

```
install.packages("ggmap")      #used to obtain various maps including google maps
install.packages("scales")     #used to define breaks and labels for graph axes and legends
install.packages("RColorBrewer") #provides colour schemes for maps and other graphics
install.packages("raster")     #reading, analyzing and manipulating gridded data
```

After installing the packages, follow up commands were required to load the packages for use; this is usually done using either `library()` or `require()`.

```
library(maptools)
library(sp)
library(rgdal)
library(rgeos)
library(ggplot2)
library(ggmap)
library(scales)
library(RColorBrewer)
library(raster)
```

Exercises

Reading and plotting the Kenya divisions map

The division map of Kenya was read into R and plotted using the commands:

```
ke.map <- readShapePoly("KenyaDivNames.shp") #importing the map from the working directory
plot(ke.map)                                #plotting the map
```

Figure 4 illustrates the output produced.

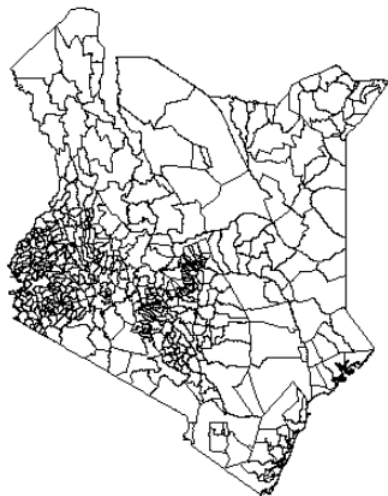


Figure 4. Map of Kenya showing divisional boundaries

Mapping RVF outbreaks aggregated at the division level

Data on RVF outbreaks in Kenya from 1979 to 2007 were used to demonstrate how to generate a choropleth map. This involved merging the outbreak data, aggregated at the division level, with the shapefile in Figure 4, and plotting the map to show the distribution of cases in space. The outbreak data were first imported to R and the first six rows of the data inspected to determine its structure. The commands used for this procedures were

```
outbreaks<-read.csv("RVF_outbreaks.csv")    #to read the data file RVF_outbreaks from the working directory
head(outbreaks)                            #inspect the first six rows of the data
```

The first six rows of the data read to R were

```
  DIVID case
1 10101    0
2 10102    0
3 10103   13
4 10104    0
5 10105    0
6 10106   11
```

DIVID represents the division ID while case represent the number of cases aggregated by division.

The shapefile and the RVF data were then merged using the following commands

```
shape.f<-fortify(ke.map, region="DIVID")    #creates a data frame from shapefile
merge.shape<-merge(shape.f, outbreaks, by.x="id", by.y="DIVID") #merges shapefile data frame and cases
final.data<-merge.shape[order(merge.shape$order), ] #sort the data by order
```

And finally, the final data were mapped using ggplot function

```
ggplot() +
  geom_polygon(data = final.data,
              aes(x = long, y = lat, group = group, fill = case),
              color = "black", size = 0.25) +
  scale_fill_distiller(palette = "YlOrRd", direction=1, limits=c(0,38), breaks = pretty_breaks(n = 8))+
  coord_map()+theme_nothing(legend=TRUE)+labs(title="RVF incidence in Kenya")
```

The map below – Figure 5 – was generated in R using the ggplot() function. Note that the parameters of the function can be changed to vary the presentation of the map.

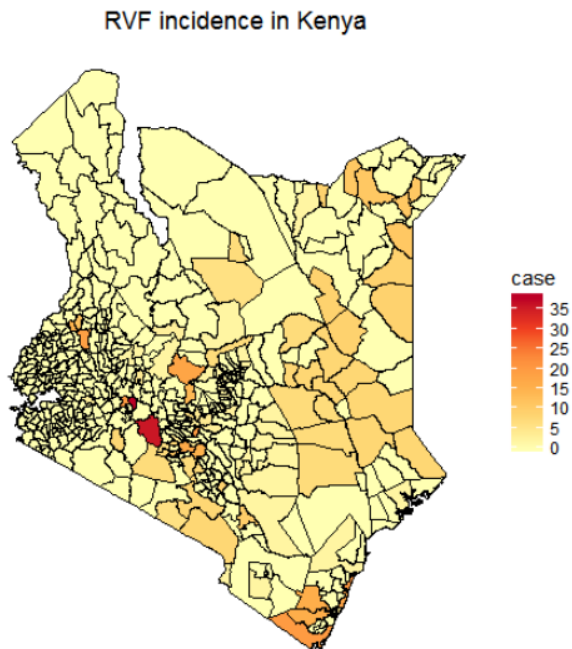


Figure 5. A map showing the distribution of the number of RVF cases in Kenya. The cases represent the total number of outbreaks by division for the period 1979 – 2007

Point referenced data

Procedures for plotting point referenced prevalence data were also demonstrated. This required installing and running an additional package “geoR” using the commands:

```
Install.packages("geoR")           #installation of a package for geostatistical analysis
library("geoR")                    #reading geoR()
```

The STSD project had generated point referenced data for a number of diseases and so the participants required skills on how to map such data and analyze them. The project team did not have existing data for running these demonstrates so comparable data were simulated theoretically for demonstrations only. This simulation used Kenya as the spatial domain and included two variables: number of animal sampled in each sampling location simulated and the number of positive cases from the total number sampled. The simulated data had this structure:

	v1	longitude	latitude	n	pos
1	1	36.90023	-1.2159840	97	15
2	2	36.78703	-1.2492350	87	9
3	3	36.79365	-1.0686140	79	30
4	4	36.62521	-1.1406090	75	31
5	5	36.63256	-1.2385320	80	38
6	6	37.21401	-0.5503597	61	10

Where: V1 – a serial number used as an ID, longitude, latitude – location of a theoretical sampling point, n and pos - theoretically generated total number of animals sampled and number of positives observed.

The data were imported into R using:

```
RVF.geo<-read.csv("pointpattern_rvf.csv", header=TRUE) #data named as pointpattern_rvf in working dir
```

Mapping point referenced data

A logit transformation of the cases was done using the command:

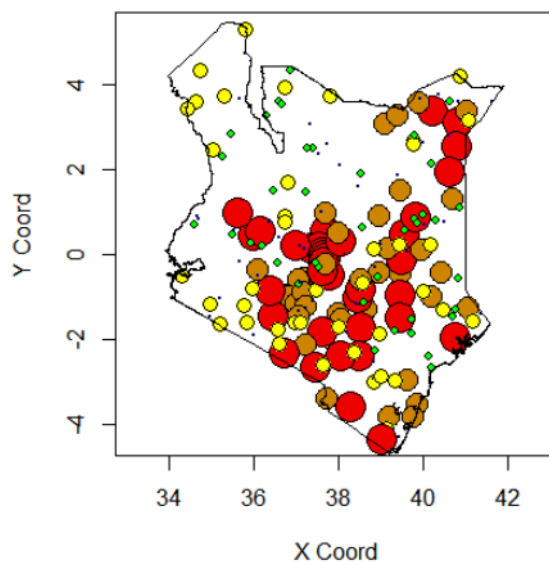
```
RVF.geo$logit <- log((RVF.geo$pos + 0.5)/(RVF.geo$n - RVF.geo$pos + 0.5)) #logit transform  
coord<-as.matrix(RVF.geo[,2:3]) #extract coordinates
```

The Kenya division map was also imported to R and its division boundaries dissolved so as to have one map with the international boundaries only. These steps were implemented using the commands:

```
ke<-readShapePoly("KenyaDiv.shp") #import map of Kenya with division boundaries  
ke_2 <- unionSpatialPolygons(ke, rep(1, nrow(ke))) #dissolve all the division boundaries
```

The transformed data were then plotted using the commands

```
data.geo<-as.geodata(cbind(RVF.geo$longitude,RVF.geo$latitude,RVF.geo$logit))  
points(data.geo,cex.min=.1,cex.max=3,pt.div="quint")  
plot(ke_2, add=TRUE)
```



The option used for dividing points, i.e. `pt.div="quint"` divides the data into 5 different sizes depending on the quintiles of the data. A description of this functionality can be found at <https://cran.r-project.org/web/packages/geoR/geoR.pdf>

Figure 6. Simulated prevalence of RVF in Kenya (for training purposes only). Data is divided into 5

Graphics publication standards

Many journals emphasize that graphics have to meet minimum resolution of 300 dots per inch (dpi). The higher the dpi value the finer the resolution but the more memory is needed for storing the graphic.

To achieve this, we create an object to store the figure in our working directory with a resolution needed and run the code for generating the graph or a plot. For instance if we are using the `ggplot()` function above, we will develop the command as follows:

```
tiff("Fig.5.tiff", width=8, height=8, units="cm", res=300)           #specify the file name (Fig.5) and its resolution
ggplot() +
  geom_polygon(data = final.data,
              aes(x = long, y = lat, group = group, fill = case),
              color = "black", size = 0.25) +
  scale_fill_distiller(palette = "YlOrRd", direction=1, limits=c(0,38), breaks = pretty_breaks(n = 8))+
  coord_map()+theme_nothing(legend=TRUE)+labs(title="RVF incidence in Kenya")
```

3 Way forward

We have covered basic procedures for risk analysis, risk factor analysis and risk mapping but more is needed to gain capacity required for analyzing the collected data. There was not enough time to cover the last session in detail. We expect that AU IBAR will follow up with the suggested plans for providing extended support to the participants.

Annex 1: Training program

Title:		
Day 1 - 12 th Dec 2016		
0800 - 0830	Registration	Facilitator
0830 - 0930	Opening session General introductions	Henry/James
930 - 1000	Risk analysis – concepts	Okuthe
1000 - 1030	Health break	
1030 - 1130	Risk analysis – applications in animal health	Okuthe
1130 - 1230	Presentation of risk analysis report from Ethiopia	Bernard + reps from Ethiopia
1230 - 1330	Lunch	
1330 - 1500	Evaluation of risk analysis report from Ethiopia	Participants
1500 - 1530	Health break	
1530 - 1700	Feedback on risk analysis report from Ethiopia	Bernard + reps from Ethiopia
Day 2 - 13 th Dec 2016		
0830 - 930	Review of the STSD survey designs and tools used	Participants
930 - 1030	Principles of survey data analysis	Silvia
1030 - 1100	Health break	
1100 - 1130	Visualization and exploration of data	Bernard
1130-1230	General discussions (Discussion on issues encountered during field activities)	Participants
1230 - 1330	Lunch	
1330 - 1500	Group exercises – data preparation, visualization and exploration	Participants
1500 - 1530	Health break	
1530 - 1600	Statistical analysis – descriptive analysis	Bernard
1600 - 1730	Group exercise – descriptive analyses	Participants
Day 3 - 14 th Dec 2016		
0830 - 0930	Statistical analysis – univariable and multivariable regressions / models	Silvia
0930 - 1030	Group work – risk factor analysis	Participants
1030 - 1100	Health break	
1100 - 1230	Group work – risk factor analysis	Participants
1230 - 1330	Lunch	
1330 - 1500	Group work – Presentation of the multivariable models	Participants
1500 - 1530	Health break	
1530 - 1630	Introduction to spatial analysis in epidemiology	Bernard
1630-1700	General discussion	
Day 4 - 15 th Dec 2016		
830 - 930	Processing spatial data	Bernard
930 - 1100	Group exercises	Participants
1100 - 1130	Health break	
1130 - 1230	Visualization and exploratory analysis of spatial data	Bernard
1230 - 1330	Lunch	
1330 - 1430	Fitting spatial models – area data	Bernard

1430 - 1500	Group work - Visualization and exploratory analysis of spatial data	Participants
1500 - 1530	Health break	
1530 - 1630	Fitting spatial models – area data	Bernard
1630-1730	Discussion	
Day 5 - 16th Dec 2016		
830 - 930	Fitting spatial models – point data	Bernard
930 - 1100	Group work	Participants
1100 - 1130	Health break	
1130 - 1230	Group work	Participants
1230 - 1330	Lunch	
1330 - 1500	Group work	Participants
1500 - 1530	Health break	
1530 - 1630	Group work	Participants
1630-1730	Wrap up and way forward	Henry/James

Annex 2: List of participants

No	Name	Affiliation	Sex
1	Dr. Hiver Boussini	Animal Health Officer, AU IBAR	M
2	Dr. Ibrahim Gashash Ahmed	Information Systems Manager, AU IBAR	M
3	Dr. James Wabacha	SMP –AH Project Coordinator, AU IBAR	M
4	Ms. Adeline Akinyi Oduor	Administrative Assistant, AU IBAR	F
5	Mr. Charles Lodiaga	Accountant, AU IBAR	M
6	Mr. Alexander Eyong	Travel Clerk, AU IBAR	M
7	Dr. James Magona	Veterinary Epidemiologist, AU IBAR	M
8	Dr. Samuel Okuthe	Regional Epidemiologist, FAO	M
9	Dr. Harry Oyas Okuom	SMP-AH Country Focal Person, Kenya	M
10	Dr. Salome Wanyoike	Deputy Director, DVS, Kenya	F
11	Miss Anima Jematia Kigen	Senior Veterinary Officer/Epidemiologist, Kenya	F
12	Dr. Aluma Araba Ameri	Director, Ministry of Livestock and Fisheries, South Sudan	M
13	Dr. Jada Rombe Wani Lokak	Deputy Director for Epidemiology and Information, South Sudan	M
14	Dr. Alor Kwaja Kuol Arop	Deputy Director for Epidemiology and Information, South Sudan	M
15	Dr. Ahmed Hachi Dirir	Veterinarian, Djibouti	M
16	Dr. Abdi Mahamoud Elmi	Veterinarian, Djibouti	M
17	Dr. Ahmed Wafi Ibrahim Allaleh	Veterinary Epidemiologist, Djibouti	M
18	Dr. Efreem Ghebremeskel Habteyohannes	Director of National Veterinary Laboratory National Animal and Plant Health Laboratory, Eritrea	M
19	Dr. Afewerki Mehreteab Ghebressassie	Director of Animal and Plant Health Division, Agricultural Extension Department, Eritrea	M

20	Dr. Uqbazghi Kefle Mesghun	Director of Debub Region, Agriculture and Land Department, Eritrea	M
21	Dr. Dan Tumusiime	Senior Veterinary Officer, Department of Animal Health, Uganda	M
22	Dr. Noelina Nantima	Principal Veterinary Officer Epidemics, Department of Animal Health, Uganda	F
23	Dr. Nicholas Kauta	DVS, Uganda	M
24	Dr. Anna Rose Ademun Okurut	Assistant Commissioner Veterinary Diagnostics and Epidemiology, Uganda	F
25	Dr. Robert Mwebe	Senior Veterinary Officer – Epidemiology, Uganda	M
26	Mr. Paul Kirabo	Data Entrant, Uganda	M
27	Dr. Sam Richard Rekuma Erechu	Senior Veterinary Officer, Uganda	M
28	Dr. Luka Selemani Makungu	Epidemiologist, Ministry of Agriculture, Livestock and Fisheries, Tanzania	M
29	Dr. Solomon Wilson Nong'ona	Zonal Epidemiologist, Veterinary Services Tanzania	M
30	Dr. Joram Elisha Mghwira	SMP – AH National Focal Point Tanzania, Veterinary Services, Tanzania	M
31	Dr. Tilahun Zenebe Alemu	Associate Researcher (Veterinary Epidemiologist) Department of Veterinary Epidemiology, Ethiopia	M
32	Dr. Wondimagegn Dejene Tuffa	Senior Inspector and Regulator for Quarantine Labs Quarantine Import Export Inspection and Certification, Directorate, Ministry of Livestock and Fisheries, Ethiopia	M
33	Dr. Kefyalew Chirkena Bali	Expert, Livestock Identification, Traceability and Animal Welfare Directorate, Ethiopia	M
34	Dr. Gashaw Beyene Asfaw	Surveillance Expert, Epidemiology Directorate, Ethiopia	M
35	Dr. Yismashewa Wegayehu Amenu	Director, Epidemiology, Ministry of Livestock and Fisheries, Ethiopia	M
36	Dr. Ibrahim Hassan Ahmed Ibrahim	Director General, Department of Animal Health and Epizootic Disease Control, Sudan	M
37	Dr. Mohamed Elsadig Ahmed Mansour	Veterinary Researcher, Rift Valley Fever Unit, Ministry of Animal Resources, Fisheries and Rangelands, Sudan	M
38	Dr. Hanan Yousif Mohamed Ahmed	Director, Department of Disease Control, Sudan	F
39	Dr. Abdullahi Araye Addow	EDMU, Department of Animal Health, Somalia	M
40	Dr. Muhamed Yusuf Isse	EDMU, Department of Animal Health, Somalia	M

Annex 3. Risk pathway and probability estimates from a quantitative risk assessment on the risk of release of *Mycoplasma mycoides mycoides* SC from northwestern Ethiopia (Source: Woube et al., 2015)

Fig. 1.3. Risk pathway

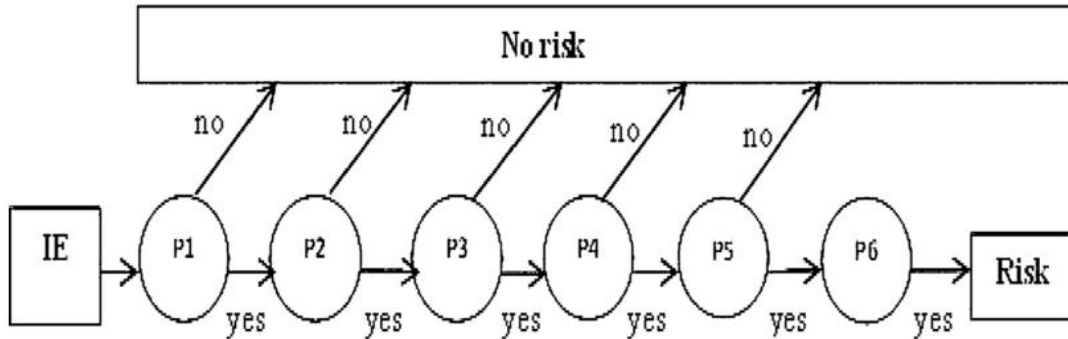


Table 1.3.

Results showing inputs and computer generated outputs for the risk of entry of CBPP through importation of cattle from northwestern DFZ of Ethiopia.

Probability	Minimum	5th Percentile	Mean	95th Percentile	Maximum
Input probabilities:					
P1	0.006	0.029	0.242	0.457	0.48
P2	0.1	0.105	0.15	0.195	0.2
P3	0.022	0.025	0.051	0.077	0.08
P4	0.0000037	0.0369	0.339	0.638	0.67
P5	0.1	0.105	0.15	0.195	0.2
P6	0.022	0.025	0.051	0.077	0.08
Other input:					
No. of export per year (<i>n</i>)	255,001	256,239	267,369	278,742	279,989
Outputs:					
Overall probability (<i>p</i>)	1.95×10^{-10}	1.48×10^{-7}	4.76×10^{-6}	1.53×10^{-5}	5.13×10^{-5}
Expected number of infected animals exported every year	0.0000279	0.0417	1.28	4.3	12.61
Probability of at least one infected animal exported	0.0000658	0.042	0.53	0.99	1
Number of animals exported to get the first infected case	5	10,674	2,725,944	5,912,841	2.14×10^9
Years until the first infected animal is exported	0.0712	0.242	10.15	23.18	4931.86