



SAARC Agriculture Centre



भाकृअनुप
ICAR



राप जा सू विस
NIVEDI

SAARC Regional Training on

ANIMAL DISEASE INFORMATICS AND BIOSTATISTICS

(3-8 May, 2018)



Organized by :

ICAR - National Institute of Veterinary Epidemiology and Disease Informatics,
Yelahanka, Bengaluru - 560 064, India



Sponsored by

SAARC Agriculture Center, Bangladesh

SAARC Regional Training
on
ANIMAL DISEASE INFORMATICS AND
BIOSTATISTICS
(3-8 May 2018)

Sponsored by SAARC Agriculture Center, Bangladesh

TRAINING COMPENDIUM

Course Director

Dr. Parimal Roy

Director, ICAR-NIVEDI, Bangalore

Course Coordinators

Dr. K.P Suresh

Dr. S.S. Patil

Dr. Md. Mudassar Chanda

Dr. Jagadish Hiremath

Organized by :



**ICAR - National Institute of Veterinary Epidemiology
and Disease Informatics (ICAR-NIVEDI)**

Post Box No. 6450

Ramagondanahalli, Yelahanka, Bengaluru - 560064, Karnataka, India

www.nivedi.res.in



© Copyright

**ICAR-National Institute of Veterinary Epidemiology and
Disease Informatics (NIVEDI)**

Printed at :

CNU Graphic Printers

Malleshwaram, Bengaluru - 560 003

Mob. : 9880 888 399





त्रिलोचन महापात्र, पीएच.डी.

एफ एन ए, एफ एन ए एस सी, एफ एन ए ए एस

सचिव एवं महानिदेशक

TRILOCHAN MOHAPATRA, Ph.D.

FNA, FNASc, FNAAS

SECRETARY & DIRECTOR GENERAL



भारत सरकार

कृषि अनुसंधान और शिक्षा विभाग एवं

भारतीय कृषि अनुसंधान परिषद

कृषि एवं किसान कल्याण मंत्रालय, कृषि भवन, नई दिल्ली 110 001

GOVERNMENT OF INDIA

DEPARTMENT OF AGRICULTURAL RESEARCH & EDUCATION

AND

INDIAN COUNCIL OF AGRICULTURAL RESEARCH

MINISTRY OF AGRICULTURE AND FARMERS WELFARE

KRISHI BHAVAN, NEW DELHI 110 001

Tel.: 23382629; 23386711 Fax: 91-11-23384773

E-mail: dg.icar@nic.in

MESSAGE

I am immensely pleased to know that ICAR- NIVEDI is organizing a regional training on 'Animal Disease Informatics and Biostatistics' under the aegis of SAARC Agriculture centre (SAC), Bangladesh from 3rd to 8th May, 2018 at Bengaluru, India.

The role of livestock in livelihood, nutritional and food security of millions of people living in SAARC countries is very significant. Formulation of effective disease control strategies is a daunting task for the animal husbandry sector in most of the countries. In this scenario, animal disease informatics is an emerging field in the livestock sector which will help in drafting livestock disease control strategies through forecasting and forewarning of these diseases which will reduce animal disease burden. In the area of animal disease informatics and biostatistics, the role of ICAR-NIVEDI is noteworthy and with the objective of sharing of these ideas, commonalities and solutions among the researchers of the South Asian Association for Regional Cooperation (SAARC) countries, ICAR-NIVEDI is organizing regional training in the relevant area.

I wish the training all success and complement the organizers for selecting appropriate theme for deliberation which will definitely provide cross learning opportunities in the field of disease informatics, statistical analysis, disease control and prevention within the SAARC region.

(T. MOHAPATRA)

Dated the 27th April, 2018
New Delhi



डॉ. जे.के. जेना

उप महानिदेशक (मत्स्य विज्ञान)

Dr J.K. Jena

Deputy Director General (Fisheries Science)
and

Deputy Director General (Animal Science)

भारतीय कृषि अनुसंधान परिषद

कृषि अनुसंधान भवन-II, पूसा, नई दिल्ली 110 012

INDIAN COUNCIL OF AGRICULTURAL RESEARCH

KRISHI ANUSANDHAN BHAWAN - II, PUSA, NEW DELHI - 110 012

Phone : (Off.) 91-11-25846738, Fax : 91-11-25841955

Email : ddgfs.icar@gov.in, jkjena2@rediffmail.com



April 27, 2018

Message

It gives me immense pleasure that National Institute of Veterinary Epidemiology and Disease Informatics (ICAR- NIVEDI) is organizing a regional training on 'Animal Disease Informatics and Biostatistics' from 3rd to 8th May, 2018 under the aegis of SAARC Agriculture centre (SAC), Bangladesh at Bengaluru, India.

The SAARC Agriculture Centre (SAC) supports livestock sector, which is an essential component for rural livelihood sustainability especially in the rural areas. As a SAARC member country, India is providing training to experts of member countries in animal disease informatics and biostatistics.

Effective control of the diseases in a country demands a thorough understanding about the epidemiology of the disease including identification of risk factors contributing to the occurrence of the diseases. Early warning of disease incidence or outbreaks and the capacity on prediction for spread to new areas is an essential pre-requisite for the effective containment and control of epidemic animal diseases.

ICAR NIVEDI plays a pivotal role in strategizing control measures by forecasting 13 prioritized livestock diseases in the country, this training on Animal disease informatics and biostatistics will provide an avenue for the researchers of the South Asian Association for Regional Cooperation (SAARC) countries to strengthen their knowledge in the area of epidemiology and disease informatics for the effective prevention and control of livestock diseases within the SAARC region.

I wish this endeavor a grand success.


(J.K. Jena)



SAARC AGRICULTURE CENTRE (SAC)



MESSAGE

I am delighted to hear that a training compendium/manual to be published for the SAARC regional training programme on "**Animal Disease Informatics and Biostatistics**" jointly organized by SAARC Agriculture Centre (SAC), Bangladesh and ICAR-National Institute of Veterinary Epidemiology and Disease Informatics (NIVEDI), Bengaluru-560 064, India.

SAARC Agriculture Centre (SAC), under the aegis of SAARC has been working for the promotion of agricultural research & development as well as technology transfer through regional networks among agricultural research/extension institutions and policy makers in the SAARC Member States. ICAR- NIVEDI is one of the specialized premier institute in India undertaking research and developmental activities on animal disease epidemiology and informatics for the promotion of animal health. The regional training on "**Animal Disease Informatics and Biostatistics**" would provide theoretical as well as hands-on knowledge to the participants on modern concept of epidemiology and disease informatics that may expedite diagnosis, prevention and control of livestock and poultry diseases in their respective countries. The training would impart knowledge for effective and timely outbreak response and surveillance for epidemic and emerging infectious diseases. I believe the contents of the compendium/ manual is certainly the store of information related to recent research and development of modern animal epidemiology and informatics. This book is unique and surely a work to treasure for anyone who is interested in pursuing research on animal epidemiology and disease informatics.

I wish all the grand success for this regional training programme and its endeavours.

(Dr. S. M. Bokhtiar)
Director, SAC

BARC Complex, New Airport Road, Farmgate, Dhaka-1215, Bangladesh
Tel: +880-2-58153152, PABX: 8141655, 8141140, 9122808, Fax: +880-2-9124596
E-mail: director@sac.org.bd, Web: www.sac.org.bd



भारतीय कृषि अनुसंधान परिषद्
Indian Council of Agricultural Research
कृषि भवन, नई दिल्ली- 110 001
Krishi Bhavan, New Delhi - 110 001

(M) +91 9412761307, Telefax (O): 011 23386668 Email: ashokkr.icar@gov.in, ashokakt@rediffmail.com

डा० अशोक कुमार

Dr Ashok Kumar

सहायक महानिदेशक (पशु स्वा.)

Assistant Director General (Animal Health)



MESSAGE

I am glad to know that ICAR-National Institute of Veterinary Epidemiology and Disease Informatics (NIVEDI), Bengaluru is again in the front for sharing the knowledge to the researchers of SAARC countries by organization of a regional training on 'Animal Disease Informatics and Biostatistics' on May 3 to 8, 2018 under the aegis of SAARC Agriculture centre (SAC), Bangladesh.

It is well known that the livestock is an important means of livelihood of the farmers augmenting the income from agriculture, especially in developing countries. Scientific management of livestock and thereby increasing the productivity is critical for global food security. This can be achieved by growing the knowledge about epidemiology of animal diseases including identification of risk factors in any specific geographic area and strategizing control measures and management practices for healthy livestock and products. I am sure that the training will be of immense help to the participants from the SAARC countries with constructive exchange of knowledge and experience of NIVEDI in this area.

I wish all the success for the training.

(Ashok Kumar)



भाकृअनुप - राष्ऱीय पशुरोग जानपदिक एवं सूचना विज्ञान संस्थान
ICAR-National Institute of Veterinary Epidemiology and Disease Informatics

रामगोडनहल्लि, येलहंका, बेंगलुरु - 560 064, भारत
Ramagondanahalli, Post Box No. 6450, Yelahanka, Bengaluru - 560 064, INDIA

Dr. Parimal Roy, M.V.Sc., Ph.D.
Director



Message

ICAR-NIVEDI, is a pioneer research institute under Indian Council of Agricultural Research (ICAR), has been entrusted to conduct Research & Development in the field of Veterinary Epidemiology and animal disease surveillance for the entire country. ICAR-NIVEDI plays an extremely pivotal role in developing models for risk analysis, animal disease forecasting, forewarning, need based diagnostics and economic impact analysis of the diseases.

Livestock plays an important role in rural livelihood of subsistence farmers dependent on the sector. Livestock disease informatics in collection, collation, analysis and dissemination of the results to stakeholders is important for prevention and control of livestock diseases. Capacity building in the field of disease informatics and bio-statistics for the SAARC member countries will be of immensely useful for the participants to contain diseases and also trans-boundary diseases of importance to the region. There is need for continuous exchange of knowledge between SAARC member countries in emergency preparedness of emerging and re-emerging livestock diseases of the region.

It is a matter of pride and responsibility for the Institute to host the SAARC sponsored training on 'Animal Disease Informatics and Biostatistics' for participants of SAARC member country from 3rd to 8th May, 2018 under the aegis of SAARC Agriculture centre (SAC), Bangladesh and hope that the deliberations in the program will be mutually beneficial to the participants as well as to the host faculty. This training will create an avenue for ICAR NIVEDI to deliver our experience in the field of disease informatics, statistical analysis, prevention and control of livestock diseases to the participants which will pave a better way to control the most economically important animal diseases within the SAARC region.

I wish the participants a pleasant stay and I wish the training a great success

(Parimal Roy)



ABOUT TRAINING PROGRAM

ICAR- National Institute of Veterinary Epidemiology and Disease Informatics (NIVEDI), Bengaluru, Karnataka, India is organizing training program on “Animal Disease Informatics and Biostatistics” from 3 to 8, May 2018, which is sponsored by SAARC Agriculture center, Dhaka, Bangladesh. The participants for the training program are from Bangladesh, Srilanka, Bhutan, Nepal and India. The objective of the training program is to impart the training to member nations in the domain of Animal Disease Informatics and Biostatistics, which is an important area that systematically studies science and technology needed for collecting, sharing, reporting, analyzing, and visualizing infectious disease data. The theme of the training program is appropriately chosen by the SAARC agriculture center considering the urgent need to strengthen this area in order to formulate effective regional disease control and prevention programs.

The current training program is designed out of expertise gained over a period of many years through the process of conceptualizing, developing and successfully managing the National Animal Disease Expert Referral System (NADRES) which is working model that epitomizes the significance of animal disease informatics. The training topics cover disease data collection/capture, data storage system, analytics, data presentation and forewarning or dissemination of disease information periodically to stakeholder. Data analysis starts with the collection of data followed by sorting and processing. Disease data presentation (using charts, maps and other methods) and analysis (spatial and temporal) forms an important part of this training program. The practical approach to train the participants to analyses the disease incidence data, cluster detection, spatial mapping by R software is provided.

In the present course, modules related to development of disease database, prediction and forewarning of disease, epidemiological study designs to understand the disease incidence and spread, biostatistical methods to analysis disease data and presentation of disease information in the form of spatial and temporal maps are designed along with hands on practical sessions. Overall the training program emphasizes more on disease data sources and collection strategies, data analysis and outbreak detection principles, data visualization, information dissemination and alerting, system assessment and evaluation. Hope this training program is immensely useful to the participants from SAARC countries with constructive exchange of knowledge and experience.



Sl. No.	Table of Contents	Page No.
1.	National Animal Disease Referral Expert System (NADRES)	1
2.	Overview of Epidemiological Concepts	8
3	Epidemiological study designs with special reference to Cross-sectional study	10
4	Research Methodology and Bio-statistics in Animal Science Research	13
5	Spatial Epidemiology: An introduction	22
6	Informatics on Protozoan Parasitic diseases in large ruminants	24
7	Geographical Epidemiology, Spatial Analysis and Livestock Disease Information System	31
8	Economic losses due to livestock diseases	35
9	Livestock Disease Diagnosis Informatics	37
10	Cohort studies in Epidemiology	40
11	An Introduction to case-control study in veterinary epidemiology	44
12	Spatial data analysis using R software - A Practical Approach	49
13	A Practical approach to calculate herd and animal level sample size	79
14	Bioinformatics applications in animal disease pathogens	84
15	Risk Factor identification for incidence of Brucellosis and its informatics	85
16	List of participants and organizing committee	92



National Animal Disease Referral Expert System (NADRES)

K.P.Suresh, D. Hemadri, SS Patil and Parimal Roy

ICAR-National Institute of Veterinary Epidemiology & Disease Informatics, Bengaluru, India

During 1987, The Indian Council of Agricultural Research (ICAR) established an All India Coordinated Research Project on Animal Disease Monitoring and Surveillance, (AICRP on ADMAS). On 1st April 2000, the AICRP on ADMAS was upgraded to Project Directorate on Animal Disease Monitoring and Surveillance (PD_ADMAS) (during IX Plan). The Directorate got further impetus with the addition of five more collaborating units in X plan and two mission mode NATP projects viz., Animal Health Information System and Data monitoring System (AHIS_DMS) and Weather based Animal Disease Forecasting (WB_ADF) having 17 and 20 collaborating units respectively. Combining the input from AHIS_DMS and WB_ADF, an interactive, dynamic online animal disease forewarning system called NADRES (National Animal Disease Referral Expert System) was developed with overall aim to improve the early warning and response capacity to animal disease threats in the country for the benefit of farmers. Presently NIVEDI is having 31 AICRP centres.

Early warning of disease incidence or outbreaks and the capacity of prediction of risk of spread to new areas is an essential pre-requisite for the effective containment and control of epidemic animal diseases, including zoonosis. Early warning is based on the concept that dealing with a disease epidemic in its early stages is easier and more economical than having to deal with it once it is wide spread. From the public health prospective, early warning of disease outbreaks with a known zoonotic potential will enable control measures that can prevent human morbidity and mortality. National Institute of Veterinary Epidemiology & Disease Informatics developed the software application, NADRES, that systematically collect, verify, analyse and respond to the information from designated AICRP-ADMAS, unofficial media reports and informal networks. NADRES builds on the added value combining the alert and response mechanisms of different organisations like state animal husbandry departments, Departments from universities, department of Animal husbandry, Dairying and Fisheries, AICRP on ADMAS and other agencies including NGOS, enhancing the capacity for the benefit of the farmers in the country and other stakeholders to assist in prediction, prevention and control of animal disease threats, including zoonosis, through sharing information, epidemiological analysis and joint missions to assess and control the outbreak, whenever needed. For Zoonotic disease events, alerts of animal outbreaks or incidence can provide the direct early warning so that human surveillance could be enhanced and preventive action can be taken. Similarly there may be cases where human surveillance is more sensitive and alerts of human cases precede known animal occurrence of disease. Sharing assessments of an outbreak will enable a joint and comprehensive analysis of the disease event and its possible consequences. Joint dissemination will furthermore allow harmonised communication by the Central and state Animal departments, ICAR-NIVEDI, regarding disease control strategies.

Regarding the joint response to disease emergencies, the three organisations will be able to respond to a larger number and cover a wider range of outbreaks or exceptional epidemiological events with the provision of a wider range of expertise. This will improve the national preparedness for epidemics and provide rapid, efficient and coordinated assistance in developing disease control strategies.

Specific Objectives of NADRES

- Allow state and central animal husbandry departments to better prepare themselves to prevent incursion of animal diseases/infection and enable their rapid containment.
- Increase timeliness and sensitivity of alerts
- Improve the detection of exceptional epidemiological events at country level

- Improve the transparency among different stakeholders
- Improve the national surveillance and monitoring systems and strengthen the networks of veterinary laboratories working in the country.
- Improve national preparedness for animal and zoonotic epidemics and provide rapid, efficient and coordinated assistance to states experiencing them.
- Provide the technical support to states on issues at the animal/human interface of outbreak control.

Disease Outbreak data base

Database on disease outbreaks were collected through the networks of AICRP on ADMAS with 31 centres across the country, provide the regular outbreak information along with date and location of outbreaks, susceptible population, deaths, attacks etc., Disease data obtained on a format is entered in to NADRES database in a double –data-entry validation mode to achieve to zero error entry. Database contains the disease events since 1990 was further improved by inclusion of additional 16 AICRP centres.

Risk factors database

Risk factors such as weather parameters from different sources includes the monthly precipitation(mm), sea level pressure (millibar), minimum temperature (°C) maximum temperature(°C) wind speed (m/s), vapour pressure (millibar), soil moisture(%) , perceptible water(mm), potential evaporation transpiration (mm), cloud cover(%) etc., extracted from National Centre for environmental prediction (NCEP), Indian Meteorological Department(IMD), National Innovations Climate Resilient Agriculture (NICRA) and other sources. The remote sensing variables like Normalised Difference vegetative index (NDVI) and Land Surface temperature were extracted from MODIS/LANDSAT/LISS III or IV satellite images. The livestock population and densities were extracted from Livestock census 2012.

Statistical Model

A multivariate logistic regression model was used to predict the probability disease risk in relation to weather parameters, remote sensing variables and livestock population or densities. The goal of logistic regression is to find the best fitting (yet biological reasonable) model to describe the relationship between dichotomous characteristic of interest (disease outbreak) and a set of predictors (weather parameters, RS variables and demographics) Logistic regression generates the co-efficients (and its standard error and significance level) of a formula to predict a logit transformation of the probability of presence of the characteristics of interest.

$$\text{logit}(p) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_k X_k$$

where p is the probability of presence of the characteristic of interest. The logit transformation is defined as the logged odds

$$\text{odds} = \frac{p}{1-p} = \frac{\text{probability of presence of characteristic}}{\text{probability of absence of characteristic}}$$

and

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$$

Rather than choosing parameters that minimize the sum of squared errors (like in ordinary regression), estimation in logistic regression chooses parameters that maximize the likelihood of observing the sample values.

Overall Fit of Model

The *null model* $-2 \text{ Log Likelihood}$ is given by $-2 * \ln(L_0)$ where L_0 is the likelihood of obtaining the observations if the independent variables had no effect on the outcome. The *full model* $-2 \text{ Log Likelihood}$ is given by $-2 * \ln(L)$ where L is the likelihood of obtaining the observations with all independent variables incorporated in the model. The difference of these two yields a Chi-Squared statistic which is a measure of how well the independent variables affect the outcome or dependent variable. If the P-value for the overall model fit statistic is less than the conventional 0.05 then there is evidence that at least one of the independent variables contributes to the prediction of the outcome. Cox & Snell R^2 and Nagelkerke R^2 are other goodness of fit measures known as pseudo R-squares. Note that Cox & Snell's pseudo R-squared has a maximum value that is not 1. Nagelkerke R^2 adjusts Cox & Snell's so that the range of possible values extends to 1.

Hosmer-Lemeshow test

The Hosmer-Lemeshow test is a statistical test for goodness of fit for the logistic regression model. The data are divided into approximately ten groups defined by increasing order of estimated risk. The observed and expected number of cases in each group is calculated and a Chi-squared statistic is calculated as follows:

$$\chi^2_{HL} = \sum_{g=1}^G \frac{(O_g - E_g)^2}{E_g(1 - E_g/n_g)}$$

with O_g , E_g and n_g the observed events, expected events and number of observations for the g^{th} risk decile group, and G the number of groups. The test statistic follows a Chi-squared distribution with $G-2$ degrees of freedom.

A large value of Chi-squared (with small p-value < 0.05) indicates poor fit and small Chi-squared values (with larger p-value closer to 1) indicate a good logistic regression model fit.

Early Warning System (EWS)

Early identification of an infectious disease outbreak is an important first step towards implementing effective disease interventions and reducing resulting mortality and morbidity. The geographic and seasonal distribution of many infectious diseases are associated with climate and therefore the possibility of using seasonal climate forecasts as predictive indicators in disease early warning system (EWS) is an interest of focus. Geographic Information system(GIS), remote sensing (RS) and Global Positioning system(GPS) are the three commonly used veterinary geo-informatics technologies employed in this digital era for rapid communication of data for better management of animal diseases.

Early warning systems are combinations of tools and process embedded within institutional structures coordinated by national or international agencies. These systems are composed of four elements depending upon they focus on specific hazard or many, namely, knowledge of risk, a technical monitoring and warning services, dissemination of meaningful warnings to at-risk areas, and farmers awareness and preparedness to act. Warning services lie at the core of these systems, and how well they operate depends on having a sound scientific basis for predicting and forecasting. As early warning systems grow in geographical coverage and sophistication, false alarms to in rise. High false alarms can undermine the public confidence, breed mistrust, dilute the impact of alerts and reduce the credibility of future warnings.

Classification table

The classification table is another method to evaluate the predictive accuracy of the logistic regression model.

In this table the observed values for the dependent outcome and the predicted values (at a user defined cut-off value, for example $p=0.50$) are cross-classified and provides the accuracy index for assessment model performance in prediction.

Table 1 : Accuracy of Prediction

Sl. No.	Diseases	Accuracy (%)
1	Anthrax	95.52
2	Babesiosis	96.14
3	Black Quarter	92.43
4	Bluetongue	97.99
5	Enterotoxemia	95.98
6	Fasciolosis	97.06
7	Foot and mouth disease	86.72
8	Haemorrhagic septicaemia	90.58
9	Peste des petits ruminants	90.12
10	Sheep & Goat pox	95.52
11	Swine fever	95.37
12	Theileriosis	96.60
13	Trypanosomosis	96.45

Internal Accuracy was performed using 10 years of data. Accuracy obtained was $> 90\%$ except Foot and mouth disease (86.72%).

The probability of disease outbreak was categorised in 6 risk levels- No risk (NR), Very low risk (VLR), Low risk (LR), Moderate risk (MR), High risk (HR) and Very high risk (VHR) for enabling the stake holders to take appropriate control measures by suitably allocating available resources.

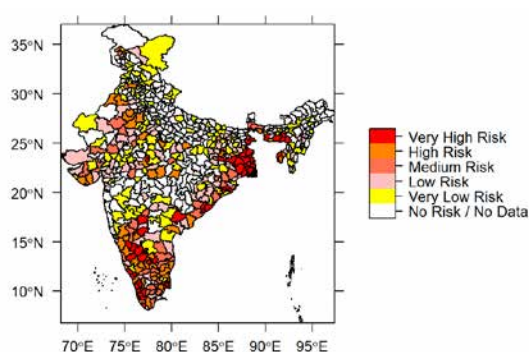


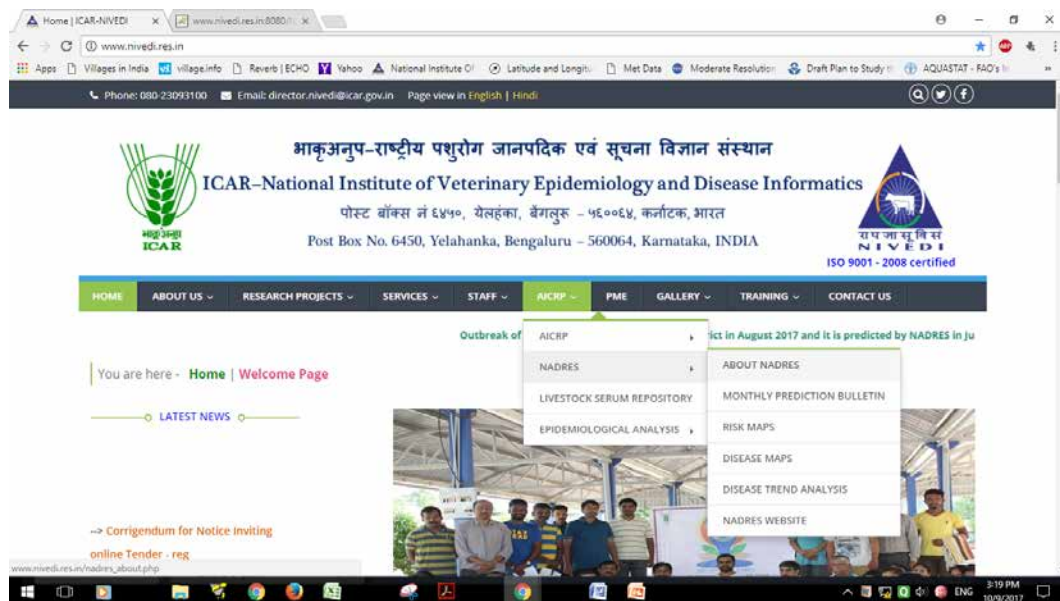
Fig. 1 Risk prediction of Anthrax for the month of June 2018

Table 2 : Probability distribution of risk interpretations.

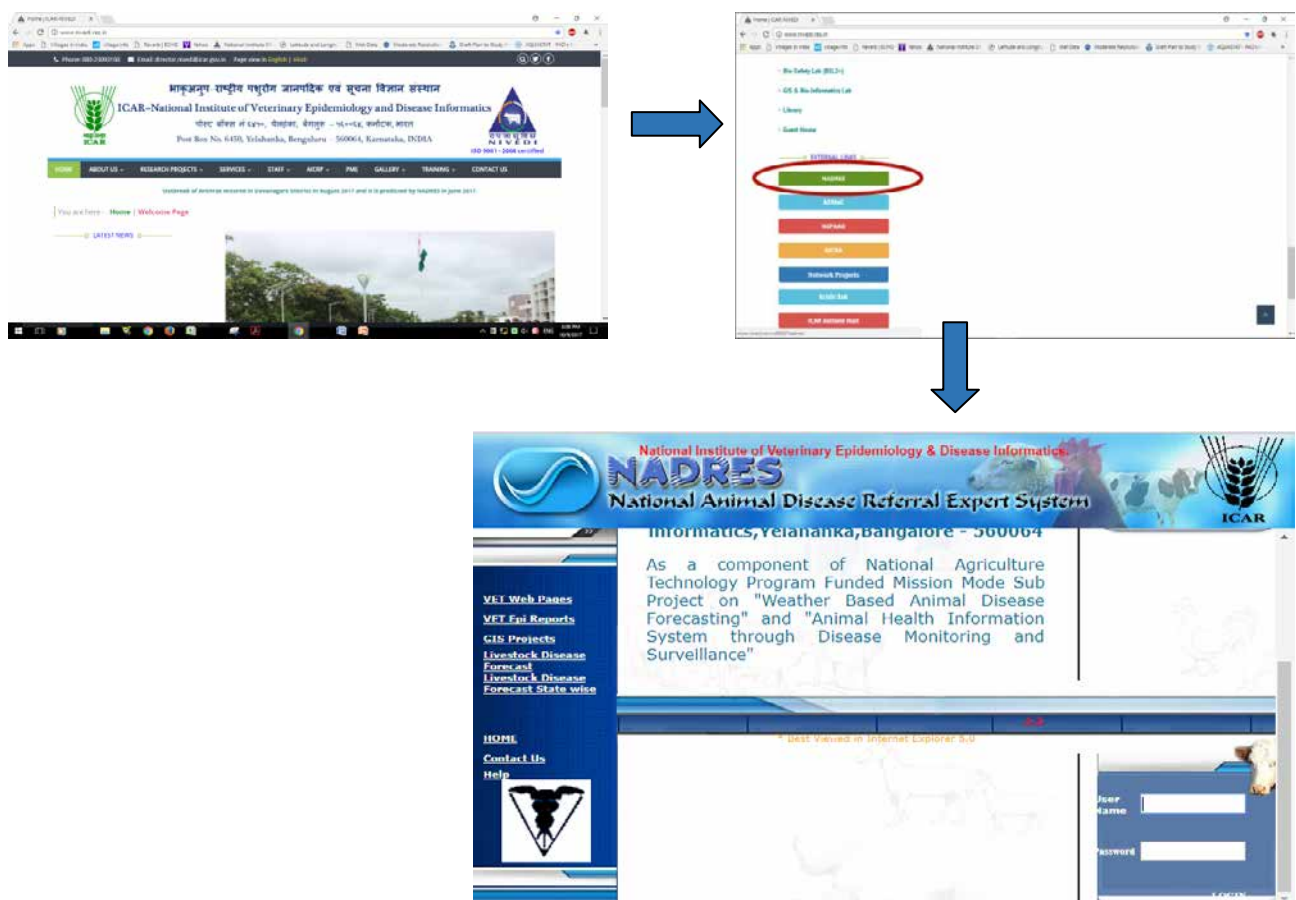
Sl. No.	Probability of risk	Interpretation
1	0	No risk/No or inadequate data
2	0-0.20	Very low risk
3	0.21-0.40	Low risk
4	0.41-0.60	Moderate risk
5	0.61-0.80	High risk
6	0.8-1.0	Very high risk

NADRES Outlook:

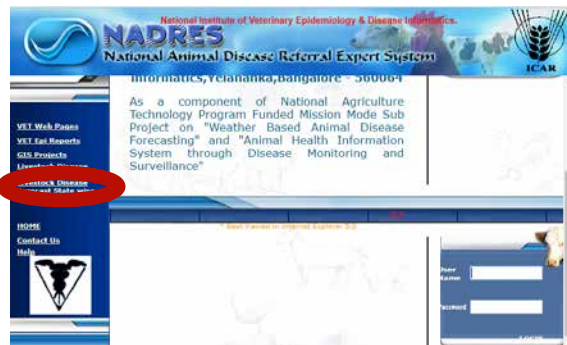
The ICAR-NIVEDI website consists of details regarding NADRES database, Monthly prediction bulletin, Risk maps, disease maps and Disease trend analysis.



Accessing the risk prediction at NADRES web page



Livestock disease forecast-State wise




Institute of Veterinary Epidemiology & Disease Informatics. NADRES National Animal Disease Referral Expert System ICAR

Anthrax Livestock Disease Forecast of December month in KARNATAKA

District Name	Cattle	Buaffalo	Goat	Sheep	Pig	Month	Result
Bangalore Rural	154626	22878	79008	121032	5889	December	High Risk
Bellary	325054	128275	175849	732810	8969	December	Very High Risk
Chamarajanagar	253838	20128	98721	124934	619	December	High Risk
Davanagere	318204	167544	99155	327631	782	December	Very High Risk
Koppal	256060	74788	153857	539842	9008	December	High Risk
Raichur	346798	128316	266878	626439	8568	December	Medium Risk
Tumkur	515530	175071	319156	1049268	5392	December	Very High Risk

Livestock disease forecast-District wise




NADRES National Animal Disease Referral Expert System ICAR

NADRES Forecast Results

Disease Name	State Name	District Name	Month	Result
Anthrax	KARNATAKA	Bangalore Rural	December	High Risk

District Livestock Population

State Name	District Name	Cattle	Buaffalo	Goat	Sheep	Pig
KARNATAKA	Bangalore Rural	154626	22878	79008	121032	5889

LDF -Mobile Application

To extend the reach of the NADRES forewarning bulletin among the various stakeholders, a Mobile Application (*app*) “LDF-Mobile App” was developed. The forewarning methodology adapted in the “mobile app” remains the same as monthly bulletin. In addition to forewarning, the LDF-Mobile App also provides the details of clinical samples to be collected in case of outbreaks of the listed diseases for laboratory confirmation. Immediate preventive measures to be taken up in case of positive prediction/disease confirmation. The LDF mobile app is available at ICAR-NIVEDI website. It will also be made available on Google play store.



Fig. 2 Mobile screen containing LDF app



Fig.3 LDF app logo



Fig.4 LDF app home page

Overview of Epidemiological Concepts

S.S. Patil & K.P. Suresh

ICAR-National Institute of Veterinary Epidemiology & Disease Informatics, Bengaluru, India

The present day veterinary profession is confronted with a different set of problems compared with the earlier period due to globalization and liberalization of the trade and commerce. Now, veterinarians often have to deal with regions that are either endemic or epidemic to many livestock diseases. In addition, it is considered necessary to take into account the economic aspects of disease control through the use of benefit/cost analyses of disease control campaigns. Veterinarians have to respond to the challenges posed by emerging and re-emerging infectious diseases. In order to address such issues, it is necessary to identify and quantify the disease determinants and establish if there is a direct or indirect causal relation to the disease condition.

Epidemiology is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems (CDC). Veterinary epidemiology deals with study of animal disease distribution and identification of disease determinants for given population in order to develop disease control strategies. The epidemiological study involves collection, analysis and interpretation of data that will aid in identifying disease frequency and pattern for a population with defined geography and time. Veterinary epidemiology is a holistic approach aimed at co-ordinating the use of different scientific disciplines and techniques during an investigation of disease or impaired productivity or welfare. There are four approaches to epidemiological investigations viz., Descriptive epidemiology, analytical epidemiology, experimental epidemiology and theoretical epidemiology.

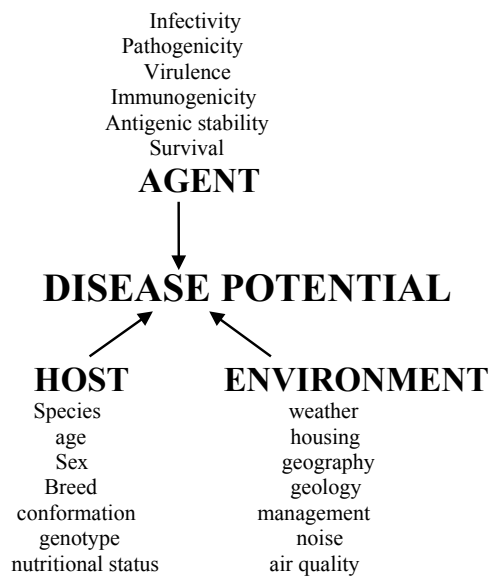


Fig. 1: Epidemiological Triad

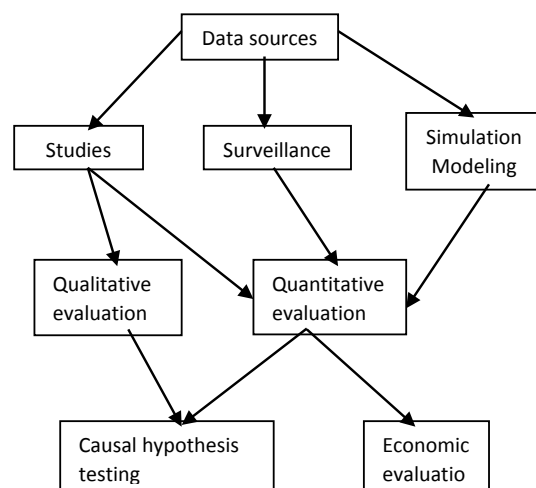


Fig. 2: Epidemiological concepts

Table. 1: Comparison of measures of disease occurrence

	Incidence density	Cumulative Incidence	Prevalence
Numerator	New cases occurring during a period of time initially free of diseases	New cases occurring during a period of time among a group initially free of disease	All cases counted on single survey a group
Denominator	Sum of time periods during which individuals could have developed disease	All susceptible individuals present of the period	All individuals examined, including cases and non-cases
Time	For each individual from beginning of follow-up until disease	Duration of period	Single point or a period
How measured	Prospective cohort study	Prospective cohort study	Cross-sectional study
Interpretation	Rapidity with which new cases develop over given time period	Risk of developing disease over period	Probability of having disease at a particular point in time

References

1. Michael Thrusfield: Veterinary Epidemiology, Third Edition, Wiley
2. Dirk U Pfeiffer: Veterinary Epidemiology-An Introduction, University of London

Epidemiological study designs with special reference to Cross-sectional study

Dr. Senthil Amudhan R

Department of Epidemiology, NIMHANS, Bengaluru, India

Study Design : Overview

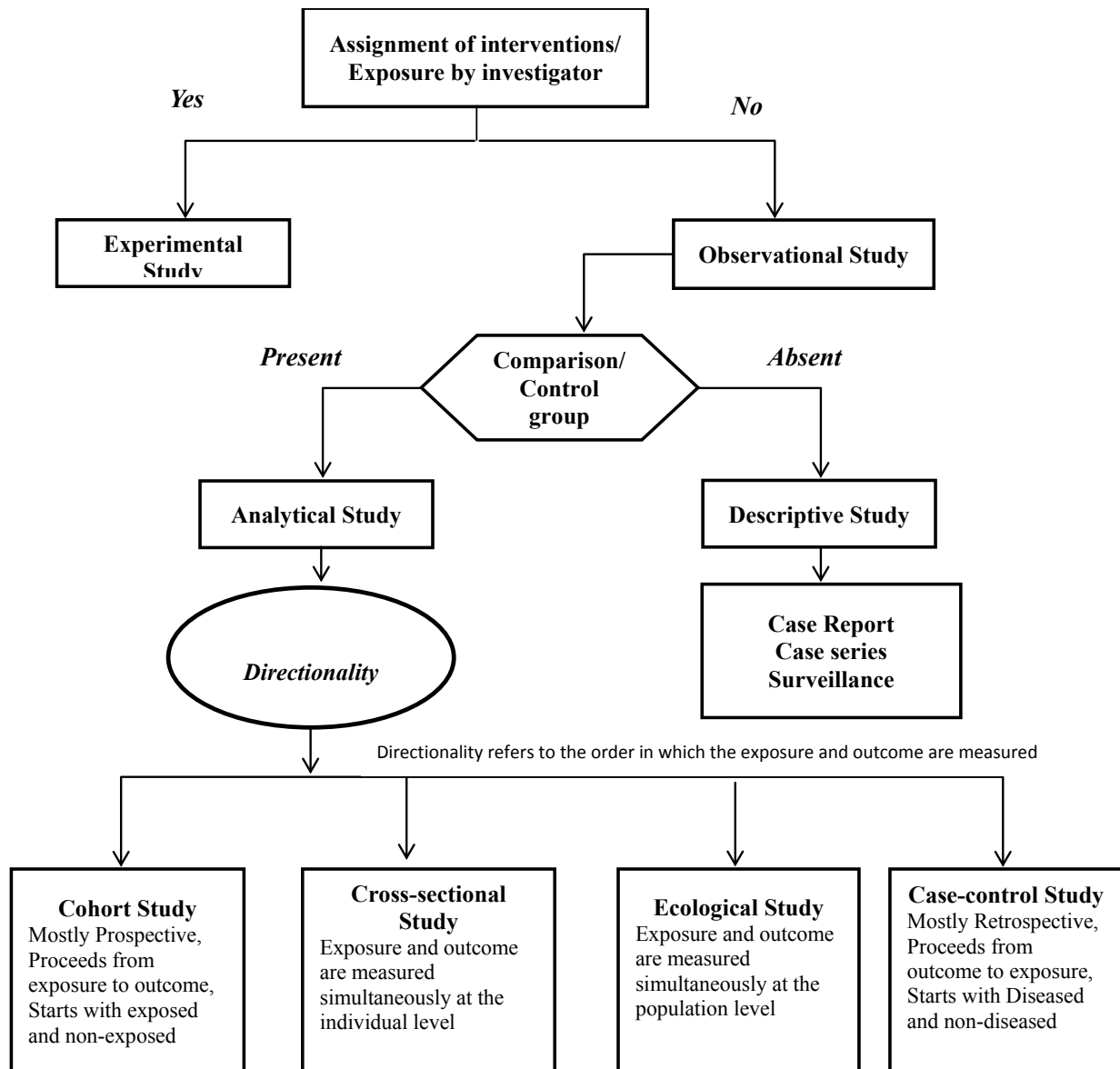


Fig. 1. Cross-sectional studies

Due to ease of implementation, cross-sectional studies are one of the most frequently chosen study designs in veterinary epidemiology. In this study, the investigator measures both exposure and outcome variables on a single occasion in a defined population (usually selected randomly) at a single point in time with no follow-up period. Thus, the outcome frequency measure is inherently prevalence. This makes them fast and inexpensive. They are well suited to estimate the prevalence (burden) and describing the characteristics or distribution pattern of disease/risk factors either in the community or hospital setting. This can be a descriptive study (without comparison group) or an analytical study (with comparison group). As the exposure and outcome

variables are measured simultaneously, temporal sequence (exposure preceded the outcome) and incidence cannot be assessed thus limiting it to provide information on disease causation and prognosis.

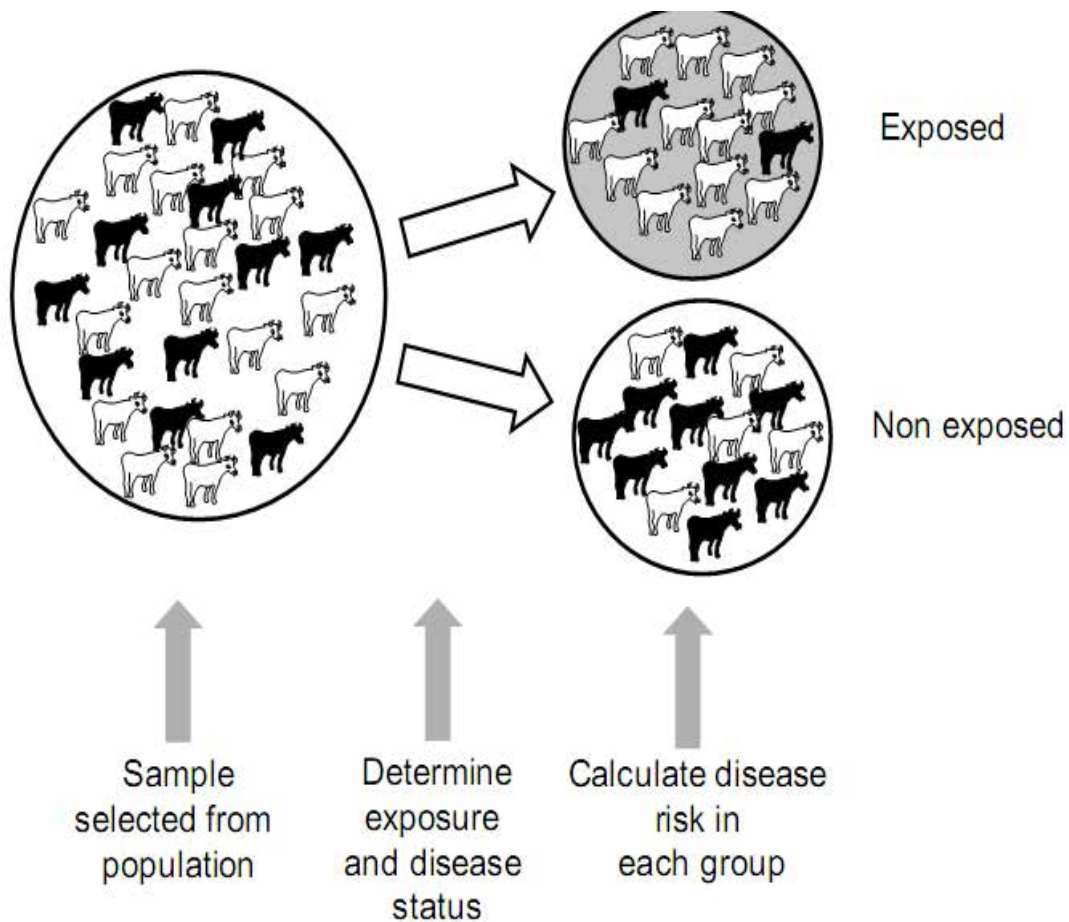


Fig. 2. Single data collection point

Descriptive cross-sectional studies are useful for planning or administering preventive or health care services, surveillance programs, and surveys and polls.

The potential drawback to this study design is that often the search for potential causes is not very focused and thus, a lot of data-mining for significant factors is used in the analysis stage.

The two major limitations of a cross-sectional study design: provides only a “snapshot in time” of the disease occurrence with Prevalence as an outcome and the reverse-causation problem (This becomes problematic for time-varying risk factors, as their effect cannot be correctly measured.)

Table 1 : Attributes of Cross-sectional

Attribute	Cross-Sectional
<i>Assignment of intervention / exposure</i>	No
<i>Directionality</i>	Outcome and exposure measured simultaneously
<i>Study group</i>	Exposed or diseased
<i>Comparison group</i>	Non-exposed or non-diseased
<i>Temporal sequence</i>	Hard to establish
<i>Multiple associations</i>	Can study multiple exposures and outcome
<i>Time and money</i>	Less expensive
<i>Sample size</i>	Large or small
<i>Measures of effect and association</i>	Prevalence Prevalence odds ratio Prevalence ratio
<i>Best when</i>	Onset of disease is prolonged
<i>Evidence of causality</i>	Only suggestive
<i>Biases</i>	Difficult to manage
<i>Other issues</i>	Cannot measure incidence of disease

(Adapted from: Kalaivani M, Amudhan S, Upadhyay AD, Kamal VK. Biostatistics. In: Hemanshu Prabhakar (ed): Essentials of Neuroanesthesia. London Wall, Academic Press (an imprint of Elsevier) 2017.p. 975-995.)

References :

1. Kalaivani M, Amudhan S, Upadhyay AD, Kamal VK. Biostatistics. In: Hemanshu Prabhakar (ed): Essentials of Neuroanesthesia. London Wall, Academic Press (an imprint of Elsevier) 2017.p. 975-995
2. Pfeiffer D. Veterinary Epidemiology: An Introduction. John Wiley & Sons; 2010. 151 p.
3. Dohoo I, Martin W, Stryhn H. Veterinary Epidemiologic Research. Prince Edward Island: University of Prince Edward Island; 2003. 704 p.

Research Methodology and Bio-statistics in Animal Science Research

K.P.Suresh, Ph.D (Biostatistics),

Principal Scientist, ICAR-National Institute of Veterinary Epidemiology & Disease Informatics

Introduction: The efficacy, safety and economics of investigational products should be demonstrated by research which follow the guidance in 'Good Research Practice(GRP): The role of statistics in biological research design and analysis is acknowledged as essential in the ICH (Intentional Conference on Harmonization) guidelines. The proliferation of statistical research in the area of animal science research coupled with the critical role of research in the treatment approval process necessitate a succinct document on statistical issues related to research. This paper is intended to give direction to researchers in the design, conduct, analysis, and evaluation of research of an investigational product(IP) in the context of its overall development. The document will also assist scientific experts charged with preparing application summaries or assessing evidence of efficacy, economics and safety, principally from research in later phases of development.

PROTOCOL

A protocol is a study plan on which all biological research is based. The plan is carefully designed to safeguard the health of the participants as well as answer specific research questions. A protocol describes what types of people or animal may participate in the research; the schedule of tests, procedures, treatments medications, and dosages; and the length of the study. While in a biological research, participants following a protocol are seen regularly by the research staff to monitor their health and to determine the safety, economics and effectiveness of their treatment.

1. Development Plan: The broad aim of the process of clinical development of a new investigational product (IP) is to find out whether there is a dose range and schedule at which the IP can be shown to be simultaneously safe and effective, to the extent that the risk-benefit relationship is acceptable. The particular subjects who may benefit from the IP, and the specific indications for its use, also need to be defined in the protocol. Satisfying these broad aims usually requires an ordered Programme of biological research, each with its own specific objectives. This should be specified in a research plan, or a series of plans, with appropriate decision points and flexibility to allow modification as knowledge accumulates. Interpretation and assessment of the evidence from the total programme of research involves synthesis of the evidence from the individual research. This is facilitated by ensuring that common standards are adopted for a number of features of the research such as dictionaries of terminologies, definition and timing of the main measurements, handling of protocol deviations and so on. A statistical summary, overview or meta-analysis may be informative when research questions are addressed in more than one research. Wherever possible this should be envisaged in the plan so that the relevant research studies are clearly identified and any necessary common features of their designs are specified in advance. Other major statistical issues (if any) that are expected to affect a number of research in a common plan should be addressed.

2. Confirmatory Research: A confirmatory research is an adequately controlled research in which the hypotheses are stated in advance and evaluated. As a rule, confirmatory research are necessary to provide firm evidence of efficacy or safety. In such research the key hypothesis of interest follows directly from the research's primary objective, is always pre-defined, and is the hypothesis that is subsequently tested when the research is complete. In a confirmatory research it is equally important to estimate with due precision the size of the effects attributable to the treatment of interest and to relate these effects to their biological significance. Confirmatory research is intended to provide firm evidence in support of claims and hence adherence to protocols and standard operating procedures is particularly important; unavoidable changes should be explained and documented.

3. Exploratory Research: The rationale and design of confirmatory research nearly always rests on earlier research work carried out in a series of exploratory studies. Like all clinical research, these exploratory studies should have clear and precise objectives. However, in contrast to confirmatory research, their objectives may not always lead to simple tests of pre-defined hypotheses. In addition, exploratory research may sometimes require a more flexible approach to design so that changes can be made in response to accumulating results. Their analysis may entail data exploration; tests of hypothesis may be carried out, but the choice of hypothesis may be data dependent. Such research cannot be the basis of the formal proof of efficacy, although they may contribute to the total body of relevant evidence.

SCOPE OF RESEARCH

1. Population: a collection of data whose properties are analyzed. The population is the *complete* collection to be studied; it contains *all* subjects of interest.

2. Sample: a *part* of the population of interest, a sub-collection selected from a population. A large population may be impractical and costly to study, collecting data from every member of the population. A sample is more manageable and easier to study. After collecting and organizing the data, a summary is made such as average values. Hopefully valid conclusions can be made on the whole population based on the sample data. Therefore, it is important that the sample data collected be representative of the population. Otherwise conclusions may be invalid. Conclusions are only as reliable as the sampling process, and information can change from sample to sample. Non-random sample will be threat to the Internal validity of the study

3. Parameters: a numerical measurement that describes a characteristic of a population, while a statistics is a numerical measurement that describes a characteristic of a sample. In general, we will use a **statistic** to infer something about a parameter.

4. Primary and Secondary Variables: The primary variable ('target' variable, primary endpoint) should be the variable capable of providing the most biologically relevant and convincing evidence directly related to the primary objective of the research. There should generally be only one primary variable. This will usually be an efficacy variable because the primary objective of most confirmatory research is to provide strong scientific evidence regarding efficacy. Safety /economics/ tolerability may sometimes be the primary variable and will always be an important consideration. Measurements relating to quality of life and health economics are further potential primary variables. The selection of the primary variable should reflect the accepted norms and standards in the relevant field of research. The use of a reliable and validated variable with which experience has been gained either in earlier studies or in published literature is recommended. There should be sufficient evidence that the primary variable can provide a valid and reliable measure of some biologically relevant and important treatment benefit in the population described by the inclusion and exclusion criteria. The primary variable should be used when estimating the sample size.

Secondary variables are either supportive measurements related to the primary objective or measurements of effects related to the secondary objectives. Their pre- definition in the protocol is also important, as well as an explanation of their relative importance and roles in interpretation of research results. The number of secondary variables should be limited and should be related to the limited number of questions to be answered in the research.

5. Composite Variables: If a single primary variable cannot be selected from multiple measurements associated with the primary objective, another useful strategy is to integrate or combine the multiple measurements into a single or 'composite' variable, using a pre- defined algorithm. This approach addresses the multiplicity problem without requiring adjustment to the type I error. The method of combining the multiple measurements should be specified in the protocol, and an interpretation of the resulting scale should be provided in terms of the size of a biologically relevant benefit.

6. Global Assessment Variables: In some cases, 'global assessment' variables are developed to measure the overall safety, overall efficacy, and/or overall usefulness of a treatment product or Investigational product (IP). This type of variable integrates objective variables and the investigator's overall impression about the state or change in the state of the subject, and is usually a scale of ordered categorical ratings.

7. Multiple Primary Variables: It may sometimes be desirable to use more than one primary variable, each of which (or a subset of which) could be sufficient to cover the range of effects of the therapies. The planned manner of interpretation of this type of evidence should be carefully spelled out. It should be clear whether an impact on any of the variables, some minimum number of them, or all of them, would be considered necessary to achieve the research objectives. The primary hypothesis or hypotheses and parameters of interest (e.g. mean, percentage, and distribution) should be clearly stated with respect to the primary variables identified, and the approach to statistical inference described. The effect on the type I error should be explained because of the potential for multiplicity problems, the method of controlling type I error

should be given in the protocol. The extent of inter correlation among the proposed primary variables may be considered in evaluating the impact on type I error. If the purpose of the research is to demonstrate effects on all of the designated primary variables, then there is no need for adjustment of the type I error, but the impact on type II error and sample size should be carefully considered.

8. Surrogate variables: When direct assessment of the relevant benefit to the subject through observing actual efficacy is not practical, indirect criteria (surrogate variables) may be considered. Commonly accepted surrogate variables are used in a number of indications where they are believed to be reliable predictors of treatment benefit.

9. Categorized Variables: Dichotomization or other categorization of continuous or ordinal variables may sometimes be desirable. Criteria of 'success' and 'response' are common examples of dichotomies which require precise specification in terms of, for example, a minimum percentage improvement (relative to baseline) in a continuous variable, or a ranking categorized as at or above some threshold level (e.g., 'good') on an ordinal rating scale. The reduction of diastolic blood pressure below 90mmHg is a common dichotomization. Categorizations are most useful when they have clear clinical relevance. The criteria for categorization should be pre-defined and specified in the protocol, as knowledge of research results could easily bias the choice of such criteria. Because categorization normally implies a loss of information, a consequence will be a loss of power in the analysis; this should be accounted for in the sample size calculation.

DESIGN TECHNIQUES TO AVOID BIAS

The most important design techniques for avoiding bias in research studies are blinding and randomization and these should be normal features of most controlled research studies intended to be included in a marketing application. Most such research studies follow a blinded approach in which treatments are pre-packed in accordance with a suitable randomization schedule, and supplied to the research centre(s) labeled only with the subject number and the treatment period so that no one involved in the conduct of the research is aware of the specific treatment allocated to any particular subject, not even as a code letter.

Bias can also be reduced at the design stage by specifying procedures in the protocol aimed at minimizing any anticipated irregularities in research conduct that might impair a satisfactory analysis, including various types of protocol violations, withdrawals and missing values. The protocol should consider ways both to reduce the frequency of such problems, and also to handle the problems that do occur in the analysis of data.

1. Blinding: Blinding or masking is intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a biological research results arising from the influence which the knowledge of treatment may have on the recruitment and allocation of subjects, their subsequent care, the attitudes of subjects /animal care takers to the treatments, the assessment of end-points, the handling of withdrawals, the exclusion of data from analysis, and so on. The essential aim is to prevent identification of the treatments until all such opportunities for bias have passed.

2. Randomization: Randomization introduces a deliberate element of chance into the assignment of treatments to subjects in a research. During subsequent analysis of the research data, it provides a sound statistical basis for the quantitative evaluation of the evidence relating to treatment effects. It also tends to produce treatment groups in which the distributions of prognostic factors, known and unknown, are similar. In combination with blinding, randomization helps to avoid possible bias in the selection and allocation of subjects arising from the predictability of treatment assignments.

DESIGN CONFIGURATIONS

1. Parallel Group Design: The most common research design for confirmatory research is the parallel group design in which subjects are randomized to one of two or more arms, each arm being allocated a different treatment. These treatments will include the investigational product at one or more doses, and one or more control treatments, such as placebo and/or an active comparator. The assumptions underlying this design are less complex than for most other designs. However, as with other designs, there may be additional features of the research that complicate the analysis and interpretation (e.g. covariates, repeated measurements over time, and interactions between design factors, protocol violations, dropouts and withdrawals).

2. **Crossover Design:** In the crossover design, each subject is randomized to a sequence of two or more treatments, and hence acts as his own control for treatment comparisons. This simple manoeuvre is attractive primarily because it reduces the number of subjects and usually the number of assessments needed to achieve a specific power. In the simplest 2×2 crossover design each subject receives each of two treatments in randomized order in two successive treatment periods, often separated by a washout period.

3. **Factorial Designs:** In a factorial design two or more treatments are evaluated simultaneously through the use of varying combinations of the treatments. The simplest example is the 2×2 factorial design in which subjects are randomly allocated to one of the four possible combinations of two treatments, A and B say. These are: A alone; B alone; both A and B; neither A nor B. In many cases this design is used for the specific purpose of examining the interaction of A and B.

4. **Multicentre Research:** a research may be designed as a multicentre (and multi-investigator) research primarily to provide a better basis for the subsequent generalization of its findings. This arises from the possibility of recruiting the subjects from a wider population and of administering the treatment in a broader range of research settings.

TYPES OF COMPARISON

1. **Research to Show Superiority:** Scientifically, efficacy is most convincingly established by demonstrating superiority to placebo in a placebo-controlled research, by showing superiority to an active control treatment or by demonstrating a dose-response relationship. This type of research is referred to as a 'superiority' research.

2. **Research to Show Equivalence or Non-inferiority:** In some cases, an investigational product is compared to a reference treatment without the objective of showing superiority. This type of research is divided into two major categories according to its objective; one is an 'equivalence' research and the other is a 'non-inferiority' research. Concluding equivalence or non-inferiority based on observing a non-significant test result of the null hypothesis that there is no difference between the investigational product and the active comparator is inappropriate.

SAMPLE SIZE ESTIMATION

The number of subjects in a research should always be large enough to provide a reliable answer to the questions addressed. This number is usually determined by the primary objective of the research. If the sample size is determined on some other basis, then this should be made clear and justified. Using the usual method for determining the appropriate sample size, the following items should be specified: a primary variable, the test statistic, the null hypothesis, the alternative ('working') hypothesis or difference of treatment, the probability of erroneously rejecting the null hypothesis (the type I error), and the probability of erroneously failing to reject the null hypothesis (the type II error), as well as the approach to dealing with treatment withdrawals and protocol violations.

STATISTICAL ANALYSIS PLAN (SAP)

A statistical analysis plan is a document that contains a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes detailed procedures for executing the statistical analysis of the primary and secondary variables and other data.

Analysis sets: The set of subjects whose data are to be included in the main analyses should be defined in the statistical section of the protocol.

1. Intention-To-Treat Principle or Full analysis set

The principle that asserts that the effect of a treatment policy can be best assessed by evaluating on the basis of the intention to treat a subject (i.e. the planned treatment regimen) rather than the actual treatment given. It has the consequence that subjects allocated to a treatment group should be followed up, assessed and analyzed as members of that group irrespective of their compliance to the planned course of treatment.

2. Per Protocol Set (Valid Cases, Efficacy Sample, Evaluable Subjects Sample)

The set of data generated by the subset of subjects who complied with the protocol sufficiently to ensure that these data would be likely to exhibit the effects of treatment, according to the underlying scientific model. Compliance covers such considerations as exposure to treatment, availability of measurements and absence of major protocol violations.

The full analysis set and the per protocol set play different roles in superiority research (which seek to show the investigational product to be superior), and in equivalence or non-inferiority research studies (which seek to show the investigational product to be comparable). In superiority research the full analysis set is used in the primary analysis (apart from exceptional circumstances) because it tends to avoid over-optimistic estimates of efficacy resulting from a per protocol analysis, since the non-compliers included in the full analysis set will generally diminish the estimated treatment effect. However, in an equivalence or non-inferiority research studies use of the full analysis set is generally not conservative and its role should be considered very carefully.

3. Missing Values and Outliers: Missing values represent a potential source of bias in a research. Hence, every effort should be undertaken to fulfill all the requirements of the protocol concerning the collection and management of data.

4. Data Transformation: The decision to transform key variables prior to analysis is best made during the design of the research on the basis of similar data from earlier research. Transformations (e.g. square root, logarithm, Box Cox Power, Angular) should be specified in the protocol and a rationale provided, especially for the primary variable(s).

5. Estimation, Confidence Intervals and Hypothesis Testing.

The statistical section of the protocol should specify the hypotheses that are to be tested and/or the treatment effects which are to be estimated in order to satisfy the primary objectives of the research. The statistical methods to be used to accomplish these tasks should be described for the primary (and preferably the secondary) variables, and the underlying statistical model should be made clear. Estimates of treatment effects should be accompanied by confidence intervals, whenever possible, and the way in which these will be calculated should be identified. A description should be given of any intentions to use baseline data to improve precision or to adjust estimates for potential baseline differences, for example by means of analysis of covariance.

6. Subgroups, Interactions and Covariates: The primary variable(s) is often systematically related to other influences apart from treatment. For example, there may be relationships to covariates such as age and sex, or there may be differences between specific subgroups of subjects such as those treated at the different centers of a multicentre research. In some instances an adjustment for the influence of covariates or for subgroup effects is an integral part of the planned analysis and hence should be set out in the protocol.

7. Integrity of Data and Computer Software Validity: The credibility of the numerical results of the analysis depends on the quality and validity of the methods and software (both internally and externally written) used both for data management (data entry, storage, verification, correction and retrieval) and also for processing the data statistically. Data management activities should therefore be based on thorough and effective standard operating procedures. The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.

REPORTING

Descriptive statistics form an indispensable part of reports. Suitable tables and/or graphical presentations should illustrate clearly the important features of the primary and secondary variables and of key prognostic and baseline variables. The results of the main analyses relating to the objectives of the research should be the subject of particularly careful descriptive presentation. When reporting the results of significance tests, precise p-values (e.g. $p=0.044$) should be reported rather than making exclusive reference to critical values.

BIOSTATISTICAL METHODS FOR HYPOTHESIS TESTING

Introduction

With inferential statistics, you are trying to reach conclusions that extend beyond the immediate data alone. For instance, we use inferential statistics to try to infer from the sample data what the population might think. Or, we use inferential statistics to make judgments of the probability that an observed difference between groups is a dependable one or one that might have happened by chance in this study. Thus, we use inferential statistics to make inferences from our data to more general conditions; we use descriptive statistics simply to describe what's going on in our data.

Here, we concentrate on inferential statistics that are useful in experimental and quasi-experimental research design or in program outcome evaluation. Perhaps one of the simplest inferential test is used when you want to compare the average performance of two groups on a single measure to see if there is a difference. You might want to know whether eighth-grade boys and girls differ in math test scores or whether a program group differs on the outcome measure from a control group. Whenever you wish to compare the average performance between two groups you should consider the t-test for differences between groups.

STATISTICAL INFERENCE

A **statistical hypothesis** is an assumption about a population parameter. This assumption may or may not be true. The best way to determine whether a statistical hypothesis is true would be to examine the entire population. Since that is often impractical, researchers typically examine a random sample from the population. If sample data are not consistent with the statistical hypothesis, the hypothesis is rejected.

There are two types of statistical hypotheses.

- **Null hypothesis.** The null hypothesis, denoted by H_0 , is usually the hypothesis that sample observations result purely from chance.
- **Alternative hypothesis.** The alternative hypothesis, denoted by H_1 or H_a , is the hypothesis that sample observations are influenced by some non-random cause.

For example, suppose we wanted to determine whether a coin was fair and balanced. A null hypothesis might be that half the flips would result in Heads and half, in Tails. The alternative hypothesis might be that the number of Heads and Tails would be very different. Symbolically, these hypotheses would be expressed as

$$H_0: P = 0.5$$

$$H_a: P \neq 0.5$$

Suppose we flipped the coin 50 times, resulting in 40 Heads and 10 Tails. Given this result, we would be inclined to reject the null hypothesis. We would conclude, based on the evidence, that the coin was probably not fair and balanced.

Hypothesis tests:

Statisticians follow a formal process to determine whether to reject a null hypothesis, based on sample data. This process, called **hypothesis testing**, consists of four steps.

- State the hypotheses. This involves stating the null and alternative hypotheses. The hypotheses are stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false.
- Formulate an analysis plan. The analysis plan describes how to use sample data to evaluate the null hypothesis. The evaluation often focuses around a single test statistic.
- Analyze sample data. Find the value of the test statistic (mean score, proportion, t-score, z-score, etc.) described in the analysis plan.
- Interpret results. Apply the decision rule described in the analysis plan. If the value of the test statistic is unlikely, based on the null hypothesis, reject the null hypothesis.

Decision Errors:

Two types of errors can result from a hypothesis test.

- **Type I error.** A Type I error occurs when the researcher rejects a null hypothesis when it is true. The probability of committing a Type I error is called the **significance level**. This probability is also called **alpha**, and is often denoted by α .
- **Type II error.** A Type II error occurs when the researcher fails to reject a null hypothesis that is false. The probability of committing a Type II error is called **Beta**, and is often denoted by β . The probability of *not* committing a Type II error is called the **Power** of the test.

The analysis plan includes decision rules for rejecting the null hypothesis. In practice, statisticians describe these decision rules in two ways - with reference to a P-value or with reference to a region of acceptance.

- **P-value.** The strength of evidence in support of a null hypothesis is measured by the **P-value**. Suppose the test statistic is equal to S . The P-value is the probability of observing a test statistic as extreme as S , assuming the null hypothesis is true. If the P-value is less than the significance level, we reject the null hypothesis.
- **Region of acceptance.** The **region of acceptance** is a range of values. If the test statistic falls within the region of acceptance, the null hypothesis is not rejected. The region of acceptance is defined so that the chance of making a Type I error is equal to the significance level.
- The set of values outside the region of acceptance is called the **region of rejection**. If the test statistic falls within the region of rejection, the null hypothesis is rejected. In such cases, we say that the hypothesis has been rejected at the α level of significance.

One tailed and two-tailed tests

A test of a statistical hypothesis, where the region of rejection is on only one side of the sampling distribution, is called a **one-tailed test**. For example, suppose the null hypothesis states that the mean is less than or equal to 10. The alternative hypothesis would be that the mean is greater than 10. The region of rejection would consist of a range of numbers located on the right side of sampling distribution; that is, a set of numbers greater than 10.

A test of a statistical hypothesis, where the region of rejection is on both sides of the sampling distribution, is called a **two-tailed test**. For example, suppose the null hypothesis states that the mean is equal to 10. The alternative hypothesis would be that the mean is less than 10 or greater than 10. The region of rejection would consist of a range of numbers located on both sides of sampling distribution; that is, the region of rejection would consist partly of numbers that were less than 10 and partly of numbers that were greater than 10.

Power of a Hypothesis Test: The probability of *not* committing a Type II error is called the **power** of a hypothesis test.

EFFECT SIZE

To compute the power of the test, one offers an alternative view about the “true” value of the population parameter, assuming that the null hypothesis is false. The **effect size** is the difference between the true value and the value specified in the null hypothesis.

Effect size = True value - Hypothesized value

For example, suppose the null hypothesis states that a population mean is equal to 100. A researcher might ask: What is the probability of rejecting the null hypothesis if the true population mean is equal to 90? In this example, the effect size would be $90 - 100$, which equals -10 .

Degrees of freedom: Elementary tests usually depend on the data sample size as well as the number of parameters, (e.g.

mean or variance) that have to be estimated from the sample, to run the test. Specifically, the number of degrees of freedom of a statistics is defined as the number of independent observations minus the number of population parameters which must be estimated from sample observations. Details will be provided for each test.

p-values: Once hypotheses H_0 and H_1 have been defined, that a test has been chosen to address these hypotheses (see below), and that parameters for this test have been calculated, one must choose a level of significance. $p < 0.05$ is the arbitrary value that is generally accepted to be significant. It means that there must be less than a 5% possibility of falsely detecting a significant difference. We will now describe how the p value relates to the different types of errors associated with elementary tests.

Type I and type II errors: If we reject a hypothesis H_0 when it should be accepted, we say that a type I error has been made. If we accept a hypothesis H_0 when it should be rejected we say that a type II error has been made. In either case a wrong decision or judgment has occurred. This is not a simple matter because decreasing one error type usually leads to increasing the other error type. One way of getting around this problem is just to set your significance level at .05 (and not at .01 or .001). In this way you are balancing between type I and type II errors in your decision making process. One way to decrease both error types is to increase the size of the sample.

Table 1 : Statistical Methods

Goal	Dataset		
	Binomial or Discrete	Continuous measurement (from a normal distribution)	Continuous measurement, Rank, or Score (from non- normal distribution)
Example of data sample	List of patients recovering or not after a treatment	Readings of heart pressure from several patients	Ranking of several treatment efficiency by one expert
Describe one data sample	Proportions or frequencies	Mean, SD	Mean (SD) or Median, IQR(Inter Quartile range)
Compare one data sample to a hypothetical distribution	χ^2 or binomial test	One-sample t test	Sign test or Wilcoxon signed rank test
Compare two paired samples	Sign test	Paired t test	Sign test or Wilcoxon signed test
Compare two unpaired samples	χ^2 test for large samples or Fisher Exact test for small sample or frequencies	Unpaired t test	Mann-Whitney test
Compare three or more unmatched samples	χ^2 test for large samples or Fisher Exact test for small samples	One-way ANOVA	Kruskal-Wallis test
Compare three or more matched samples	Cochrane Q test	Repeated-measures ANOVA	Friedman test
Quantify association between two paired samples	Contingency coefficients	Pearson correlation	Spearman correlation

References:

1. Bernard Rosner (2000), Fundamentals of Biostatistics, 5th Edition, Duxbury.
2. Sunder Rao P S S , Richard J : An Introduction to Biostatistics, A manual for students in health sciences , New Delhi: Prentice hall of India.
3. Griner PF, Mayewski RJ, Mushlin AI, Greenland P (1981) Selection and interpretation of diagnostic tests and procedures. Annals of Internal Medicine.

4. Altman D.G and Bland J M (1999). Treatment allocation in controlled trials: Why randomize? British Medical Journal, 318: 1209.
5. Altman D.G and Bland J M (1999). How to use randomize. British Medical Journal, 319:703-704.
6. Altman DG. Practical statistics for Medical Research, London, UK; Chapman & Hall 1991.
7. Agresti A. Categorical Data Analysis, New York: John Wiley & Sons 1990.
8. Elise Whitley Jonathan Ball. Statistics review 4: Sample Size calculations. Critical care 2002; 6:335-341.
9. Donner A. Approaches to Sample size estimation in the design of clinical trials-A review, Statistics in Medicine 1984;3:199-214.
10. Suresh K.P. and Chandrasekhar S (2012). Sample Size estimation and Power analysis for Clinical research studies. Journal Human Reproduction Science,5(1), 7-13.

Spatial Epidemiology: An introduction

Divakar Hemadri

ICAR-National Institute of Veterinary Epidemiology & Disease Informatics, Bengaluru, India

Spatial Epidemiology

Spatial epidemiology is the description and analysis of geographic variations in disease with respect to demographic, environmental, behavioral, socioeconomic, genetic, and infectious risk factors

Why to analyze disease data spatially ?

It may be noted that many diseases are spatially constrained; for example, vector-borne and zoonotic diseases occur where and when vectors, animal hosts, pathogens and susceptible human populations overlap. Vectors, pathogens and animal populations are unevenly distributed in space and time and as a result risk for exposure to vector-borne diseases is spatially heterogeneous. Therefore, spatial models for the study and management of vector-borne disease risk have become common with the development of digitally encoded environmental data and computational tools such as geographical information systems (GIS). It is estimated that over 90% of health data (animal and human) has a spatial or geographical component.

- Variations in disease outbreaks are better understood if neighbouring farm networks are taken into account
- geography can also be used as a proxy
eg., measuring the distribution of outbreaks against vaccination activity, i.e “is there more FMD in non-vaccinated areas”
- and finally geography is a very useful framework for communication for example “how does the livestock health in Mizoram compare to Sikkim is more easily understandable when mapped

What do you require to do spatial analysis ?

Disease data

Time (if available datewise very good)
No of cases (no. affected)
Total population
No of deaths

Demographic Data

Specieswise data

Environmental Data

Rainfall
Temperature
Wind speed
Wind direction
Relative humidity
Vector Data

Socio-economic Data

Land Cover and Land Use Data

Soil Data

Digital Maps

Lat Long
Real time satellite images
Finally, skill and will

What can be achieved in spatial epidemiology ?

Disease mapping

Disease mapping is a field that concentrates on the spatial variation in the risk of disease. Basic map styles commonly used for disease mapping include Dot maps, Choropleth maps, Isopleth maps.

Dot maps, are suitable for presenting point data referenced in two-dimensional coordinate space, such

as the locations of disease events.

Choropleth maps are tremendously common and useful. These use some existing system of boundaries (countries, states, counties, voting districts, etc.). In choropleth maps, data is grouped into or more levels or classes using slicing values. These maps show spatial variation of one or two variables at a time by using color, shades of grey and/or patterns

Isopleth maps are especially well fitted for inspecting continuously varying phenomena, such as temperature, rainfall etc. These maps feature continuously varying color values basically by a line on a map that connects points of equal value.

Geographic correlation studies

These look at correlations between variables. For example, a study may investigate the occurrence of meningococcal infection. In doing this it might correlate the economical status and personal hygiene of people. One would expect that as level of infection to be more among economically weaker class. Some studies may focus on habitat mapping of a particular species of insect based on the data obtained from a small survey. Another example of geographical correlation studies is relating human taeniasis infection with heavily infected cysticerctic pigs.

The assessment of risk in relation to a point or line-source

Point and line-source studies assume a risk source that has the shape of a point or a line, such as a chimney or an electrical wire. Highly localized studies are conducted in order to discover possible increases in ill-health due to increased exposure from these specific types of sources.

Cluster detection and disease clustering

Cluster detection is carried out in order to detect raised levels of incidence of disease. If disease cases seem to form non-random patterns, that is, clusters, then there is reason to suspect that the underlying effect is non-random. It can provide information on the etiological background of the disease. A spatial disease cluster may be defined as an area with an unusually elevated disease incidence rate. The identification of a cluster of disease can help epidemiologists determining putative environmental risk factors and lead to improved understanding of etiology.

Commonly encountered problems in spatial epidemiological studies

- Non availability of good quality of data which may lead to Garbage in = Garbage out kind of situation.
- Georeferencing of attribute data
- Government data rich but information poor
- Lag time between environmental exposure and presentation
- Biological plausibility – does it make sense (critical thinking skills)
- Spatial data is expensive
- Spatial data must be accurate and up to date
- Complexity of analysis - requires training
- Complexity of activity – multiple exposures points
- Interpretation - ecological fallacy, aggregating data

Summary

Spatial epidemiology is the study of the spatial patterns, processes and determinants of health and disease

Spatial epidemiological methods use GIS and spatial based analytical tools

A GIS can place data in an environmental context – but don't overlook human input

Informatics on Protozoan Parasitic diseases in large ruminants

P.P. Sengupta & S.S. Jacob

ICAR-National Institute of Veterinary Epidemiology & Disease Informatics, Bengaluru, India

Protozoa are ubiquitous throughout aqueous environments and the soil, and play an important role in their ecology. Farm animals are usually infected with several species of parasites and they are also confined to pasture or pens. Many times they are transmitted through vector; however some time direct transmission is also observed. Some are also transmissible to humans.

The important protozoan diseases in large ruminants are :

Babesiosis

Babesiosis is an infectious tick-borne disease of livestock that is characterised by fever, anemia, haemoglobinuria and weakness. The disease is also known by such names as bovine babesiosis, piroplasmosis, Texas fever, red water fever, tick fever etc. The disease also is a hemoparasitic disease caused by protozoa of the genus *Babesia* which infects mainly ruminants. Infection of a vertebrate host is initiated by inoculation of sporozoite form of parasites into the blood stream during the taking of a blood meal by tick vectors.

Organisms	Host affected	Geographical distribution
<i>B. bigemina</i>	Cattle	Americas, Europe, Africa, Australia, Middle East
<i>B. bovis</i>	Cattle	Americas, Europe, Africa, Asia Australia
<i>B. major</i>	Cattle	Europe, North Africa, Middle East
<i>B. jakimovi</i>	Cattle	Asia
<i>B. ovata</i>	Cattle	Asia
<i>B. divergens</i>	Cattle	Europe

Bovine babesiosis

Bovine babesiosis associated with *B. bigemina* and *B. bovis* is the most important disease of tropical and subtropical regions. Both species are transmitted transovarially by *Boophilus* ticks, but only tick larvae transmit *B. bovis*, whereas nymphs and adults transmit *B. bigemina*. In Europe, babesiosis is caused by *Babesia divergens*, an intraerythrocytic parasite that can persist for >13 months in the organs of infected cattle. The distribution of *B. divergens* reflects its triphasic telotrophic tick vector, *Ixodes ricinus*, which is widespread across Western Europe and North Africa. *B. major* occur in Europe, North Africa and South America. *B. major* is transmitted by the three host tick *Haemaphysalis* sp. (Taylor et al., 2007).

Pathogenesis

Babesia spp. are a various group of tickborne, obligate, intra-erythrocytic Apicomplexan parasites infecting a wide variety of animals. Ticks are most often infected transovarially. The female tick becomes infected by the ingestion of parasites during engorgement. After it drops off the host, the babesial agents reproduce within the tick's tissues. Some of the reproducing organisms are incorporated within developing tick embryos, and the disease agents are transmitted to new hosts by the feeding of ensuing tick larvae, nymphs, or adults (Zaugg, 2009). *B. bovis* is the most pathogenic of the bovine *Babesia*. *B. bigemina* infections are not as virulent as those of *B. bovis*, however the parasites may infect 40% of the red cells (Taylor et al., 2007).

Clinical findings

Incubation period is 2-3 weeks. *B. bigemina* and *B. bovis* produce acute syndromes which are clinically

indistinguishable, and are characterized by high fever (41°C), anorexia, depression, weakness, cessation of rumination, and a fall in milk yield. Hemoglobinuria can be seen, the color of urine is dark-red to brown. Respiratory and heart rates are increased, and the red conjunctivae and mucous membranes change to the extreme pallor of severe anemia. Abortion occurs in pregnant animals. Subacute syndrome also occurs in young animals, but fever is mild and hemoglobinuria is absent (Radostits et al., 2008). In cerebral babesiosis, hyperexcitability, convulsions, opisthotonos, coma, and death, may be observed in cattle infected with either *B. bigemina* or *B. bovis*, but especially with the *B. bovis*. Central nervous system signs are caused by brain anoxia resulting from severe anemia (Zaugg, 2009).

Diagnosis

Blood smears and clinical findings are useful in acute cases of piroplasmosis, but are not sufficient in subclinical cases. The complement fixation test is used as a serological test for bovine babesiosis. The most commonly used tests are ELISA, PCR and a DNA probe, which can detect specific parasitemias at very low levels of infection (Radostits, 2008). Recently, the 'reverse line blot (RLB)' is a versatile technique for simultaneous detection and identification of small ruminant piroplasm species, based on the recognition of specific gene regions by oligonucleotide probes.

Treatment

After the hemoglobinuria or cerebral signs, prognosis is not well. In acute cases that PCV values are above 12%, treatment will be successful. Supportive therapy such as blood transfusions (4 L of whole blood per 250 kg of body weight), fluids, hematinics, and prophylactic antibiotics are important (Zaugg, 2009). Babesiosis can be treated using diminazene aceturate (3-5 mg/kg), phenemidine diisethionate (8-13 mg/kg), imidocarb dipropionate (1-3 mg/kg), and amicarbalide diisethionate (5-10 mg/kg) (Zaugg, 2009).

Control

The control of the disease depends on effective quarantine to prevent the introduction of the vector tick. The control of ticks by dipping or spraying animals at risk with recommended acaricides. In routine surgery, care should be taken to prevent accidental transfer of blood from one animal to another (e.g., castration, dehorning). In addition, in cattle, the selection and breeding of cattle which acquire a high degree of resistance to ticks is practiced. Widespread use of tick vaccines may also have a significant influence on the incidence of infection in cattle (Taylor et al., 2007; Radostits et al., 2008; Zaugg, 2009).

Theileriosis

Theileriosis is caused by *Theileria* spp. in cattle, goats, sheep and wild and captive ungulates. Theileriosis is a hemoparasitic disease caused by protozoa of the genus *Theileria* (Apicomplexa). *Theileria* species affect domestic and wild ruminants, especially in Africa, Europe, Australia, and Asia. The parasites are transmitted by tick. These parasites undergo repeated merogony in the lymphocytes ultimately releasing small merozoites, which invade the red cells to become piroplasms. Theileriosis has a variety of tick vectors which cause infections ranging from clinically inapparent to rapidly fatal.

Organisms	Host affected	Geographical distribution
<i>T. annulata</i>	Cattle, domestic buffalo	Mediterranean countries, Middle East, Indian subcontinent and China
<i>T. orientalis</i>	Cattle	Southern Europe, Middle East, Asia, Australia
<i>T. mutans</i>	Cattle	Africa, Caribbean Islands
<i>T. taurotragi</i>	Cattle	Africa
<i>T. velifera</i>	Cattle	Africa

Epidemiology

In bovine, Tropical Theileriosis (Mediterranean coast fever), occurs in Mediterranean countries, Middle East, Indian and China, and is caused by *T. annulata* (Radostits et al., 2008). The tick vectors are *Hyalomma dentritum* in north Africa and in the Mediterranean countries, *H. dentritum* and *H. dromedari* in central Asia, and *H. marginatum* in India (Taylor et al., 2007; Radostits et al., 2008). In endemic areas indigenous cattle are relatively resistant while cross breed cattle, especially European breeds, are highly susceptible (Taylor et al., 2007). *Theileria orientalis* complex is a milder disease than East coast fever and Tropical Theileriosis, and called benign theileriosis in cattle (Radostits et al., 2008). The tick vectors are *Amblyomma variegatum*, *A. cohaerens*, *A. haebraeum*, *Haemaphysalis bispinosa* are the probable vectors in Australia (Taylor et al., 2007).

Pathogenesis

Pathogenesis of different form of theileriosis is based on the production of schizonts in lymphocytes and piroplasms in erythrocyte. *T. parva*, *T. annulata* and *T. hicri* produce highly schizonts and piroplasma and are very pathogenic; *T. mutans*, *T. buffeli*, and *T. ovis* unusually produce schizonts but may cause varying degree of anemia when piroplasms are highly in red blood cells (Radostits et al., 2008).

Clinical findings

Theileia spp. are classified in to 2 groups. In first group (*T. parva* and *T. annulata*), proliferate is seen in lymphocytes but in the second group (*T. orientalis*) it is seen in erythrocytes that causes hemolytic anemia (Magona et al., 2010). In *Theileria parva* incubation period is approximately 1-3 weeks. One or two days later, the first clinical sign is generalised swelling of the superficial lymph nodes, eyes, ears and submandibular regions. After few days there is anorexia, decreased milk production, loses condition, ceases rumination, rapid hearth beat, petechial haemorrhages under the tongue and on the vulva. In cerebral theileriosis there are localized nervous signs and convulsions, tremor, salivation and head pressing (Radostits et al., 2008). In *Theileria annulata*, pyrexia, anorexia, enlargement of superficial lymph nodes, nasal and ocular discharges and salivation is the most common signs. Constipation is recorded in some cases. Respiratory distress, coughing and pulmonary oedema are seen.

In *Theileria orientalis* clinical signs are associated with anemia, jaundice and lymphadenopathy. Clinical signs in *Theileria velifera* is not reported but in *Theileria taurotragi* mild fever and anemia are reported.

Diagnosis

Investigation of giemsa-stained blood smears and lymph node biopsy will reveal piroplasms in erythrocytes and schizonts in lymphocytes. For diagnosis, indirect fluorescent antibody test (IFAT) and indirect enzyme-linked immunosorbent assay (ELISA) are the most commonly used techniques. The ELISA tests is more sensitive than IFAT.

Treatment and control

Buparvaquone is the most effective drug and the recommended dose in cattle, sheep and goat is 2.5mg /kg BW. In control of the disease use of genetically resistant breed, a judicious and selective application of acaricides at 3- week intervals and the use of vaccines are recommended.

Trypanosomosis

Trypanosoma evansi, a haemoflagellate extra cellular protozoan parasite causes a disease known as trypanosomosis ('surra'). It affects a number of species of domesticated animals in Asia, Africa and central and south America. The host species affected are bubalines, bovines, dromedarines, equines, felines, canines. Recently it has also been reported several times from human host. The tryps are transmitted mechanically by haematophagous flies – *Tabanus* sp., *Stomoxys* sp. and *Lyperosia* sp.

Tsetse-transmitted trypanosomosis is a disease complex caused by several of these species, mainly transmitted cyclically by the genus *Glossina* (tsetse flies), but also mechanically by biting flies. Tsetse infest 10 million square kilometres and affect 37 countries, mostly in Africa, where it is known as 'nagana'. The disease infects various species of mammals but, from an economic point of view, tsetse-transmitted trypanosomosis is particularly important in cattle (also referred as tsetse-fly disease in southern Africa). It is mainly caused by *Trypanosoma congolense*, *T. vivax* and, to a lesser extent, *T. brucei brucei*. *Trypanosoma vivax* is also transmitted mechanically by biting flies, among which tabanids and stomoxes are presumed to be the most important, as exemplified by its presence in South and Central America, but also as observed in some areas of Africa free or cleared of tsetse (Ethiopia, Chad, etc.). Tsetse-transmitted trypanosomosis can affect camels and is a natural barrier preventing the introduction of this mammalian species into the southern Sahel region of West Africa. Horses are also highly sensitive. Very rare human cases have been observed caused by animal *Trypanosoma* species. However, tsetse transmitted trypanosomosis also affects humans, causing sleeping sickness, through infection with either *T. brucei gambiense* or *T. brucei rhodesiense*. A large range of wild and domestic animals can act as reservoirs of these human parasites; particular care must be taken for people handling biological material that can contain infective human parasites, for example in livestock.

Chagas disease is caused due to *Trypanosoma cruzi* infection. The disease is a public health threat in most Latin American countries, although cases due to blood derivatives or blood transfusion has been reported in non-endemic regions. According to WHO the overall prevalence of human *T. cruzi* infection is estimated in 18 million cases and 100 million people are living at risk. The vectors are reduviidae bugs which are haematophagous and the most important are *Triatoma infestans* (Argentina, Chile, Brazil, Bolivia, Paraguay, Uruguay, Peru), *T. sordida* (Argentina, Bolivia, Brazil, Paraguay), *Rhodnius prolixus* (Colombia, Venezuela, Mexico, Central America), *T. dimidiata* (Ecuador, Mexico, Central America), and *Panstrongylus megistus* (northeast Brazil). The transmission by the vector is faecal contamination.

Pathogenesis

After entering through the skin, parasites reach the bloodstream via the lymphatic system. Infection is characterized with parasitemia. Some *Trypanosoma* spp. invade extravascular spaces such as the ocular aqueous humor and cerebral spinal fluid. Some *trypanosoma* spp. may produce hemolysin that causes anemia in the host. Then, phagocytic activity increased because of the massive erythrocyte failure.

Clinical findings

Clinical findings are based on the speed of onset of anemia and the grade of organ impairment. Trypanosomosis can be acute, subacute, or chronic. In acute form abortion, drop in milk, depression, anorexia can be seen. Hyperemic mucous membranes and lacrimation also can occur. In subacute form clinical signs include weight loss, enlargement of lymph nodes and dry hair coat. In chronic form dull, dry hair coat, inelastic skin, lethargy, pale mucous membranes and exercise intolerance may be seen.

Diagnosis

Diagnosis can be based on the clinical findings, presence of vectors, appearance of trypanosomes on a fresh blood smear, or a Giemsa-stained blood smear. Indirect fluorescent antibody test (IFA) and the enzyme-linked immunospecific assay (ELISA) test are used for diagnosis. Different ELISAs employing recombinant antigens and monoclonal antibody have already been developed (Sengupta et al., 2014, 2016; Ligi et al., 2016; Rudramurthy et al., 2015; 2017). PCR based assay targeting VSG and ISG gene have also been developed (Sengupta et al., 2010; Rudramurthy et al., 2013).

Treatment and control

The most common drugs that is used for treatment of trypanosomosis are:

- Diminazene aceturate 3.5-7 mg/ kg BW.
- Homidium bromide and chloride 1mg /kg BW -Pyrrithidium bromide 2mg / kg BW.
- Isometamidium 0.25-1mg / kg BW (Radostits et al., 2008).

Vector control can help to control or prevent trypanosomosis. Insecticides can be used.

Sarcocystosis

Sarcocystosis caused by *Sarcocystis* species in cattle, sheep and goats. The names of *Sarcocystis* species are according to their intermediate and final.

Parasite	Intermediate Host	Definitive Host	Geographical Distribution
<i>S. bovicanis</i> (<i>S. cruzi</i>)	Cattle	Dog	World wide
<i>S. bovifelis</i> (<i>S. hirsuta</i>)	Cattle	Cat	World wide
<i>S. bovi hominis</i> (<i>S. hominis</i>)	Cattle	Human	World wide

Pathogenesis

Sarcocystis spp, are protozoon parasites with a requisite two-host life cycle. Sexual reproduction phase occur in the intestine of a carnivore (dogs, cats) later, cysts in the muscles of cattle, sheep or goats. Sporocysts are shedded by carnivore's feces and then ingested by cattle, sheep, or goats. Then sporocysts hatch in the ruminant small bowel and invade the vascular endothelium during three phases of asexual reproduction. After third phase, merozoites enter the ruminant's muscle tissue and encyst as sarcocysts.

Clinical findings

In cattle is usually asymptomatic, but in heavy infections in nonimmune cattle, clinical signs include fever, anorexia, anemia, weight loss, lameness, abortion and diarrhea may occur (Taylor et al., 2007; Radostits et al., 2008). Neurologic signs are occasionally seen in cattle that include ataxia, tremors, muscular weakness, hypersalivation, blindness, opisthotonos and nystagmus (Radostits et al., 2008). In infections that caused by *Sarcocystis bovicanis* in cattle there is usually loss of hair at the end of the tail (Taylor et al., 2007). In sheep and goats clinical signs may be asymptomatic. In heavy infections there is fever, loss of weight, anemia and weakness. Abortion may occur. In chronic infections clinical signs include poor weight gain, edema of the limbs, anemia and abortion. Neurological symptoms, such as depression, in coordination, hind leg paralysis and coma can be seen in Encephalitic Sarcocystosis .

Diagnosis

Generally *Sarcocystis* infections are diagnosed at meat inspection with grossly visible sarcocysts in the animal's muscle. When infection is very heavy in intermediate hosts, clinical signs and histological evidence of schizonts in the blood vessels of organs and the presence of cysts in the muscles at necropsy will be used for diagnosis (Urquhart et al., 1987). Indirect hemagglutination test (IHA) and ELISA test are used for serological diagnosis. In acute form of disease titers of antibodies are not high but 1 week to 3 months later will be at diagnostic levels. For certain diagnosis immunohistochemistry, electron microscopy and PCR techniques are available (Radostits et al., 2008).

Treatment and control

There is no effective treatment for sarcocystosis. (Urquhart et al., 1987; Radostits et al., 2008;). Treatment of infected calves and sheep with salinomycin (4 mg /kg and 1-2 mg /kg BW; respectively) has been recommended.

Amprolium 100 mg /kg, for 30 days reduces the severity of infection.

After beginning nervous system symptoms in sheep, Encephalitic Sarcocystosis recovery has not been observed. Control of disease is based on protection the food supply of ruminants. Feed bunk should be kept clean, also farm dogs and cats that have access to the feed or pastures should not be fed uncooked meat (Urquhart et al., 1987).

Neosporosis

Neosporosis caused by *Neospora caninum* in cattle, sheep and goats. *Neospora caninum* is a protozoan parasite of the phylum Apicomplexa in the family Sarcocystidae (Radostits et al., 2008).

Epidemiology

The protozoa *Neospora caninum* is an important parasite that cause abortion in cattle, sheep and goats . The majority of *N. caninum*- positive cattle prenatally infected via their dams. Transplacental transmission is considered the major route of transmission of *N. caninum* in cattle. In nonfatal infection in the fetus, the fetus is born with neurologic disorder (Smith and George, 2009). *N. caninum* has a worldwide distribution, the prevalence of infection in cattle and sheep approaches 100% with a lower ((Taylor et al., 2007; Radostits et al., 2008). In India, it has been observed that the seroprevalence is more among cattle from unorganized herds than organized. The sero-positive cattle are 8.4 times more likely to have abortion history than their sero-negative counterparts (Sengupta et al., 2013).

Pathogenesis

Definitive host: Dogs are the final host and sexual phase occur in them but, they are also intermediate host in prenatal infections.

Intremediate host: Cattle are the major intermediate hosts and asexual reproduction phase occur in them. Infection can be transmitted from dam to calf in utero and lactogenically. Infection of cattle can also occur from the ingestion of food or water contaminated with dog feces containing *Neospora caninum* oocysts (Taylor et al., 2007; Radostits et al., 2008). *Neospora caninum* is a major cause of abortion in cattle, however, sporadic abortions can occur in beef cows that have been infected congenitally (Dubey et al., 2006; Taylor et al., 2007; Radostits et al., 2008).

Asexual phase has 2 stages:

1. Tachyzoites: Tachyzoites penetrate host cell like central nervous system, muscles, macrophages and other cells, where they divide rapidly. Tachyzoites can also be transmitted either with contaminated food and water or transplacentally to the fetus in pregnant animals. Tissue cyst containing bradyzoites that are found only in the nervous system. After the asexual phase, sexual phase occur in definite host. It results in production of oocysts, which is shed in the dog feces.
2. Tissue cysts (Taylor et al., 2007). Infection in sheep and goats is infrequently (Radostits et al., 2008).

Clinical findings

In cattle, neosporosis causes stillbirth, fetal resorption, mummification, abortion and decreases in their milk production . Abortions in cows are seen between 5-7 month gestations . Fetus may born alive but congenitally diseased . Neurological symptoms are different because of the widespread distribution of the parasite in the central nervous system. Calves are born with neurological symptoms, which these symptoms initially are mild but after birth become progress. In calves with neurologic dysfunction clinical signs are included of unable to stand, unable to suckle, domed skull and torticollis.

Diagnosis

The diagnosis of neosporosis is based on the examination of maternal and fetal sera ideally combined with the examination of fetal tissues.

- Immunofluorescent antibody test (IFAT) and indirect enzyme-linked immunosorbent assay (ELISA) are used for diagnosis.

- Histopathology of fetus: In histopathological examination of brain characteristic nonsuppurative encephalitis is suggestive of *Neospora* infection and also the lesions in the heart are characteristic for diagnosis .

Treatment and control

At present, there is no effective treatment for bovine neosporosis.

Control of abortion in infected cattle depends on saving food and water sources and the grazing environment from feces of any animal. Aborted fetuses and placentas should be removed or incinerated. The feces of dogs should be prevented from contaminating animal foodstuffs. Congenitally infected cows are at high risk for abortion thus seropositive animals should be culling from a herd (Radostits et al., 2008).

References :

1. Buxton, D., Thomson, K.M., and Maley, S., 1993. Treatment of ovine toxoplasmosis with a combination of sulphamezathine and pyrimethamine. *Veterinary Record*. 132 (16), 409-411.
2. Dubey, J.P., Buxton, D., and Wouda, W., 2006. Pathogenesis of bovine neosporosis. *Journal of Comparative Pathology*. 134 (4), 267-289.
3. Dubey, J.P., Lindsay, D.S., and Speer, C.A., 1998. Structure of *Toxoplasma gondii* tachyzoites, bradyzoites and sporozoites, and biology and development of tissue cysts. *Clinical Microbiology Reviews* 11, 267–299.
4. Ligi, M., Sengupta, P.P., Rudramurthy, G.R. and Rahman, H. (2016). Flagellar antigen based CI-ELISA for sero-surveillance of surra. *Veterinary Parasitology* 219: 17-23.
5. Radostits, O.M., Gay, C.C., Hinchcliff, K.W., and Constable, P.D., 2008. Diseases associated with protozoa. 10th Edn. In: *Veterinary Medicine: A Textbook of Diseases of cattle, horses, sheep, pigs, and goats*. Saunders Elsevier; pp: 1483-1540.
6. Rudramurthy, G.R., Sengupta, P.P., Metilda, B., Balamurugan, V., Prabhudas, K., and Rahman, H. (2015). Development of an enzyme immunoassay using recombinant invariant surface glycoprotein (rISG) 75 for serodiagnosis of bovine trypanosomosis. *Indian Journal of Experimental Biology* 53: 7-15.
7. Rudramurthy, G.R., Sengupta, P.P., Balamurugan, V., Parbhudas, K. and Rahman, H. (2013). PCR based diagnosis of trypanosomiasis exploring invariant surface glycoprotein (ISG) 75 gene. *Veterinary Parasitology* 193: 47– 58.
8. Rudramurthy, G.R., Sengupta, P.P., Ligi, M. and Rahman, H. (2017). An inhibition enzyme immune assay exploring recombinant invariant surface glycoprotein and monoclonal antibodies for surveillance of surra in animals. *Biologicals* 46: 148- 152.
9. Sengupta, P.P., Balumahendiran, M., Suryanaryana, V.V.S., Raghavendra, A.G., Shome, B.R., Gajendragad, M.R., and Prabhudas, K. (2010) PCR-based diagnosis of surra-targeting VSG gene: Experimental studies in small laboratory rodents and buffalo. *Veterinary Parasitology* 171:22-31.
10. Sengupta, P.P., Balumahendiran, M., Raghavendra, A.G., Honnappa, T.G., Gajendragad, M.R. and Prabhudas, K. (2013). Prevalence of *Neospora caninum* antibodies in dairy cattle and water buffaloes and associated abortions in the plateau of southern peninsular India. *Tropical Animal Health and Production* 45:205–210.
11. Sengupta, P.P., Rudramurthy, G.R., Ligi, M., Roy, M., Balamurugan, V., Krishnamoorthy, P., Nagalingam, M., Singh, L. and Rahman, H. (2014) Sero- Diagnosis of surra exploiting recombinant VSG antigen based ELISA for Surveillance. *Veterinary Parasitology* 205:490-498.
12. Sengupta, P.P., Ligi, M., Rudramurthy, G.R., Balamurugan, V. and Rahman, H. (2016). Development of ELISA exploring recombinant variable surface glycoprotein for diagnosis of surra in animals. *Current Science* 111(12): 2022 - 2027.
13. Shuzhen, G., Zhengphu, Y., Guoxue, W., Wenbiao, W., Denglu, M., and Hongde, D., 2002. Epidemiology of ovine theileriosis in Ganan region, Gansu Province, China. *Parasitology Research*. 88 (Suppl. 1). Berlin: Springer- Verlag, 536-537.
14. Taylor, M.A., Coop, R.L., and Wall, R.L., 2007. *Veterinary Parasitology*. Third Edn. Blackwell Publishing.
15. Urquhart, G.M., Armour, J., Duncan, J.L., Dunn, A.M., and Jennings, F.W., 1987. *Veterinary Parasitology*. Longman Group UK Limited, Essex CM20 2JE, England.
16. Zaugg, J.L., 2009. Babesiosis. In: Smith, B.P. (Eds): *Large Animal Internal Medicine*. Mosby, Elsevier, St. Louis, p 1157.

Geographical Epidemiology, Spatial Analysis and Livestock Disease Information System.

K.P.Suresh, S.S. Patil and D. Hemadri

ICAR-National Institute of Veterinary Epidemiology & Disease Informatics, Bengaluru, India

Place is an important element in understanding the geographic distribution of livestock diseases and means of formulating and testing etiological hypothesis. The term 'place' is usually considered as surrogate for interaction of genetic factors, species type or host and environment. In the prospective of veterinary health practices, knowledge that the veterinary health issues is concentrated in identifiable places is essential for the efficient distribution of resources for prevention, minimizing the disease burden or prevalence, treatment or amelioration. In recent years, availability of geographical information system and databases provide the excellent environment to explore the disease pattern in the given study area.

Age, breed, species type, environment material and other variables vary from one place to another and hence influences the varying risk of disease incidence. The observed differences of risk of diseases is likely to be confounded by these variables and hence the comparison of risk variability is an important issue in the process of disease epidemiology. The adjustment of potential confounding variables is important for evaluation of spread of risk of disease incidence.

Data for Spatial analysis

Point and layer (area) data are two important types of spatial data. Each item in the veterinary health data system viz., population, breed, environmental exposure, morbidity and mortality relate to a point data while data like village, herd, state, districts, blocks etc. indicated as layer data. The data for spatial analysis comes from different sources, it is advisable collect these data taking help of veterinary epidemiologists or Bio-statisticians. It is necessary to ensure that precise and complete point or area data are used in spatial epidemiology. The data related to diagnosis, collection, coding and reporting of a given health events may vary between geographical regions and time.

The quality of data must be given important in spatial analysis of data, presence of missing cases, underreporting, under ascertainment of cases, inaccurate baseline population data, may lead to inaccurate or misleading interpretation of estimated high or low risk. The confidentiality is also an important issue in spatial epidemiology, breaching the confidentiality may cause concern, particularly when it discloses the area with high rates of morbidity/mortality, outbreaks or risk of disease incidence.

Disease Clustering

Scanning for disease clustering is one of the important topic in geographical epidemiology, involves the assessment local or global concentration of disease events. There are two types of clustering, general clustering and specific clustering. Analysis of disease events for overall assessment of clustering tendency and its spatial autocorrelation refers to general clustering, whereas the specific clustering methods are designed to investigate the exact location of clustering. Methods for detection of clusters in Point format data are more compared to a real format of data, and are usually divided into three groups, global, localized and focused. Large number of statistical tests available to assess the different kinds of clusters in point format data.

A Cluster is being a bounded group of occurrences, i) of a disease already known to occur characteristically in clusters, or ii) of sufficient size and concentrations to be unlikely to have occurred by chance, or iii) related

to each other through some social or biological mechanism or having common relationship with some other event or conditions. The statistics of disease clustering are useful detect and monitor potential public hazards.

Disease mapping

Mapping of disease events is one of the best method for better visualization for exploring the complex structure of data. Data visualization is not only creating interest and also attract the attention of viewer and provide the way for discovering the pattern. Disease mapping is one of the tools of geographical epidemiology, fulfilling the need to generate accurate and precise maps of disease events. For Example, dot or dot density maps are used to display point data, while areal data were presented by Choropleth maps and for continuous surface data, contour maps or isopleth maps are used. In the veterinary epidemiology, the presentation of maps is established as a basic tool for analysis and interpretation. The selection of appropriate of administrative unit for mapping viz., village, block, district or state, selection of suitable data classification in the map viz., high risk, moderate risk, low risk no risk etc., and selection of color schemes are the important issues associated with creation of disease maps.

Ecological analysis

Ecological analysis refers to the analysis of association of disease incidence or outbreak with covariates of interest viz. anthropogenic variables, socio-economic variables, host related information, environmental variables and remote sensing variables etc. In the ecological analysis, the variables are defined on aggregated groups of individuals rather than individuals themselves. In the ecological analysis, the important issue to be attended is scale of covariate. The optimum choice of scale is a trade-off between making the groups (regions) large enough to have stable estimates and small enough to make them homogenous in terms of their socio-economic and other important covariates. For large regions, there is greater possibility that associations measured at the aggregate level will differ from the same association measured at individual level. This can lead to a problem known as ecological fallacy, or ecological bias, in this scenario, inference is made incorrectly at individual -level association from one that is observed at regional level. On other hand, if the regions or units chosen are too small, the results may show spurious spatial correlation or pattern due to random variation in small number of disease events.

The uncertainty induced by the aggregation procedure may result from scale dependency data and cause modifiable unit problem. Hence it is important to consider the scale of analysis of ecological variables and analyze the data at two or more levels of aggregation. It is possible to overcome the issues like scale dependency and ecological bias, by the se of multi-level approach using individual level and group level together.

Geographical epidemiology

Geographical epidemiology is defined as the description of spatial pattern of livestock disease events such as outbreak, incidence, prevalence, mortality etc. It is an analytical epidemiology wherein specific hypothesis will be formulated for testing of etiology of disease events. Disease mapping, disease clustering and ecological analysis are important predominant methods of geographical epidemiology. Most of the geographical epidemiological studies depends on scale of ecological variables, leads to serious spatial problem when the results of geographic aggregate level data are combined with those at individual level. The combination of aggregate and individual level data analysis is possible with multilevel modelling, hierarchical regression, and contextual analysis. Multilevel modelling is more powerful and a new technique, that can be used to determine the ecological effects explained by individual risk factors.

Geographical Information systems

Geographical Information System (GIS) can be defined as software system for the automated system to capture, storage, retrieval, analysis and display of spatio-temporal data. GIS has dramatically changed the ability of epidemiologists and public health specialists to work with spatial data. The advantages of GIS include an ability to operate repetitive tasks, handle large volume of data and quickly compare the spatial & temporal data. GIS has traditionally been used for maps when analyzing associations between locations, environment and disease events, recently GIS is used in the surveillance and monitoring of vector-borne diseases, water-borne disease, environmental health, modelling the exposures, prediction of disease events and analysis of disease policy and planning.

Veterinary geographic information system provides a strong framework for increasing our ability to monitor the disease pattern and identify their causes. The evolution of GIS from early disease maps to digital maps is a long journey in the making and continues to evolve. These maps have enabled us to gain insight about temporal and spatial pattern of livestock diseases and increasing knowledge of worldwide health issues.

Global Positions systems

Global Positions systems(GPS) is a system consists of group of powered satellites orbiting earth every 12 hour and transmitting radio pulses at every time intervals. To determine the position of specified location in three dimensions, latitude, longitude and elevation, a receiver needs signals from at least four satellites. GPS has become standard method for data capturing in geographical epidemiology and public health studies. Using GPS, it is currently possible to obtain the latitude and longitude of location at higher resolution. The impact of GPS in providing spatially referenced case data is potentially greatest in the developing countries where the absence of quality paper maps in that data are not often collectible. The other technical advance that has made obtaining spatially referenced case data increasingly easy for geocoding using postal codes.

Disease modelling and Smoothing

A disaster, precipitated by a natural hazard, can be defined as a “serious disruption of the functioning of community or society or livestock system causing widespread human, animal, material, economic or environmental losses which exceed the ability of the affected community or society to cope using its own resources (ISDR, 2004, Terminology: basic terms of disaster risk reduction, International strategy for disaster reduction, Geneva). The expression of ‘early warning’ is used in many fields to mean the provision of information on an emerging dangerous circumstance where the information can enable action in advance to reduce the risks involved. Early warning systems exist for natural geophysical and biological hazards, complex socio-political emergencies, industrial hazards, and many other related fields.

Early identification of an infectious disease outbreak is an important first step towards implementing effective disease interventions and reducing resulting mortality and morbidity. The geographic and seasonal distribution of many infectious diseases are associated with climate and therefore the possibility of using seasonal climate forecasts as predictive indicators in disease early warning system (EWS) is an interest of focus. Geographic Information system (GIS), remote sensing (RS) and Global Positioning system (GPS) are the three commonly used veterinary geo-informatics technologies employed in this digital era for rapid communication of data for better management of animal diseases.

The ability to detect disease incidence or outbreaks early is important to minimize morbidity as well as mortality through timely implementation of disease prevention and control measures. The attacks on world trade center

& terrorist attack using anthrax spores in 2001, as well as recent SARS outbreaks, have motivated many public health authorities to develop system for early detection of outbreak using non-diagnostic information, often derived from electronic data collected for other purposes. Emerging infectious disease poses a threat to livestock population, climate change like increased temperature and altered rainfall patterns are likely to increase the burden of vector-borne diseases resulting into emergence of zoonotic diseases. Forecasting is the monitoring of specific risk parameters helping to predict the situation that could lead to occurrence of disease and its subsequent spread. The forecasting of disease helps to predict the course of disease, warn health care workers and adopt control measures to prevent disease outbreaks.

Spatial autocorrelation

Spatial autocorrelation refers to lack of independence of neighboring areas. The correlation or dependency refers to the index for geographically close areas are more related than those areas that are geographically distant in respect of disease events. Detecting the spatial dependency would help researchers to justify the application of statistical models or smoothing techniques for disease mapping of rare disease or mapping is done for small boundaries. Spatial autocorrelation statistics provide a very useful summary information about the spatial arrangement of disease events in a map. Moran I and Geary's C statistic are two commonly used spatial autocorrelation statistics for detecting the global clustering in continuous areal data. There is also many spatial autocorrelation statistics available like Getis and Ord's G, which measures the local clustering.

Mobile Phone Technology

Global spread of mobiles phones and networks at higher internet bandwidth, this technology is increasingly used to communicate early warning services and coordinate preparation activities. SMS alerts can be used to alert the local veterinary officers about the possible of occurrences of disease outbreaks. Advanced mobiles phones can also be used for video conferencing of disease information activities.

Early warning systems are combinations of tools and process embedded within institutional structures coordinated by national or international agencies. These systems are composed of four elements depending upon they focus on specific hazard or many, namely, knowledge of risk, a technical monitoring and warning services, dissemination of meaningful warnings to at-risk areas, and farmers awareness and preparedness to act. Warning services lie at the core of these systems, and how well they operate depends on having a sound scientific basis for predicting and forecasting. As early warning systems grow in geographical coverage and sophistication, false alarms to in rise. High false alarms can undermine the public confidence, breed mistrust, dilute the impact of alerts and reduce the credibility of future warnings.

Reference :

1. Developing Early Warning Systems: A checklist. Report from the Third International Conference on Early Warning (Report, UNISDR, 2006).
2. *Global Assessment Report for Disaster Risk Reduction*. (Report, UNISDR, 2011).

Economic losses due to livestock diseases

Dr. G.Govindaraj

ICAR-National Institute of Veterinary Epidemiology and Disease Informatics

Introduction

The disease in animals reduces the efficiency with which inputs are converted into outputs. There are several direct and indirect effects, of which, some can be valued or quantified easily and some are difficult to quantify. Besides quantification using the reliable data collected from primary or secondary sources, the effect of a disease can be modelled with certain assumptions like increased death rate, lower yield, decreased body weight, increased calving/kidding interval etc. It is also sometimes referred as simulation assessment of the impact of the disease.

Disease impact

Any disease in animals has direct and indirect impacts on the productivity of animals and on animal keepers. The direct impact is further classified into visible and invisible impacts. The visible impacts can be easily quantified eg. mortality loss, reduction in milk, wool reduction in draught power availability etc whereas invisible impacts like reduced fertility, changed herd structure etc., are relatively difficult to capture. The indirect impact includes societal and financial impact. The death of animals due to a disease in large numbers in general or death of certain breeds/ species due to outbreaks in a certain geographical region has variable societal impact. The indirect impact like change in dung availability has direct bearing agriculture and allied activities productivity. The death of animal also affects the availability of animal products like milk to the farm family especially in subsistence agriculture families. It also affects the human nutrition and thereby reduces the longevity. It also ends up in various social problems in the long run. The indirect impact of zoonotic diseases affects humans and has larger societal ramifications. The indirect financial impact includes lesser price for the diseased animals in the market or lower value for the normal animals also in a particular locality due to an outbreak. The price of the complement and substitute goods also changes due to an outbreak and hence the price effect should be a part of impact analysis.

Quantifying the direct impact of any disease is relatively easy compared to quantifying the indirect impact. However, with use of implicit assumptions and indirect valuation methods the overall impact can be quantified.

Information/factors required for assessing the impact of a disease

- i) The foremost information required is about the disease itself, i.e.
 - a) Whether the disease is bacterial or viral or parasitic or combination thereof or any other agent?
 - b) incidence rate of the disease
 - c) duration of infection
- ii) Diseased animal information
 - a) The species information (Eg. bovine, small ruminants, poultry, pig, horse, etc.)
 - b) The breed details (Eg. indigenous, crossbred, exotic)
 - c) The age and sex of the diseased animal
 - d) The susceptibility information agewise, sexwise, breedwise
- iii) The physiological parameters at normal and diseased state
- iv) Productive and economic traits at normal and diseased state

Indicators/data required for assessing the losses

The indicators required for assessing the losses due to animal diseases may vary for macro (national) or micro (farm) level estimation. The important indicators required for macro level disease impact estimates are as follows:

- a) The number of animals died in different species and across different ages
- b) The number of animals culled due to disease in different species and across different ages
- c) The number of animals disabled due to disease in different species and across different ages
- d) The productivity loss in animals like milk loss/day/animal
- e) The number of days of milking loss in a lactation etc.
- f) The number of days of draught power unavailability
- g) Information on decreased levels of fertility

Besides, the basic requirements listed above the economic information are also necessary to assess the impact of the disease. For time series disease impact assessment, the data on different time periods on various productivity impairment levels and price levels are necessary whereas for cross section analysis the information during a particular time on different variables are required.

Some of the very important economic information required are given below:

- a) Price of milk, wool, meat, etc.
- b) Price of live animal for different breeds, for different age groups and for different sex
- c) Price of the culled animal for different breeds, for different age groups and for different sex
- d) Price of disabled animal across breeds, age and sex
- e) Rates of bullock labour per day
- f) Rates of hired labour per day for male and female

Factors to be considered for projection of losses

1. Only the affected animal categories in the surveyed samples are to be considered for estimation.
2. Livestock inventory, affected, died and recovered animals of sample population are to be used for calculating mortality and morbidity rates.
3. For calculating the treatment cost and opportunity cost of labour for calves appropriate provision are to be made for different age categories and susceptibility levels before projections
4. Latest available census need to be considered for projections. If latest census is not available, then the population projections need be estimated before projecting the losses.

Conclusion

The analysis of losses due to livestock disease helps policy makers and planners. The estimation of losses species-wise helps to identify the disease severity between species by how much and what action need to be taken to mitigate. The disease loss information guide the policymakers in prioritizing the control options.

LIVESTOCK DISEASE DIAGNOSIS INFORMATICS

P. Krishnamoorthy and K.P. Suresh

ICAR-National Institute of Veterinary Epidemiology & Disease Informatics, Bengaluru, India

Diagnosing animal disease quickly and accurately has the economic benefit to the farmers. The expert system for animal disease diagnosing can meet the farms for urgent needs of veterinary experts, since there are very few experts at the farms (Wan and Bao, 2010). An expert system is a computer system that emulates the decision making ability of a human expert (Peter, 1998). The idea behind creating an expert system is that it is useful for many people from the knowledge of one person - the expert. Expert system simulates the judgment and behavior of a human that has expert knowledge and experience in a particular field. It consists of the inbuilt facilities to write the rules that build the knowledge base.

Nowadays farmers are very conscious about their health as well as health of their animals. The owner of the cattle has to observe the daily routine of the animals and check for any changes in their routine or behavior. Based on these observation, owners can decide that the animal is suffering from which disease and what treatment use for them. Most of the time it is difficult for the owners of the animals to take any action against the observation observed by them. Animal owners take help of book or other experienced owner or veterinarian. It is very time consuming and costly (Saurkar and Watane, 2012).

Most of the systems developed so far are based on Animal disease surveillance to improve the disease analysis, early warning and predicting disease emergence and spread (Patil et al., 2012). The expert system will be useful for the animal owners for healthcare of their animals and also for the veterinary doctors to take the advice from this system as it is expert and knowledge based system. The expert system is available to everyone and user can check only for the symptoms or signs observed in the animals and the disease name and treatment or control will be easily available to the owners.

Development of Screening tool/ device for diagnosis of cattle diseases:

To develop a screening tool or expert system to identify the cattle diseases on the farm level by Veterinarians

Identified clinical signs/symptoms for thirteen cattle diseases (Anaplasmosis, Anthrax, Babesiosis, Black quarter, Foot and Mouth Disease, Haemorrhagic septicemia, Theileriosis, Trypanosomiasis, Rabies, Infectious Bovine Rhinotracheitis, Leptospirosis, Brucellosis and Mastitis)



Prepared questionnaire containing 52 signs/symptoms for the thirteen cattle diseases for collecting the scores based on Likert scale scoring (1-10)



Face validity of questionnaire by pilot survey conducted with veterinarians or experts

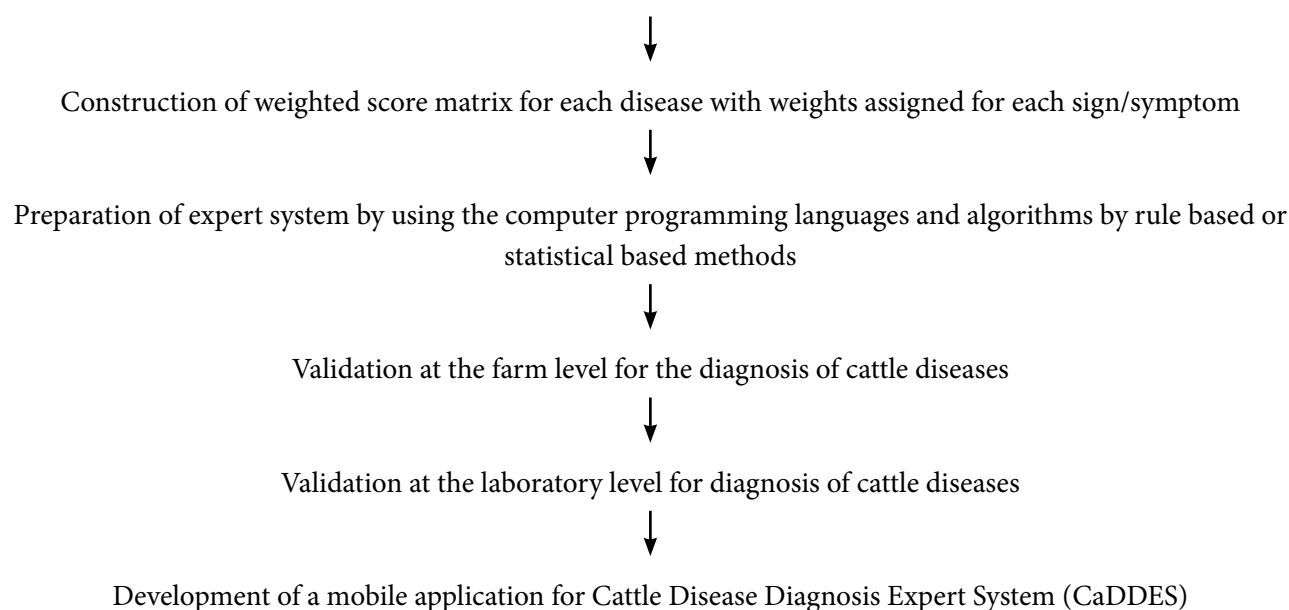


Scoring (1-10) for clinical signs on the priority based importance in screening a Cattle disease by Veterinarian



Content validity was done by analysis of scores given by Veterinarians using Aiken's value

Ranking of the scores for signs/symptoms



Guidelines for Scoring:

1. Clinical signs of their respective diseases are arranged in random
2. Score range will be 1-10
3. The highest score can be given to the most relevant and observed clinical signs at field conditions.
4. The least score can be given to the least relevant clinical signs
5. If there are two or more clinical signs appear as relevant for the particular disease at the same level can be given the same score.
6. If any clinical sign is not relevant to a particular disease you can give no score.

Content Validity:

The content validity was ascertained by calculation of Aiken's Value. Content validity, also known as logical validity refers to the extent to which a measure represents all facets of a given social construct. The generally accepted quantitative index for content is the Aiken's V index. This index will be used to quantify the ratings of panel experts constituted for evaluating the items in the instrument. The Aiken's V index with 0.80 indicates the good content validity of the measure. The Aiken's V index for content validity can be calculated using the formula given below (Aiken, 1980).

$$\text{Aiken's V is then } V = S / [n \cdot (c-1)]$$

Where S = the sum of s for the n raters

Let $s = r - l_o$

Let l_o = the lowest possible validity rating (usually, this is 1 on the Likert-scale)

Let r = the rating by an expert

n experts rate the degree to which the item taps an objective on a 1 to c on Likert-scale,

where c is the maximum score in grading scale

The range will be from 0 to 1.0, A score of 1.0 is interpreted as all Veterinarians giving the item the highest

possible rating.

The expert system for animal disease diagnosis is available for the dogs or canines mainly in many countries in the world, few expert system for swine in Thailand is available. But in India, there is no expert system available for the animal disease diagnosis and hence we are planning to develop expert system for the cattle diseases diagnosis in India.

References:

1. Aiken, L.R. (1980). Content validity and reliability of single items or questionnaires, *Educational and Psychological Measurement*, 40 : 955-959.
2. Patil AB and Kulkarni RV (2012). A Review of expert systems in Animal health care. *International Journal of computer Science and Information Technologies*, 3: 5014-5018.
3. Saurkar AV and Watane HN (2012). An expert system for disease diagnosis in pet. *Advances in Medical Informatics*, 2: 18-21.
4. Wan and Bao W (2010). Animal disease diagnosis expert system based on SVM. *Computer and Computing technologies in Agriculture III*, 317, Springer, pp:539-545.

Cohort studies in Epidemiology

Jagadish Hiremath

ICAR-National Institute of Veterinary Epidemiology & Disease Informatics, Bengaluru, India

Theory Topic

Introduction: Observational studies, as the name suggests it aids in making inferences about the effect of an exposure (infectious agent/toxin) or intervention (vaccination/treatment) on subjects which is observed in their natural setting rather than manipulated by the investigator. There are three measure observational study designs viz., cohort, case-control and cross-sectional study designs. These observational study designs are distinguished by the objective of the research study, how subjects are sampled, and the timeline of data collection. This write-up describes the Cohort study design, advantages and disadvantages, types of cohort studies, methodological issues, calculating the basic measure (Relative risk), and appreciate the strength and weaknesses.

Cohort Study: A study design, where the investigator identifies the subjects at a point in time when they are free of an outcome or disease and compare the incidence of the outcome of interest among groups of exposed and unexposed subjects (Fig.1). The groups compared are called cohorts and these can be identified retrospectively or prospectively and in either case the outcome status needs to be established at least twice. This is an appropriate study design under following situations

1. There is good indication of having an association between an exposure and an outcome, many times based on the observations of cross-sectional studies
2. The duration of time to see the outcome after exposure is relatively short so as to reduce the loss of subjects during follow-up
3. The outcome should not be too rare as this impacts the size of the cohort

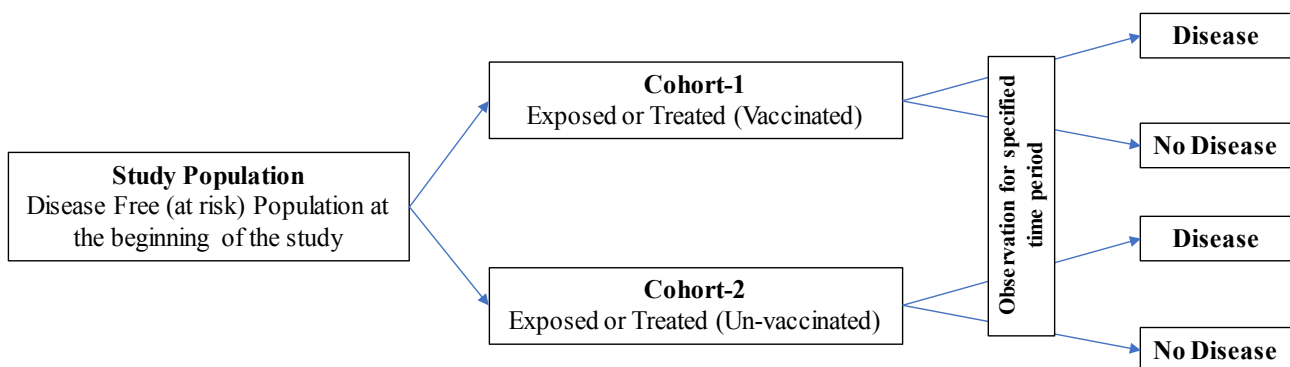


Fig.1 Cohort Study Design

Advantages:

1. For a given exposure multiple outcomes can be measured.
2. Exposure is measured before the onset of disease (in prospective cohort studies).
3. Good for measuring rare exposures, for example among different occupations.
4. Demonstrate direction of causality.
5. Can measure incidence and prevalence.

Disadvantages:

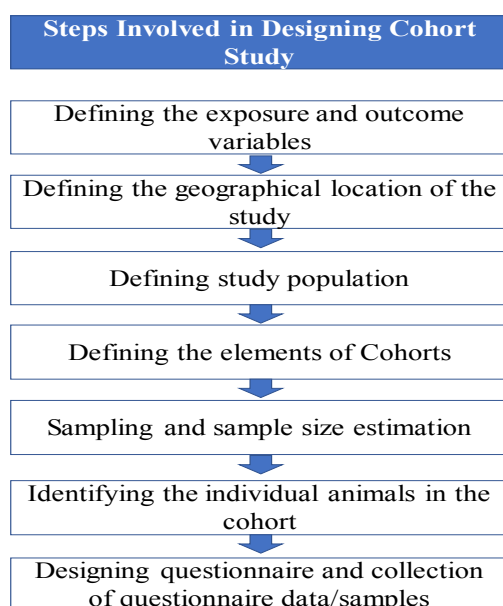
1. Study involves more time and hence resource.
2. Prone to bias due to loss to follow-up.
3. Prone to confounding.
4. This may not be appropriate choice for very rare disease since this will increase the size of the cohorts.
5. Classification of individuals (exposure or outcome status) can be affected by changes in diagnostic procedures.

Types of Cohort study designs: The cohort studies may be prospective or retrospective. In prospective cohort studies the groups of animals are chosen that do not have the outcome of interest (Disease). The investigator then measures a variable that might be relevant to the development of outcome of interest over a period of time. When study uses two cohorts, one group is exposed to or treated with agent of interest and other is unexposed or un-treated which acts as control.

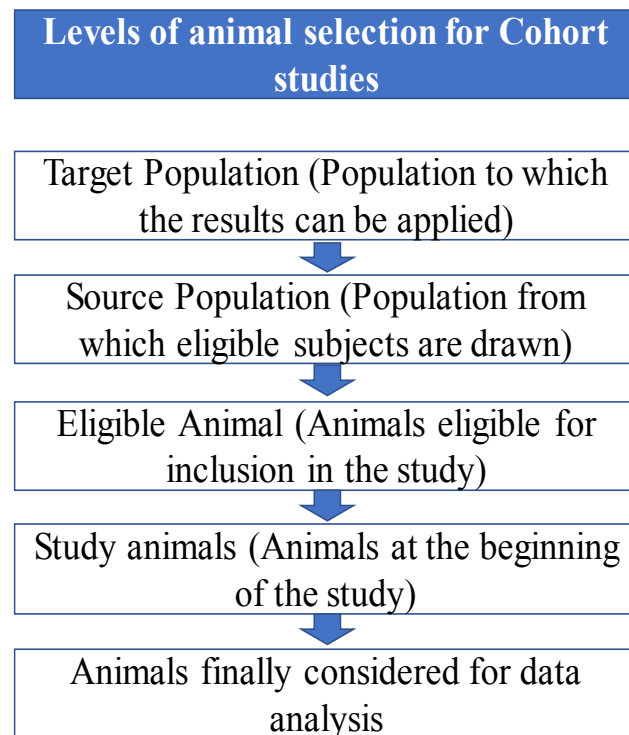
Retrospective cohort studies use data already collected for other purposes but the methodology is same except for the study is performed retrospectively as the cohorts are followed-up retrospectively. The study period may be many years but the time to complete the study is subjected to time required to collate and analyze the data.

Methodological Issues: Selection of subjects (Animal) and attrition bias are two major methodological issues to be addressed while planning for Cohort study. Selection of animals involves defining the selected groups of animals by exposure at the beginning of the study. A critical step in the study is to select the exposed and unexposed group of animals from the same source population having equal probability risk of developing outcome. A sampling technique and sample size estimation will aid in final identification of animals for each group (exposed and unexposed). Attrition bias is another major issue to be addressed as the selected animals may be sold or dead during the observation period. This is more an issue in prospective cohort studies where it requires long follow-up periods. This can be overcome by selecting the number of animals in each group after adjusting for such losses. In veterinary epidemiological studies timing of the study has greater influence for such losses as the rate of animal selling is more in certain times of the year.

General Steps involved in Epidemiological study design with special reference to Cohort Study:



Levels of animals selected for Cohort Study:



Typical Observations recorded from Cohort Studies: Analysis of a cohort study uses either the relative risk of disease in the exposed cohort compared with the risk in the unexposed cohort. At the end of the study the data obtained will be as shown in the 2X2 contingency table shown below along with formula to calculate the relative risk.

		Outcome	
		Yes	No
Exposure	Yes	A	B
	No	C	D

$$\text{Relative Risk (RR)} = \text{RR} = \frac{(A/(A+B))}{(C/(C+D))}$$

Higher the value of RR stronger is the association between exposure factor and outcome.

Exercise:

Part-I: Vet-bio, a veterinary biological company located in Surugur, India which produces range of veterinary vaccines for 15 years. Porcine reproductive and respiratory syndrome (PRRS) vaccine is recent addition to their list of vaccines. The vaccine is known to have efficacy of 95% which is based on vaccine efficacy studies using 2-3months of Yorkshire pigs that are C-section delivered.

The company has decided to do a study in Mizoram to estimate the vaccine efficacy at field level. If you have been asked to give a consultation service to design, plan and execute the study, explain how you go about providing the service.

Part-II:

Total number of vaccinated animals: 800

Total number of unvaccinated animals:800

Period of observation: 6 months

Number of cases in vaccinated: 100

Number of cases in unvaccinated:400

Based on the above observation prepare 2X2 contingency table and estimate the relative risk of getting disease in vaccinated as against unvaccinated

References:

1. Jae W. Song and Kevin C. Chung (2010) **Observational Studies: Cohort and Case-Control Studies**. Plast Reconstr Surg. 2010 Dec; 126(6): 2234–2242.
2. Melissa D.A. Carlson and R. Sean Morrison (2009) **Study Design, Precision, and Validity in Observational Studies**. J Palliat Med. 2009 Jan; 12(1): 77–82.
3. <https://www.healthknowledge.org.uk/e-learning/epidemiology/practitioners/introduction-study-design-cs>
4. http://sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/EP713_CohortStudies/EP713_CohortStudies_print.html
5. <http://dx.doi.org/10.1136/emj.20.1.54>
6. Methods Future Research Needs Reports, No. 8. Carey TS, Sanders GD, Viswanathan M, et al. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012 Mar.

An Introduction to case-control study in veterinary epidemiology

Md. Mudassar Chanda

ICAR-National Institute of Veterinary Epidemiology & Disease Informatics, Bengaluru, India

Case-control study is a retrospective study to find the relative risk between a specific exposure (e.g. carcass disposal) and an outcome (e.g. Anthrax). A control group of animals which do not have the disease or was exposed to the particular risk factor is used for comparison with animals which did not come down with the disease or not exposed. The aim of case-control studies is to identify the relationship between risk factors (Odd's ratio) and disease or the outcome variable. The following steps can be followed in Case-control studies

1. Enrolling animals which have a certain disease or outcome are considered to be cases
2. Control animals of similar size can be sampled randomly from an identical population without the disease or outcome
3. Data collection
4. Data is entered and analysed to calculate Odd's ratio

There are two types of case-control studies

Non-matched case-control study: Animals with and without disease are enrolled in the study and then their exposure status is determined.

Matched case-control study: Identify animals with the disease and enrolled them in the study. Animals with matched criteria (e.g. same age, same breed etc) are then enrolled in the study. Matched case-control studies are conducted to minimize confounding variables.

Advantages of case-control study

1. Case-control studies are preferred in case of rare diseases as cohort will involve selection of more number of animals selected and may miss the animals with disease.
2. The study does not require long time waiting as in cohort study for events/disease to happen.
3. The study is relatively in-expensive compared to cohort study
4. Multiple risk factors can be studied simultaneously
5. Case-control studies are useful in outbreak investigations to identify the association between the risk factors and the disease with smaller sample size
6. It is stronger than cross-sectional studies for establishing causation of the disease and their contributing risk factor

Disadvantages in case control study

1. Control group is difficult to identify in case of homogenous population, which is common in a village with similar management practices and environmental conditions.
2. It can be affected by recall bias where farmers whose animals are affected by the disease are likely to remember compared to the farmers whose animals did not come down with the disease.
3. Case-control study is weaker for establishing a causation compared to cohort study
4. The results (Odd's ratio) cannot be generalized to other areas and/or population

Sample Size Estimation:

The total number of sample size to study the various risk factors associated with a particular disease occurrence can be carried out using software Epi-Info version 7.0 on the basis of following assumptions as per the description.

For example

Power = 80%

Odds Ratio (OR) = 4

An equal number of cases and control (1:1)

The proportion exposed in the control group = 30%

The calculated total sample size comes to **82**, i.e. 41 for each group (Case: Control)

Exercise to identify risk factors for occurrence of Anthrax

The case study was developed by using hypothetical values for the risk factors and hence is not real risk factors. However, the probable risk factors and to create hypothetical data we referred this paper (Mongoh, M. N., Dyer, N. W., Stoltenow, C. L., & Khaita, M. L. (2008). Risk factors associated with anthrax outbreak in animals in North Dakota, 2005: A retrospective case-control study. *Public Health Reports*, 123(3), 352-359.)

Case study: Anthrax outbreak occurred from July 1 to October 12 2005 in 109 farms of 16 counties. The outbreaks were confirmed by Laboratory diagnosis (staining, isolation and PCR). Questionnaire was developed and sent to 419 farms, which included 109 farms confirmed for Anthrax and 130 farms negative for Anthrax. Out of 419 questionnaires 137 responded, which included 52 confirmed farms (Cases) and 85 negative farms (Control). Species affected were Bison, Cattle, Horses, Elk, Sheep, Deer.

Risk factors: 23 risk factors were considered for the questionnaire preparation and out of which 10 were selected as important by selection method. The answers of the 10 variables are given to calculate Odd's ratio and your interpretation.

Study design: Retrospective case control study

Total respondents: 137

Cases: 52

Controls: 85

Exercise: Farms which were exposed to the risk factor and not exposed to the risk factor is given in the table.

- a) Calculate the values of controls for farms exposed and unexposed to the risk factors
- b) Calculate the Odd's ratio for the risk factor
- c) Interpret the Odd's ratio

1. **Death reported on neighbouring pasture:** 31 farmers respondent reported death on neighbouring pasture and 21 farmers not reported death on neighbouring pasture

Risk factor	Anthrax occurred		Total
	Cases	controls	
Exposed	31		50
Unexposed	21		87
Total	52	85	137

Odd's ratio:

Interpretation:

2. **Animal moved off pasture:** 25 farmers moved their animals off the pasture and 27 farmers did not move their animals off pasture

Risk factor	Anthrax occurred		Total
	Cases	controls	
Exposed	25		44
Unexposed	27		93
Total	52	85	137

Odd's ratio:

Interpretation:

3. **Antibiotic use:** 23 farmers used antibiotic and 29 farmers did not use antibiotic

Risk factor	Anthrax occurred		Total
	Cases	controls	
Exposed	23		42
Unexposed	29		95
Total	52	85	137

Odd's ratio:

Interpretation:

4. **Carcass of unknown origin:** 5 farmers reported carcass of unknown origin and 47 farmers did not report carcass of unknown

Risk factor	Anthrax occurred		Total
	Cases	controls	
Exposed	5		11
Unexposed	47		126
Total	52	85	137

Odd's ratio:

Interpretation:

5. **Dry conditions:** 12 farmers reported dry conditions on the farm and 40 farmers did not report dry conditions on the farm

Risk factor	Anthrax occurred		Total
	cases	controls	
Exposed	12		25
Unexposed	40		112
Total	52	85	137

Odd's ratio:

Interpretation:

6. **Multiple vaccinations:** 29 farmers used multiple vaccinations and 23 farmers did not use multiple vaccination

Risk factor	Anthrax occurred		Total
	cases	controls	
Exposed	29		48
Unexposed	23		89
Total	52	85	137

Odd's ratio:

Interpretation:

7. **Presence of burial site:** 7 farmers had burial site on the farm and 45 farmers did not had burial site at the farm

Risk factor	Anthrax occurred		Total
	cases	controls	
Exposed	7	8	15
Unexposed	45	77	122
Total	52	85	137

Odd's ratio:

Interpretation:

8. **Standing water** : 33 farmers had standing water on their farm and 19 farmers did not had standing water

Risk factor	Anthrax occurred		
Standing water	cases	controls	Total
Exposed	33	34	67
Unexposed	19	51	70
Total	52	85	137

Odd's ratio:

Interpretation:

9. **Vaccination period**: 33 farmers vaccinated after the outbreak and 19 farmers vaccinated before the outbreak

Risk factor	Anthrax occurred		
Vaccination period	cases	controls	Total
Exposed	33	42	75
Unexposed	19	43	62
Total	52	85	137

Odd's ratio:

Interpretation:

10. **Wet conditions**: 36 farmers had wet farms compared to 16 farmers which did not had wet farms

Risk factor	Anthrax occurred		
Wet conditions	cases	controls	Total
Exposed	36	54	90
Unexposed	16	31	47
Total	52	85	137

Odd's ratio:

Interpretation:

Spatial data analysis using R software- A Practical Approach

Rashmi R. Kurli, Dheeraj R, Mainak Mondal and K P Suresh

ICAR-National Institute of Veterinary Epidemiology & Disease Informatics, Bengaluru, India

I. R Language

A. Introduction

R is a programming language and software environment for statistical analysis, graphics representation and reporting. R was created by Ross Ihaka and Robert Gentleman at the University Of Auckland, New Zealand, and is currently maintained by the R Development Core Team.

The programming language was named **R**, based on the first letter of first name of the two R authors (Robert Gentleman and Ross Ihaka), and partly a play on the name of the Bell Labs Language **S**. R is an integrated suite of software facilitating for data manipulation, calculation and graphical display, it also has an effective data handling and storage facility

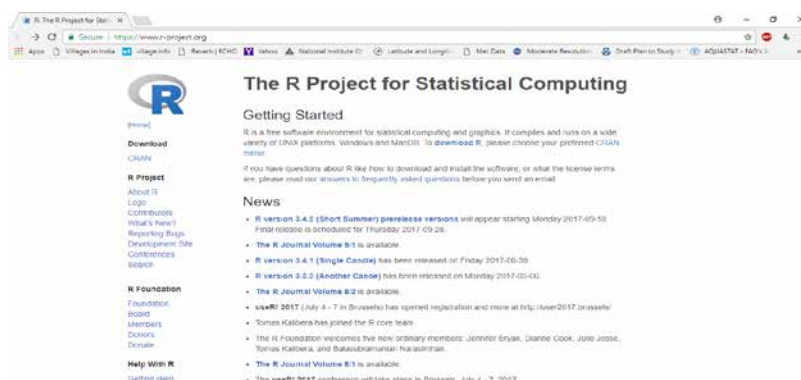
B. Installation Protocol

i. R Software installation

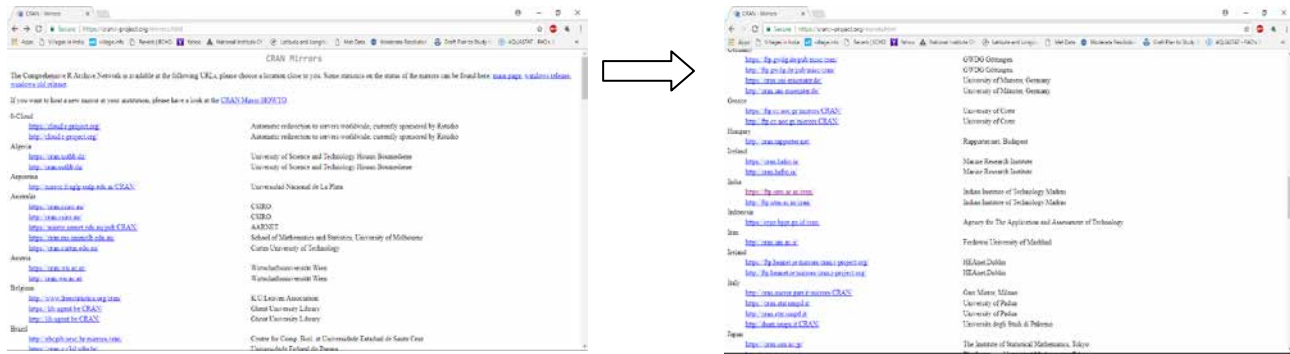
1. Open an internet browser and go to www.r-project.org. and Click the “download R” link in the middle of the page under “Getting Started.”



2. Select a CRAN location (a mirror site) and click the corresponding link.



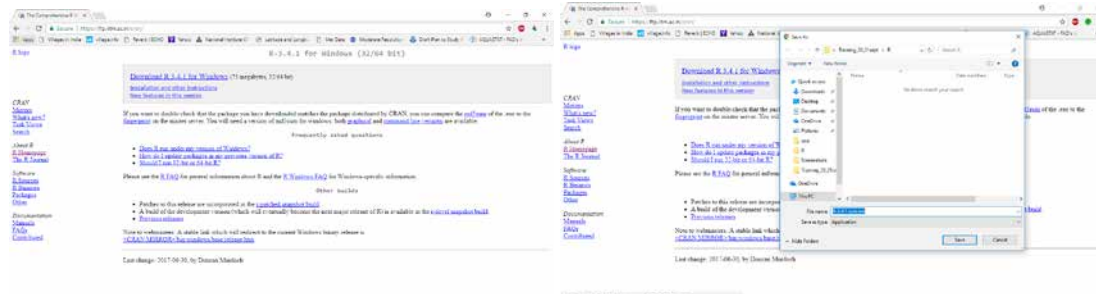
3. Scroll down to select the closest place for you



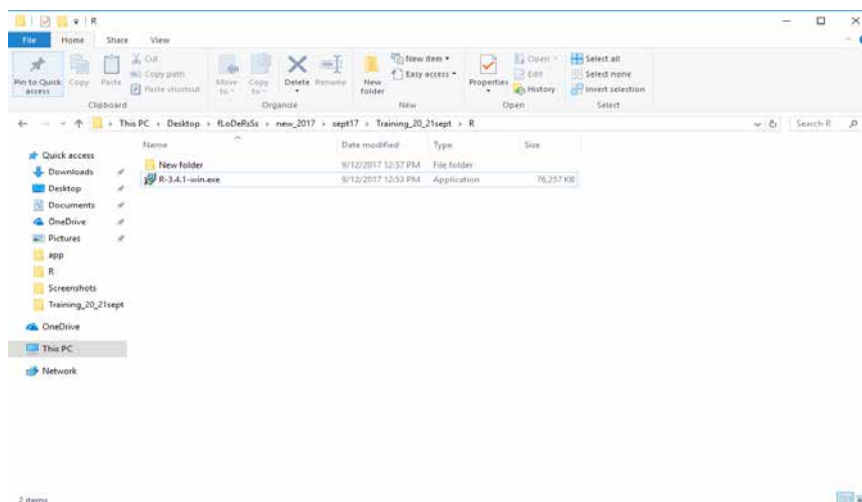
4. Click on the link “Download R for Windows” for suitable Operating System.



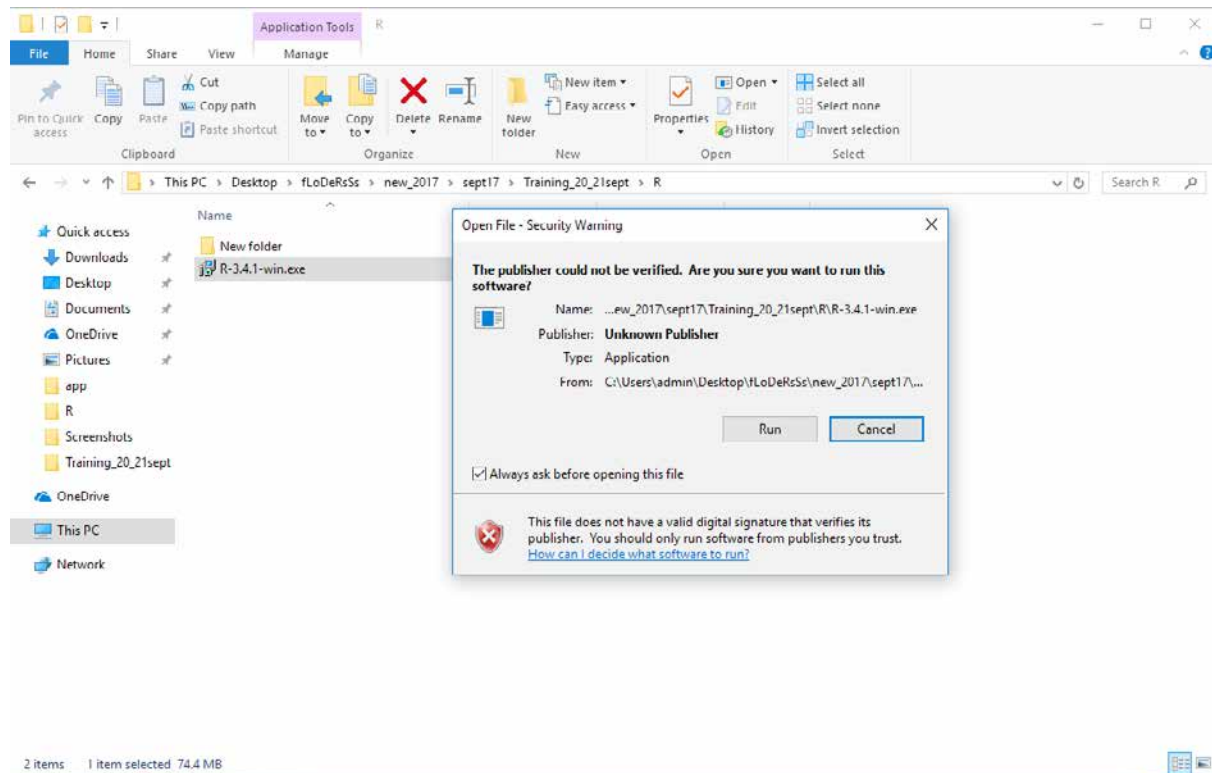
5. Click on install R software and Save the file in the desired folder



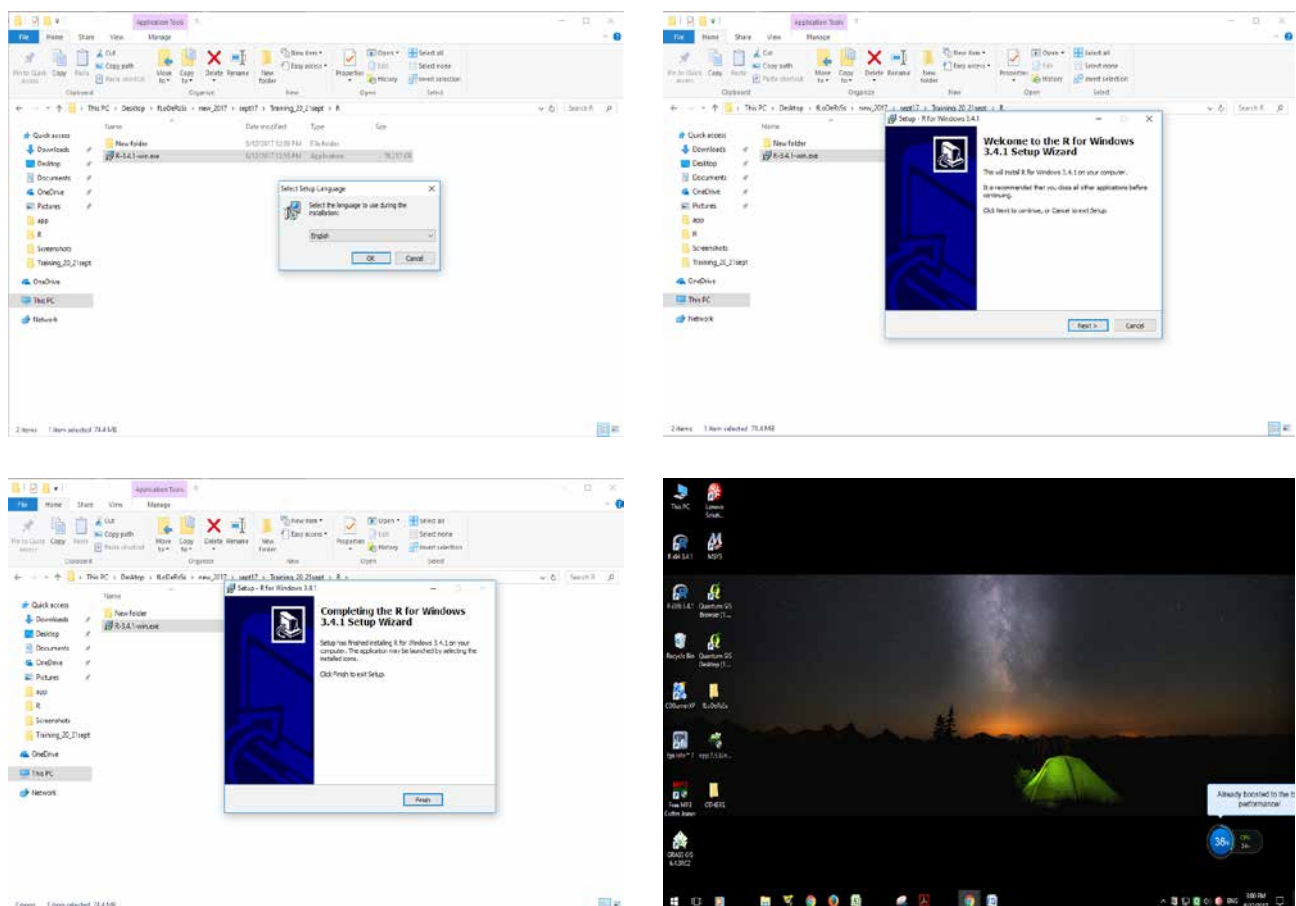
6. Click on the file



7. Run it

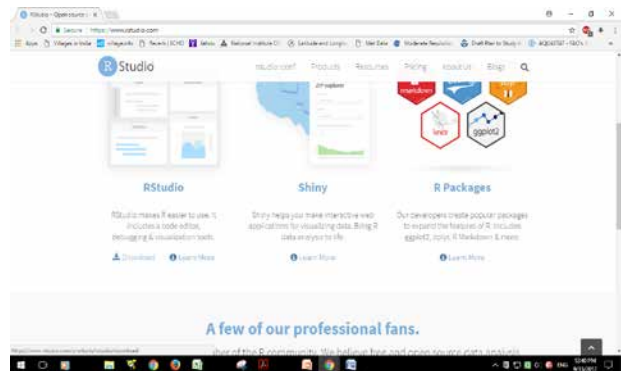
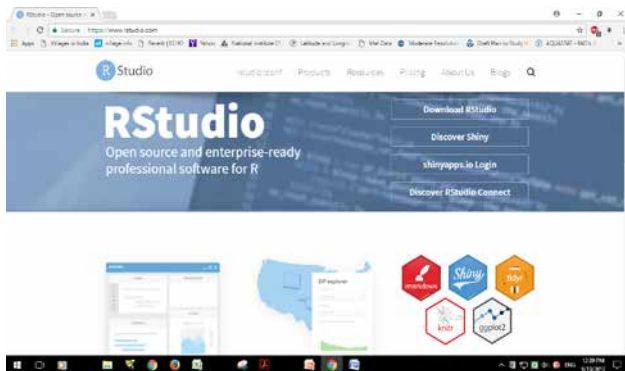


8. Follow the instructions to install the software on your PC

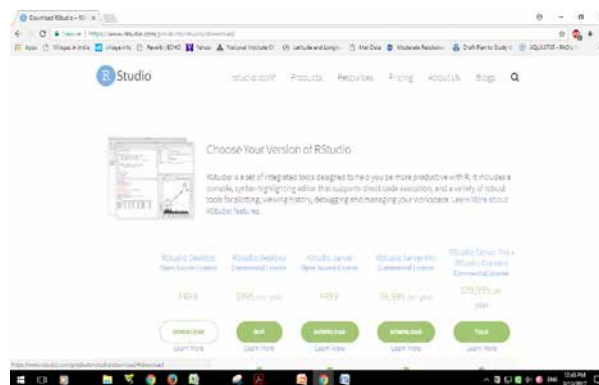


ii. R Studio Installation

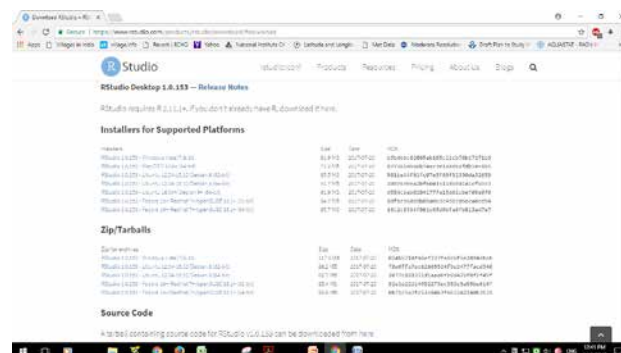
1. Go to www.rstudio.com and Click on the “Download RStudio” button.



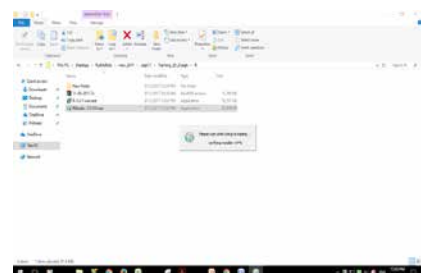
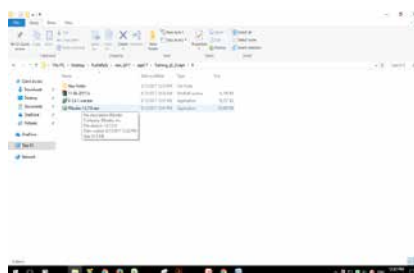
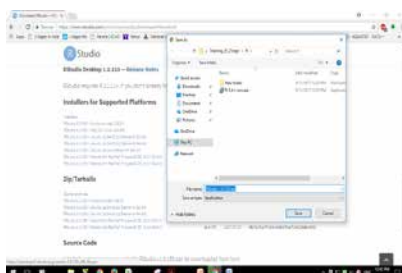
2. Click on “Download RStudio Desktop.”



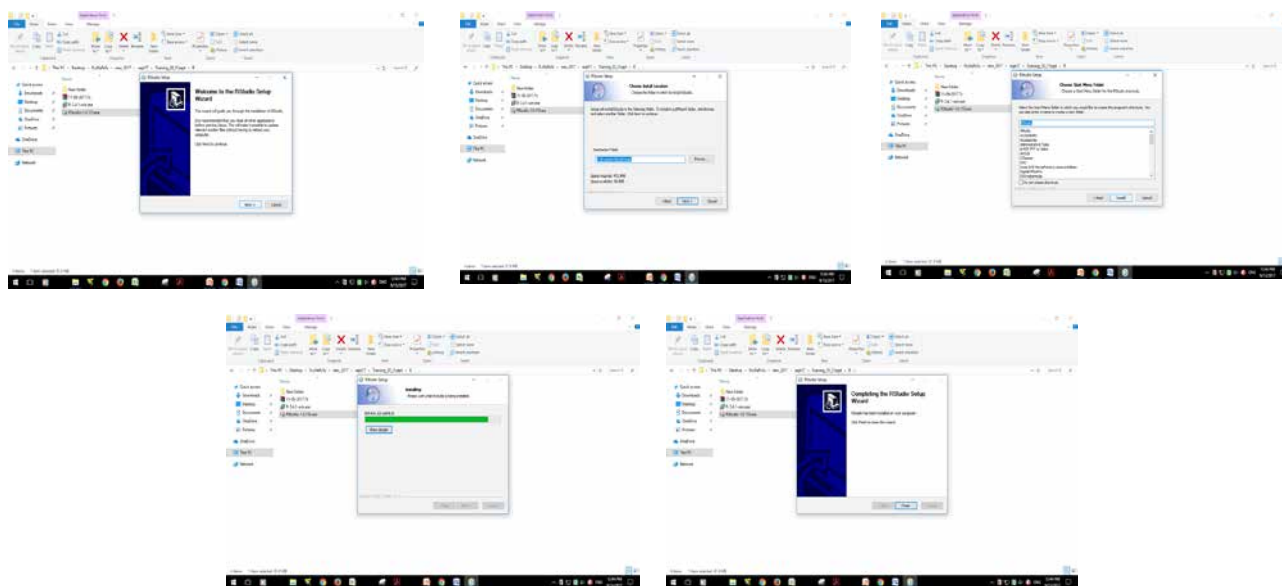
3. Click on the version recommended for your system, or the latest Windows version



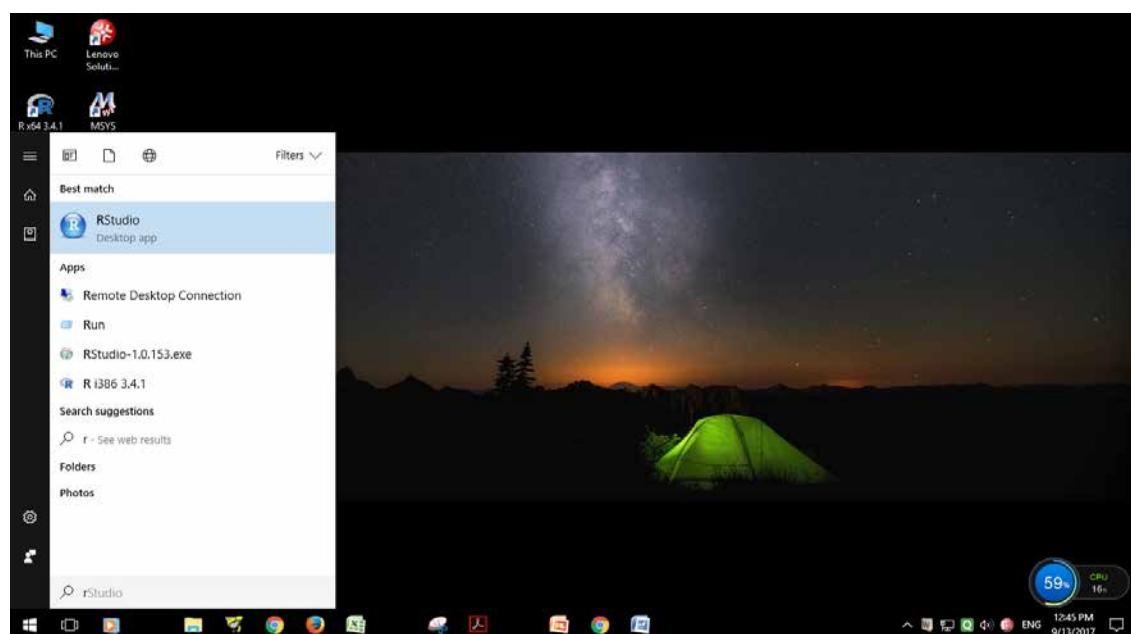
4. Save the executable file and Run the .exe file



5. follow the installation instructions and click on finish option when the window pops up



6. Now the R studio is installed on your PC



II. Geographic Coordinate System

A. Introduction

A geographic coordinate system (GCS) uses a three-dimensional spherical surface to define locations on the earth. A point is referenced by its longitude and latitude values. Longitude and latitude are angles measured from the earth's center to a point on the earth's surface. The following illustration shows the world as a globe with longitude and latitude values.

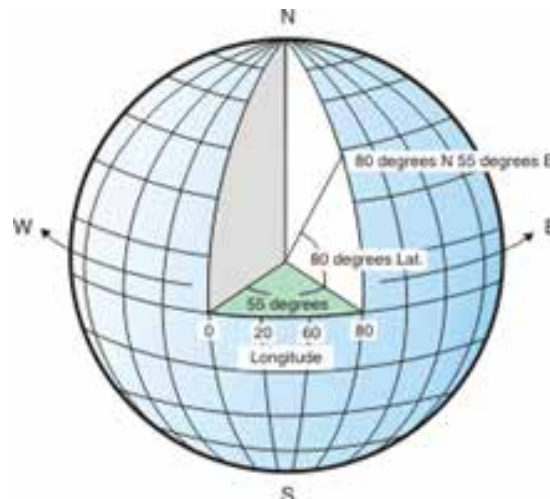


Fig.1 Illustration showing longitude and latitude values in the world as a globe

In the spherical system, horizontal lines, or east–west lines, are lines of equal latitude, or parallels. Vertical lines, or north–south lines, are lines of equal longitude, or meridians. These lines encompass the globe and form a gridded network called a graticule.

The line of latitude midway between the poles is called the equator. It defines the line of zero latitude. The line of zero longitude is called the prime meridian. For most geographic coordinate systems, the prime meridian is the longitude that passes through Greenwich, England. Other countries use longitude lines that pass through Bern, Bogota, and Paris as prime meridians. This illustration shows the parallels and meridians that form a graticule.

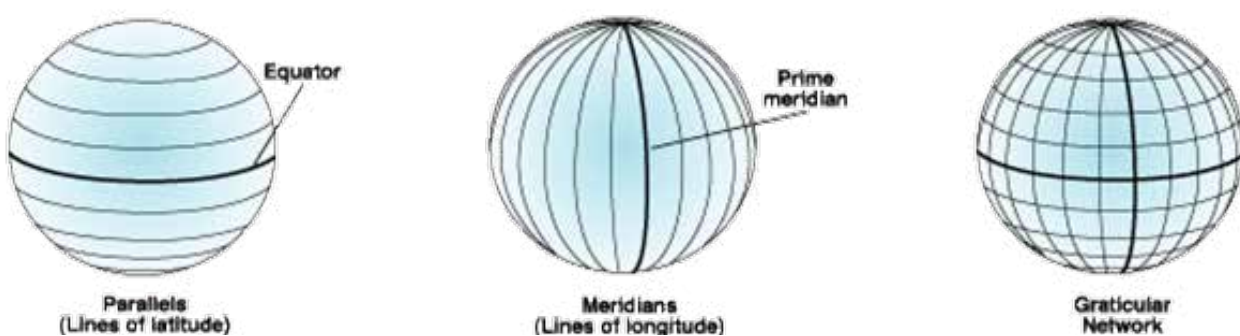
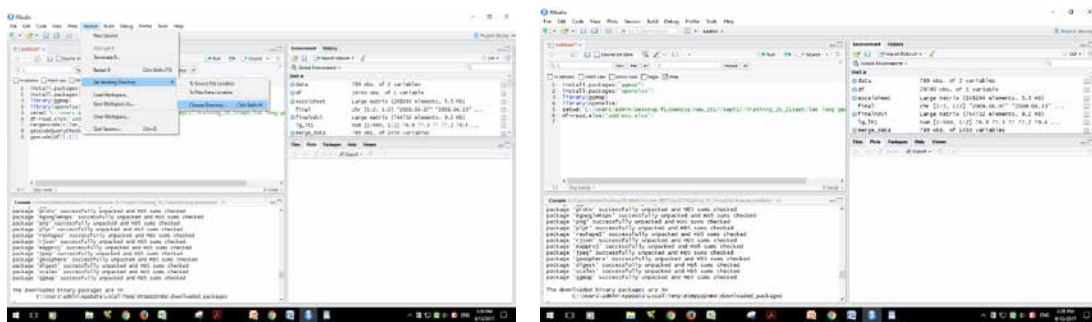


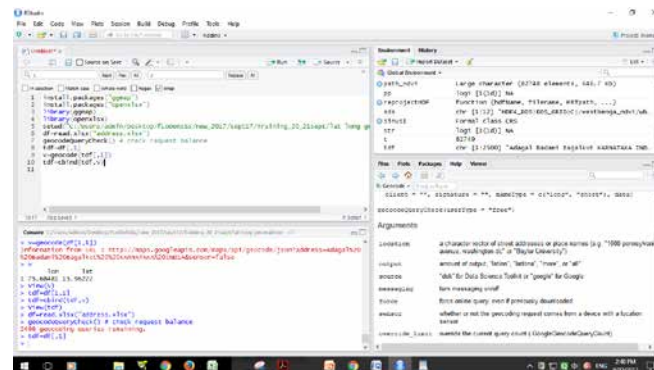
Fig.2 The parallels and meridians that form a graticule.

Latitude and longitude values are traditionally measured either in decimal degrees or in degrees, minutes, and seconds (DMS). Latitude values are measured relative to the equator and range from -90° at the South Pole to $+90^\circ$ at the North Pole. Longitude values are measured relative to the prime meridian. They range from -180° when traveling west to 180° when travelling east. If the prime meridian is at Greenwich, then Australia, which is south of the equator and east of Greenwich, has positive longitude values and negative latitude values. It may be helpful to equate longitude values with X and latitude values with Y.

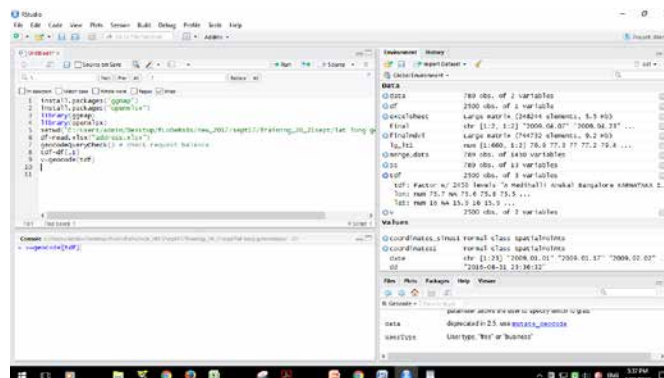
- Go to session in the menu bar and click on set working directory and select choose directory and read the input file



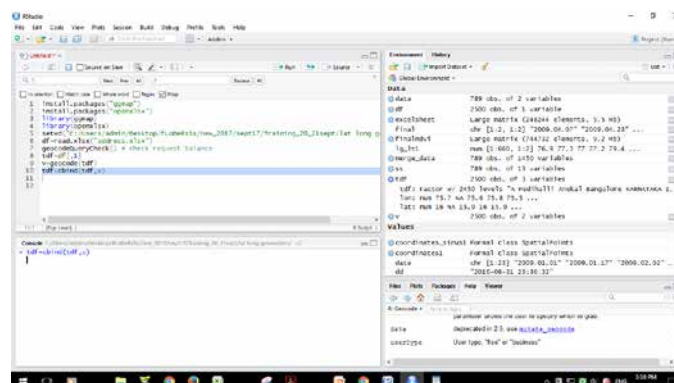
- `geocodeQueryCheck()` tells the number of request remaining to generate latitude and longitude.



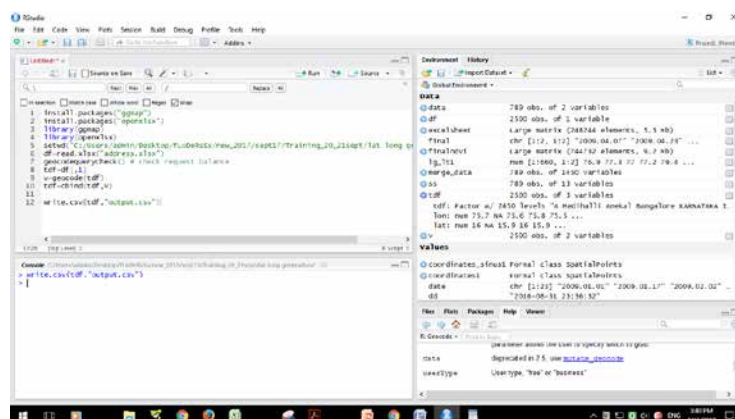
- `geocode()` generates the lat long for the given address.



- the generated lat long data gets combined with the address provided.



9. Write the output to a csv file. This will be saved in the selected directory.



III. Remote sensing and GIS

A. Introduction

Remote sensing (RS), also called earth observation, refers to obtaining information about objects or areas at the Earth's surface without being in direct contact with the object or area. The **science of acquiring information about the earth** using instruments which are remote to the earth's surface, usually from aircraft or satellites. Instruments may use visible light, infrared or radar to obtain data. Remote sensing offers the ability to observe and collect data for large areas relatively quickly, and is an important source of data for GIS. Remote Sensing is basically a multi-disciplinary science which includes a combination of various disciplines such as optics, spectroscopy, photography, computer, electronics and telecommunication, satellite launching etc. All these technologies are integrated to act as one complete system in itself, known as Remote Sensing System.

Applications of Remote Sensing widespread in many fields Oceanography, Glaciology, Geology, Topography and cartography, Agriculture, Hydrology, Oil and mineral exploration and Climate.

i. NDVI (Normalised Difference Vegetative Index)

Remote sensing phenology studies use data gathered by satellite sensors that measure wavelengths of light absorbed and reflected by green plants. To determine the density of green on a patch of land, researchers must observe the distinct colours (wavelengths) of visible and near-infrared sunlight reflected by the plants. When sunlight strikes objects, certain wavelengths of this spectrum are absorbed and other wavelengths are reflected. The pigment in plant leaves, chlorophyll, strongly absorbs visible light (from 0.4 to 0.7 μm) for use in photosynthesis. The cell structure of the leaves, on the other hand, strongly reflects near-infrared light (from 0.7 to 1.1 μm).

Many sensors carried aboard satellites measure red and near-infrared light waves reflected by land surfaces. Using mathematical formulas (algorithms), the raw satellite data about these light waves is transformed into vegetation indices. A vegetation index is an indicator that describes the greenness, the relative density and health of vegetation for each picture element, or pixel, in a satellite image.

Calculations of NDVI

NDVI is calculated from the visible and near-infrared light reflected by vegetation. Healthy vegetation (left) absorbs most of the visible light that hits it, and reflects a large portion of the near-infrared light. The Normalized Difference Vegetation Index (NDVI) can be calculated by

$$NDVI = (NIR - VIS) / (NIR + VIS)$$

Calculations of NDVI for a given pixel always result in a number that ranges from minus one (-1) to plus one (+1); however, no green leaves gives a value close to zero. A zero means no vegetation and close to +1 (0.8 - 0.9) indicates the highest possible density of green leaves.

i. Land Surface Temperature

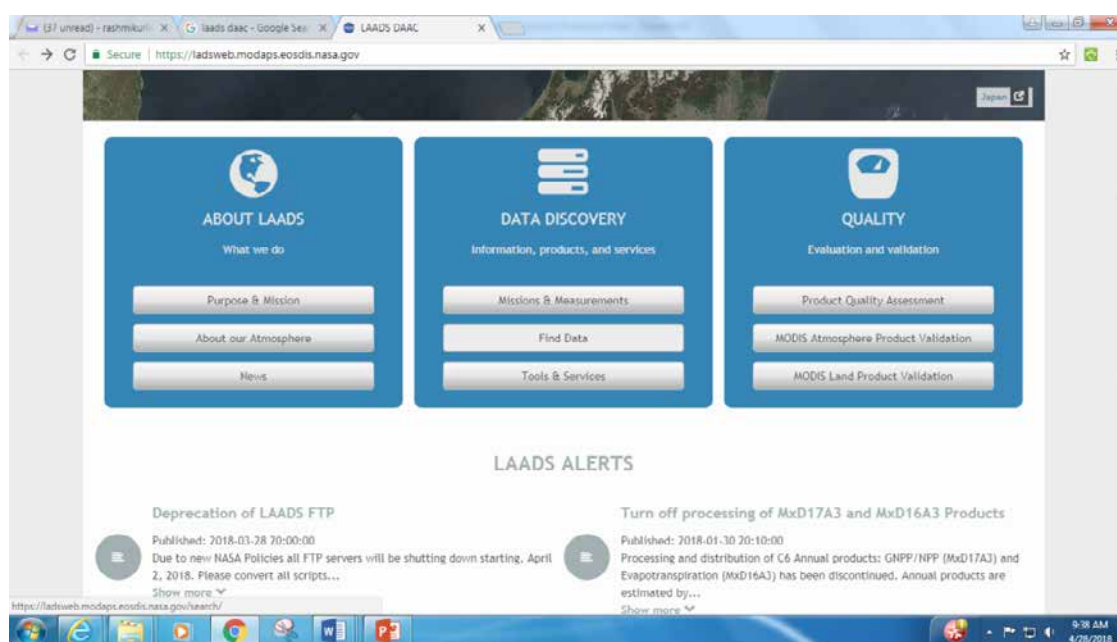
Land surface temperature is how hot the “surface” of the Earth would feel to the touch in a particular location from a satellite’s point of view, the “surface” is whatever it sees when it looks through the atmosphere to the ground. It could be snow and ice, the grass on a lawn, the roof of a building, or the leaves in the canopy of a forest. Thus, land surface temperature is not the same as the air temperature that is included in the daily weather report

The data is collected by the Moderate Resolution Imaging Spectroradiometer (MODIS) on NASA’s Terra satellite. Temperatures range from -25 degrees Celsius (deep blue) to 45 degrees Celsius (pinkish yellow). At mid-to-high latitudes, land surface temperatures can vary throughout the year, but equatorial regions tend to remain consistently warm, and Antarctica and Greenland remain consistently cold.

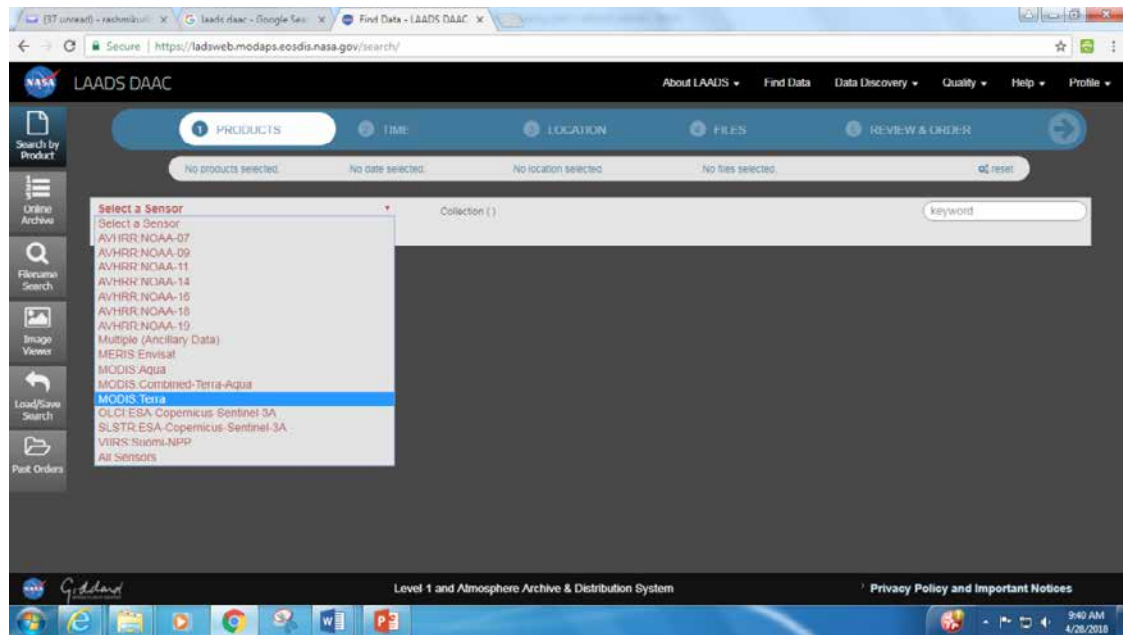
B. Protocol to Generate NDVI and LST

The HDF files for LST (°C), NDVI were downloaded from the MODIS website using the MOD11A2 and MOD13A1 products respectively by specifying the coordinates and time period (dates). HDF files were then converted to TIF files using gdalUtils package of R software. The pixel values were converted to index value for NDVI (pixel value X 0.0001) and LST were converted to degree centigrade (pixel value X 0.002- 273.15 Kelvin). In NDVI the index values were considered negative for water, 0-0.1 for rock, soil and barren land. 0.2- 0.4 was taken as low vegetation, 0.41-0.6 as moderate and 0.6-0.8 as high vegetation.

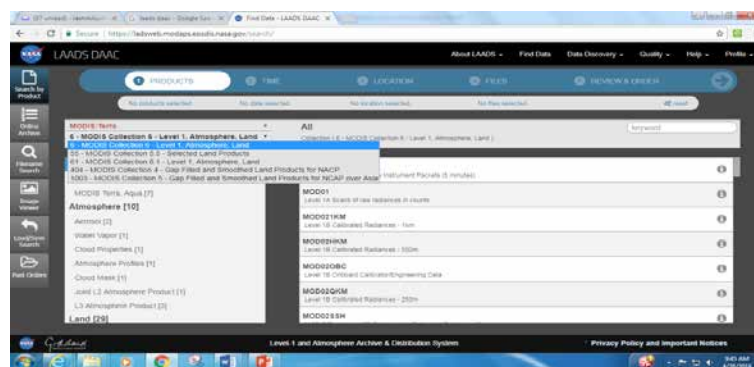
1. Arrange the excel sheet(CSV/XLSX) comprising latitude & longitude and its address.
2. Install the following Packages in R software “rjava”, “ raster” , “Rcurl” and “gdalUtils”
3. Download the Data from an Online Source LAADS DAAC- Click on Find data



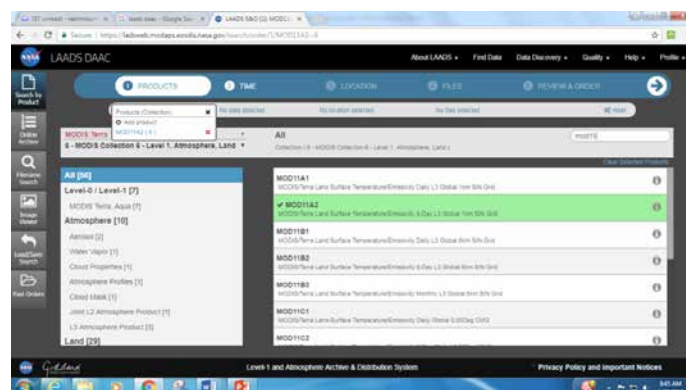
4. Select MODIS:TERRA Sensor.



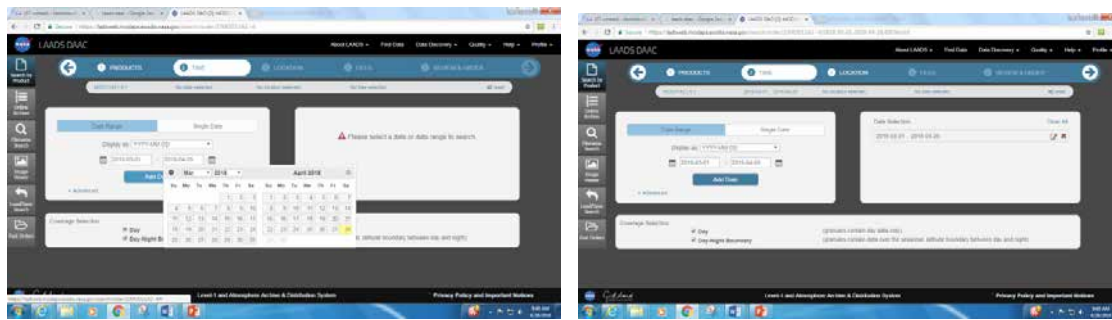
5. Select the collection- MODIS collection 6 Atmosphere, Land



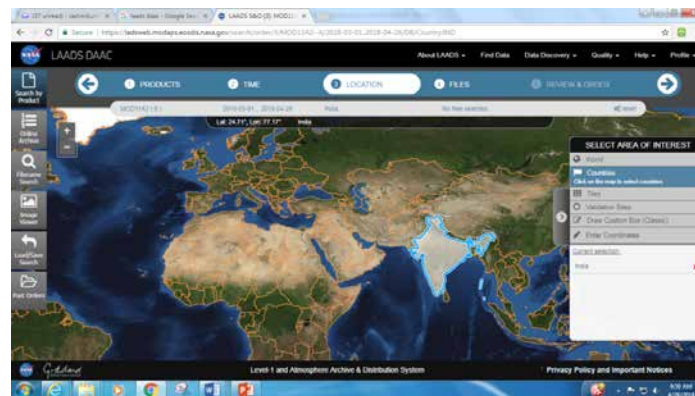
6. Select the product MOD11A2 for LST and MOD13A1 for NDVI



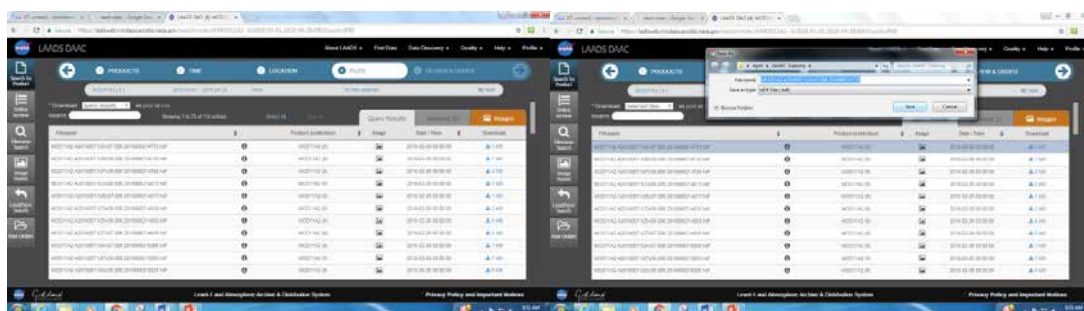
7. Specify the time period to download the values and click on add date



8. Select the location by specifying the area or selecting the country boundaries



9. The files will appear in the Files menu which can be downloaded by clicking on it



10. TIF file generation

library(gdalUtils)

library(MODIS)

files <- list.files(path="LST",pattern = glob2rx("*.hdf"),full.names = T) #specify the directory name containing HDF files

j<-length(files)

date=extractDate(files,asDate = T)

filename <- paste0("LST/", substr(files,23,28),date\$inputLayerDates,"tif") # specify the directory name to store TIF files

i <-1

while(i<=j){

 sds <- get_subdatasets(files[i]);

 gdal_translate(sds[1], dst_dataset = filename[i]); # sds[1] LST

```

i<-i+1;
}

```

11. NDVI and LST measurement

```

ss<-read.csv("Points/BNG.csv",sep=";",header=T,check.names = F) # Specify the
geocoordinates file
#filename <- list.files(path="LST/",pattern = ".tif",full.names = T) #specify the directory name
containing TIF files
x<-ss$lat
y<-ss$long
data<-data.frame(y,x)
latlon1<-CRS('+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0')
coordinates1 = SpatialPoints(data,latlon1)
sinus1 = CRS("+proj=sinu +lon_0=0 +x_0=0 +y_0=0 +a=6371007.181 +b=6371007.181 +units=m
+no_defs")
coordinates_sinus1 = spTransform(coordinates1,sinus1)
#my<-raster(filename[2])
my=raster('LST/26v06.2018-02-26.tif')
dd<-extract(my,coordinates_sinus1)
dd=dd-273.15
df=cbind(data,dd)
write.csv(df,"BNG_LST.csv") #Output file name

#NDVI Measurements
ss<-read.csv("Points/BNG.csv",sep=";",header=T,check.names = F) # Specify the
geocoordinates file
#filename <- list.files(path="NDVI/",pattern = ".tif",full.names = T) #specify the directory name
containing TIF files
x<-ss$lat
y<-ss$long
data<-data.frame(y,x)
latlon1<-CRS('+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0')
coordinates1 = SpatialPoints(data,latlon1)
sinus1 = CRS("+proj=sinu +lon_0=0 +x_0=0 +y_0=0 +a=6371007.181 +b=6371007.181 +units=m
+no_defs")
coordinates_sinus1 = spTransform(coordinates1,sinus1)
#my<-raster(filename[1])
my=raster('NDVI/26v06.2018-03-06.tif')
dd<-extract(my,coordinates_sinus1)
dd=dd*0.00000001
df=cbind(data,dd)
write.csv(df,"BNG_NDVI.csv") #Output file name

```

12. Output file containing NDVI /LST values will be saved in the set directory.

IV. Meteorological Parameters

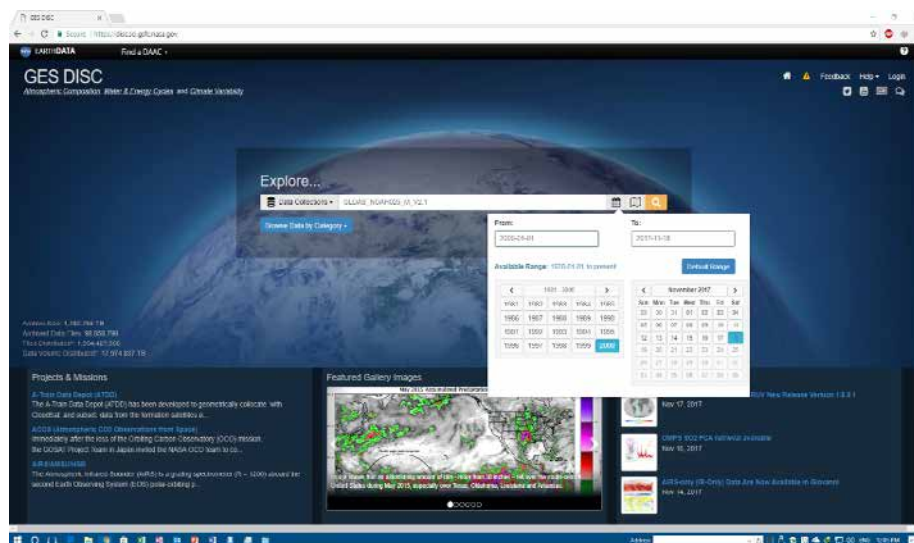
A. Introduction

All meteorological parameters are subject to short-term variations, normally caused by turbulences within the atmosphere. They are influenced by solar radiation, directly or indirectly, and this results in typical daily or yearly trends. The main meteorological parameters in this field are: Temperature, Pressure, Sea level pressure, Precipitation, Perceptible water, Zonal wind, Meridonal wind, Relative Humidity etc.

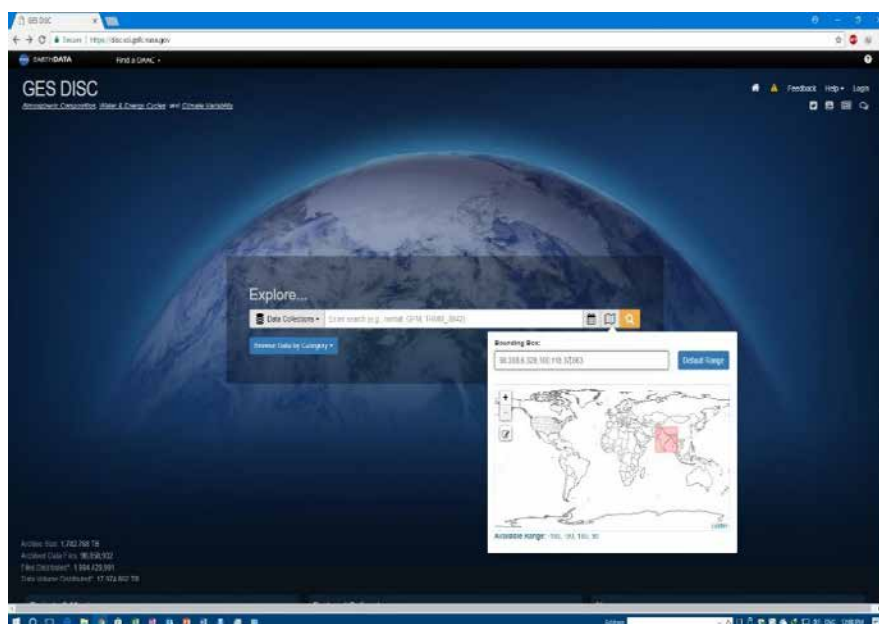
B. Protocol for generating Environmental parameters

GLDAS Noah Land Surface Model was used to download the environmental parameters.

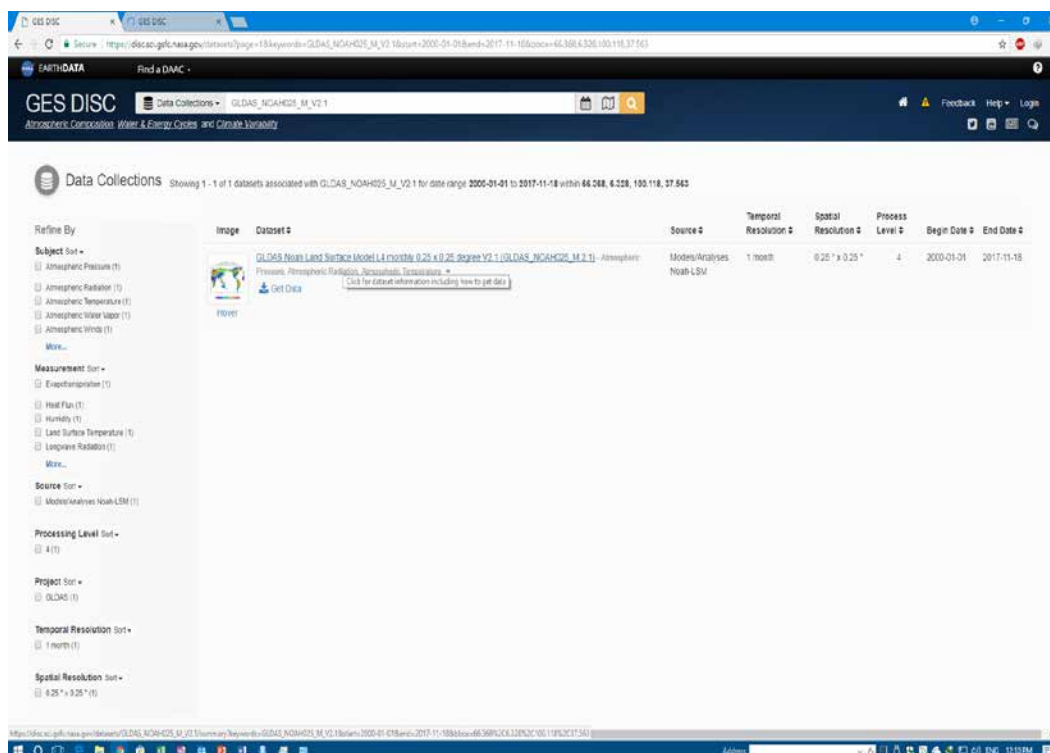
1. Arrange excel sheet (CSV/XLSX) comprising latitude, longitude and address.
2. Data was downloaded from the “GLDAS_NOAH025_M_V2.1” Dataset (<https://disc.sci.gsfc.nasa.gov/>) by setting the start and end dates.



3. The extent of map was set to the boundary of India map by drawing rectangular box.



4. The dataset was downloaded from search results.



Environmental parameters such as Potential evaporation rate (W m^{-2}), Pressure (Pa), Specific humidity (kg/kg), Total precipitation rate ($\text{kg m}^{-2} \text{s}^{-1}$), Soil moisture (kg m^{-2}), Temperature (K), Wind speed (m/s) from GLDAS Noah Land Surface Model requires using of `ncdf4` and `raster` packages of R software.

The input file containing latitude and longitude were read and corresponding parameter values were extracted from the downloaded datasets.

5. Creation of CSV/XLSX files containing remote environmental values.

[illegible]

V. Generation of maps using R

A. Introduction

A map is a graphic representation or scale model of spatial concepts. It is a means for conveying geographic information. Maps are a universal medium for communication, easily understood and appreciated by most people, regardless of language or culture. Incorporated in a map is the understanding that it is a “snapshot” of an idea, a single picture, a selection of concepts from a constantly changing database of geographic information (Merriam 1996).

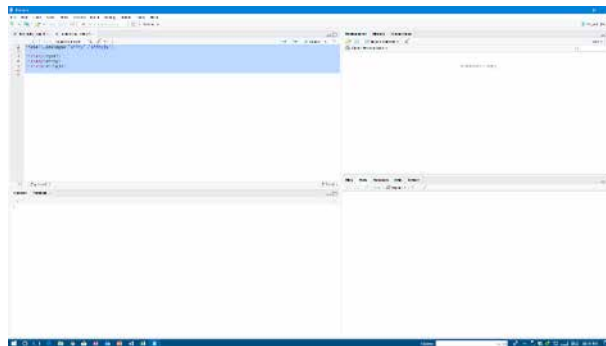
Maps are useful visual tools, from displaying sample sites to performing spatial analyses.

B. Protocol for generating Maps

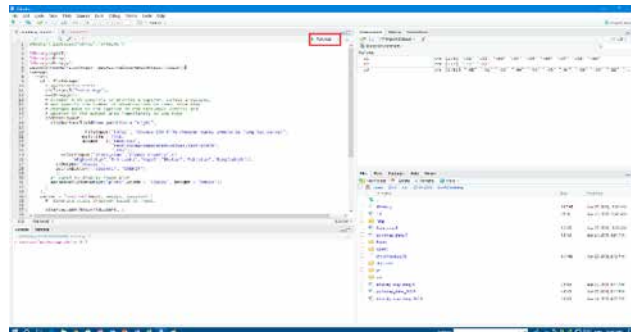
i. Generating maps using User Interface (web page)

i. Point Map using User Interface (web page)

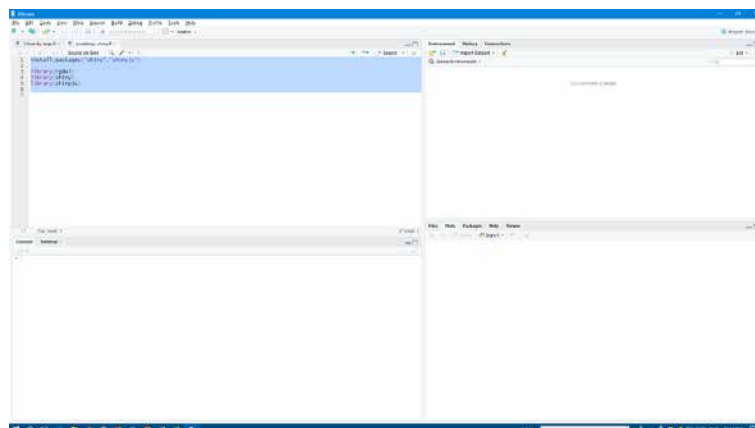
1. Install and load the following libraries from code **Point_map.R**



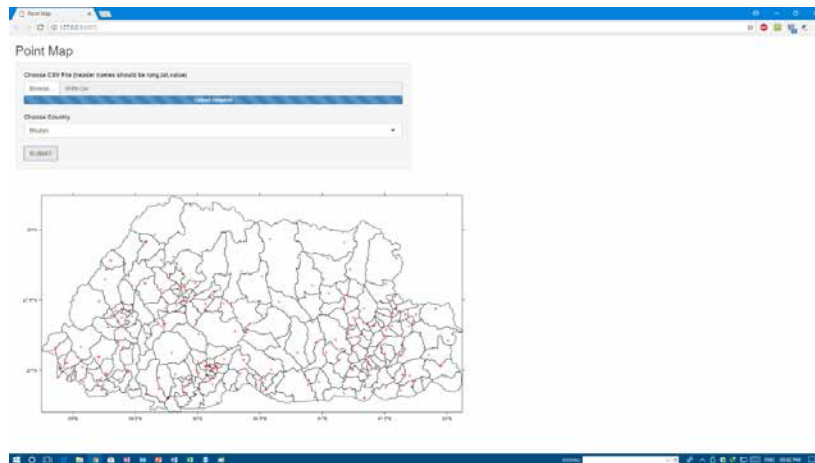
2. Run the shiny app (Annexure A) using Run App button as highlighted below



After running the app, it opens a web page in a browser

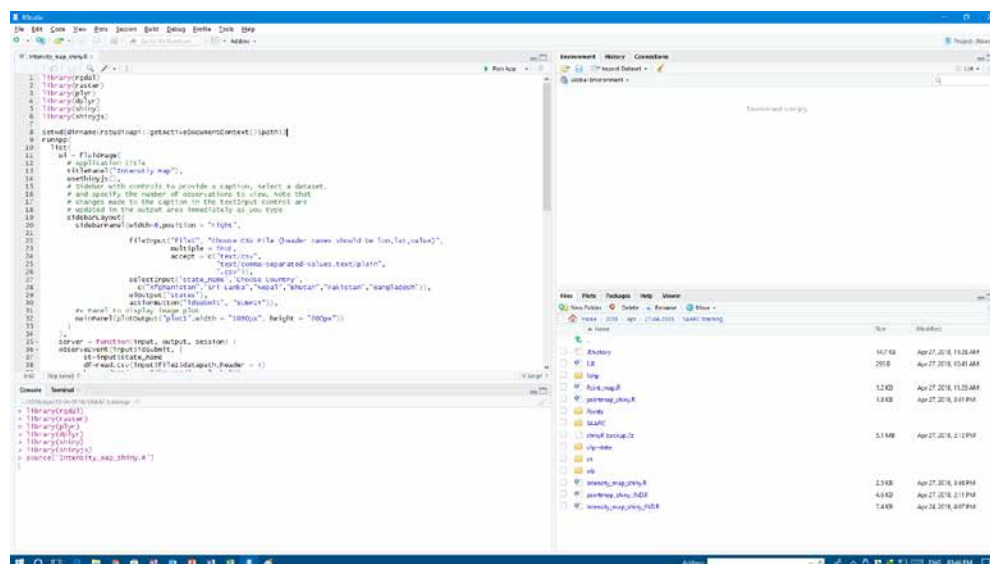


3. Choose the file containing geocoordinates of the specific country and select that country from dropdown and click submit.

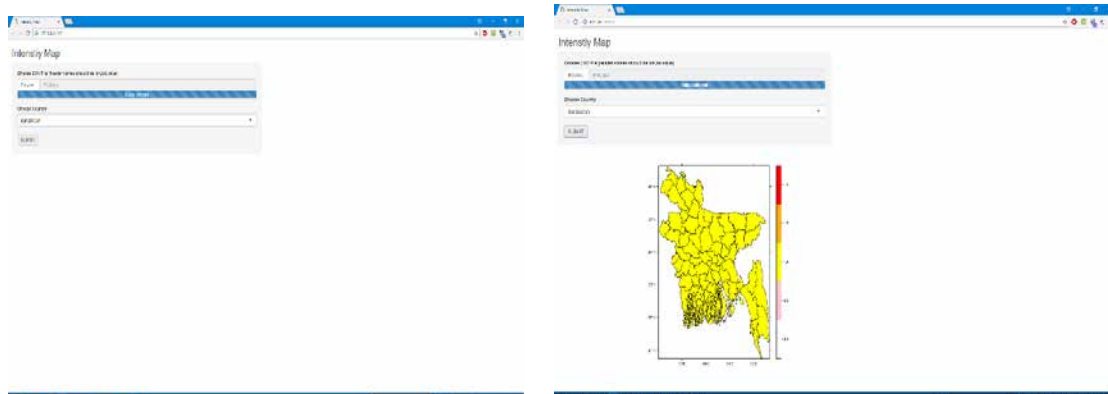


ii. Intensity map using User Interface (R shiny package)

1. Run the Intensity map code(Intensity_map_shiny.R)



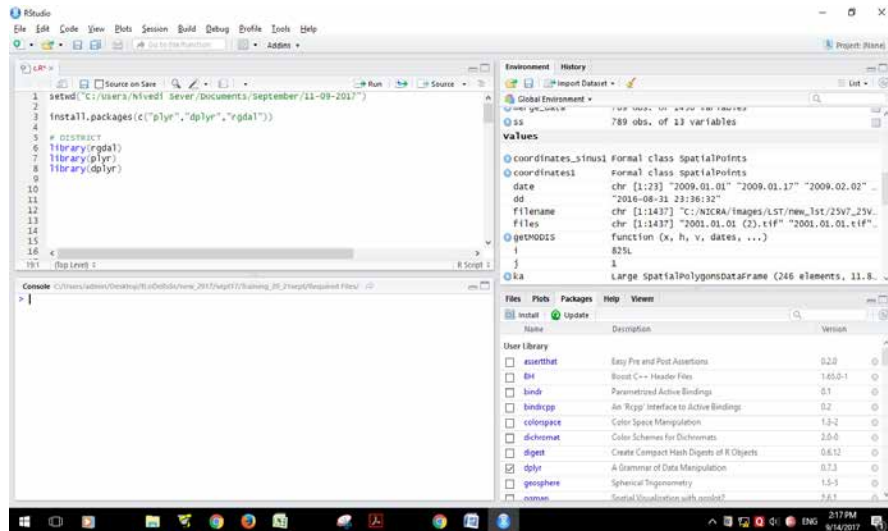
2. Choose the file containing geocoordinates of the specific country and select that country from dropdown and click submit.



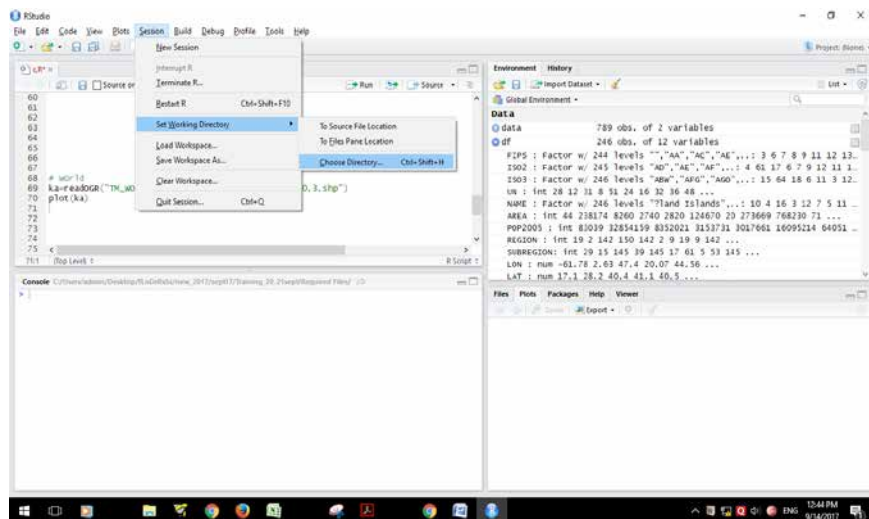
ii. Generating Maps when input is place name - World Intensity map

(Refer Annexure B for the code)

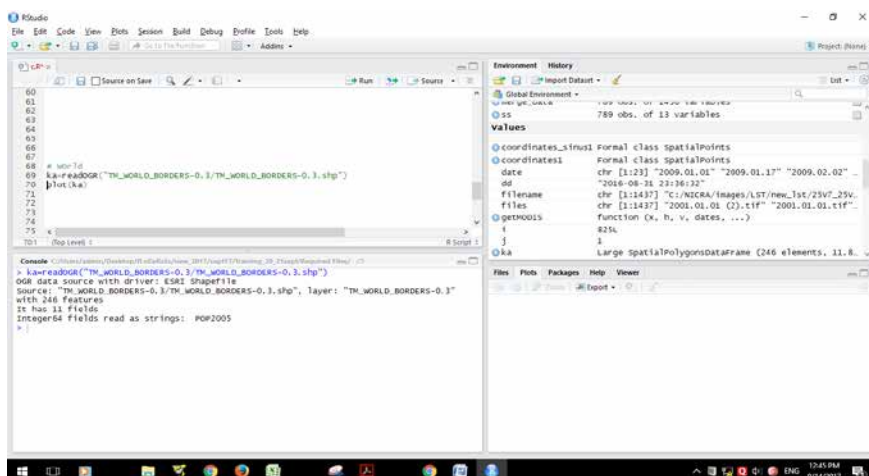
1. Install and load packages rgdal, plyr and dplyr



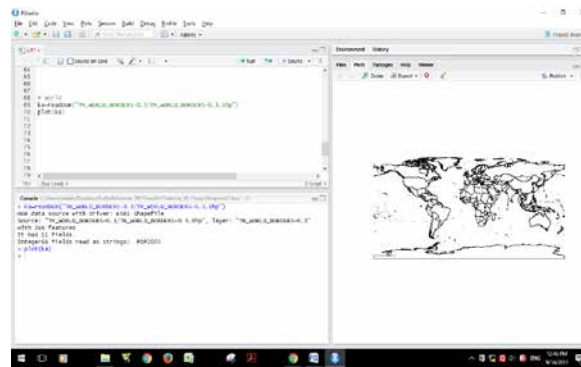
2. Set directory



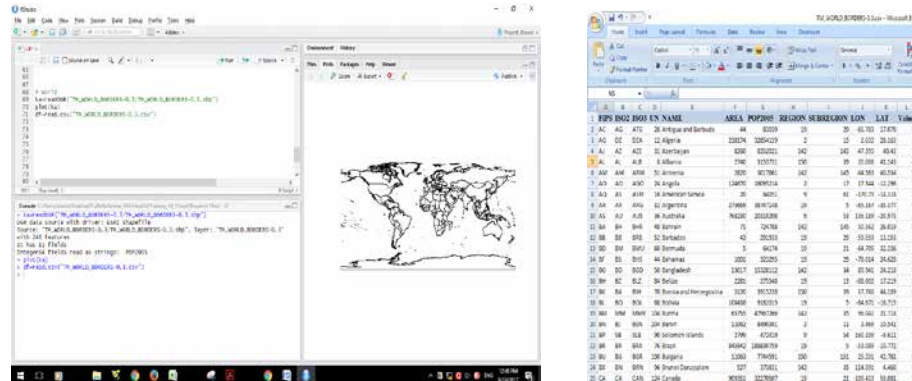
3. Read the world map shape file



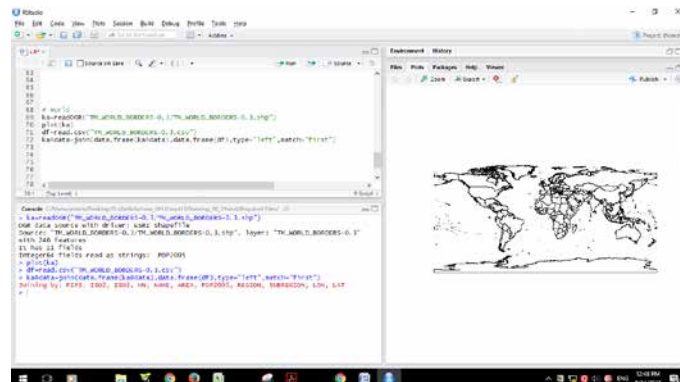
4. Plot the outline of the world map



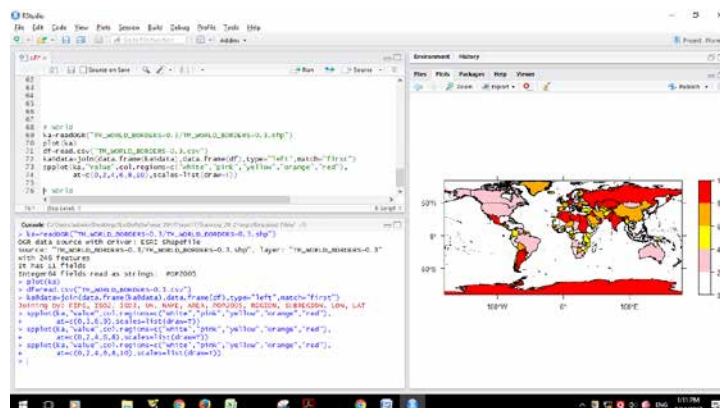
Read the csv file containing data according to shapefile.



5. Joining the shape file with input data

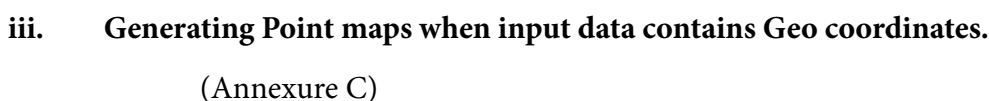
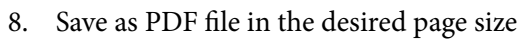


6. Plotting the map with colouring the output data

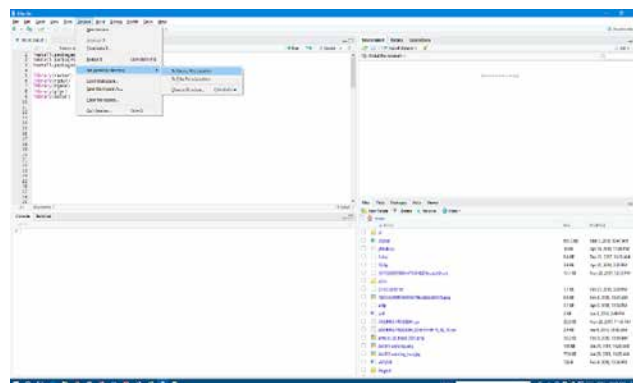


col.regions will represent the colours for the range of values, scale indicates lat long of the world map, at divides

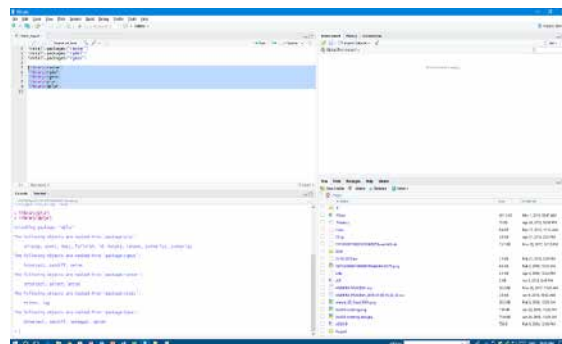
7. Exporting the map



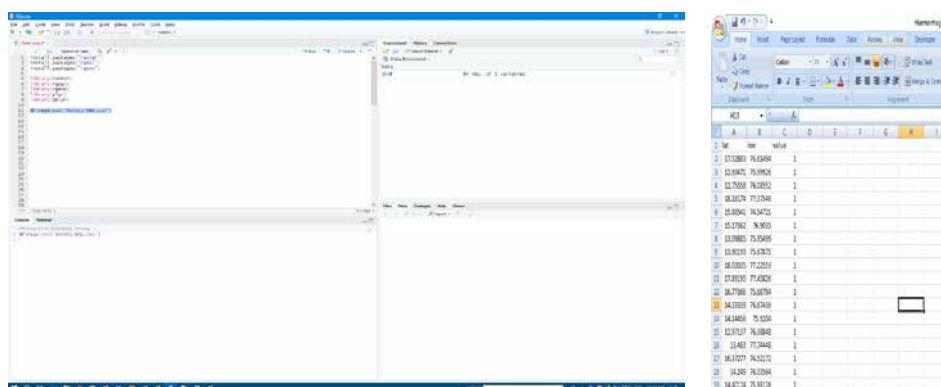
1. Set working directory to the location where source code is stored or Choose directory manually-go to session in the menu bar and click on set working directory and select choose directory.



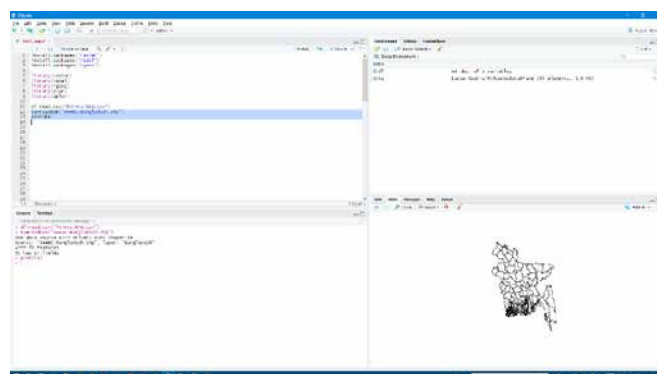
2. Install required packages raster,rgdal and rgeos and Load the libraries using library() function



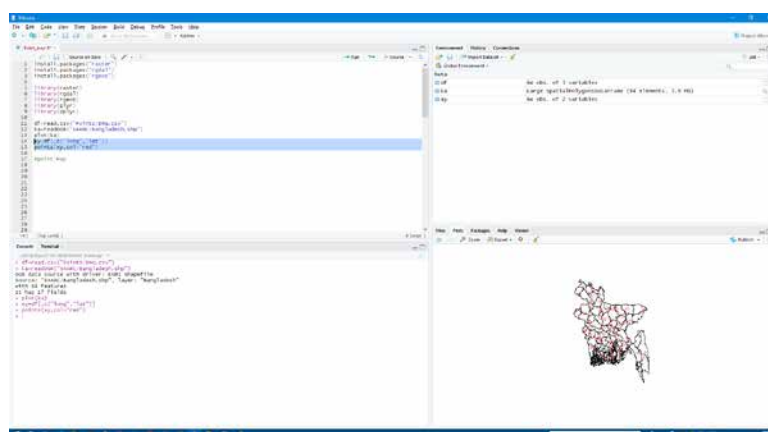
3. Read the csv file containing geocoordinates (Columns should be long, lat, value)



4. Read the district level shapefile of the country. Plot the shapefile and preview

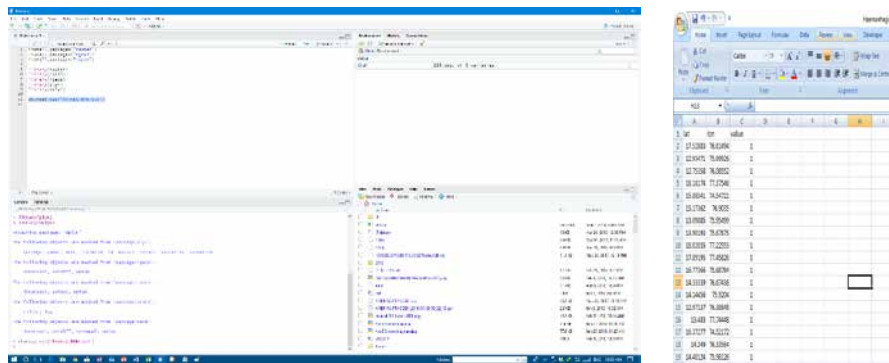


5. Extract geocoordinates and store in xy variable. Plot points on the map.

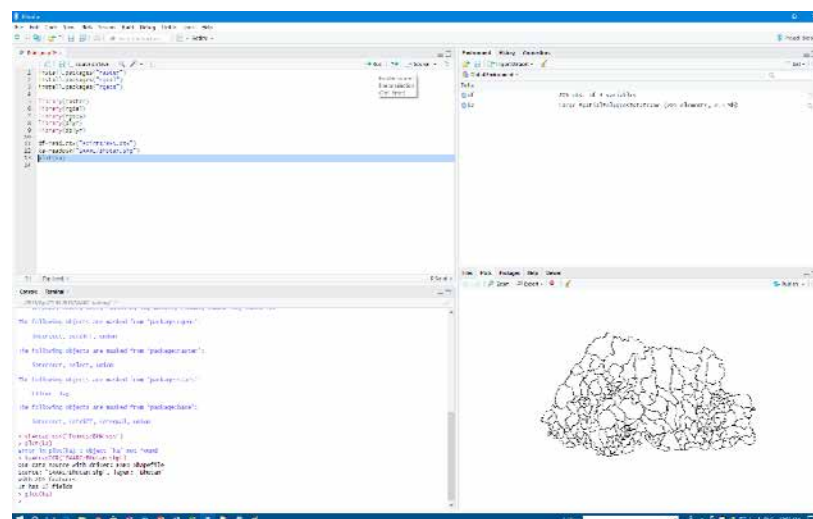


iv. Intensity map when input data contains Geo coordinates.

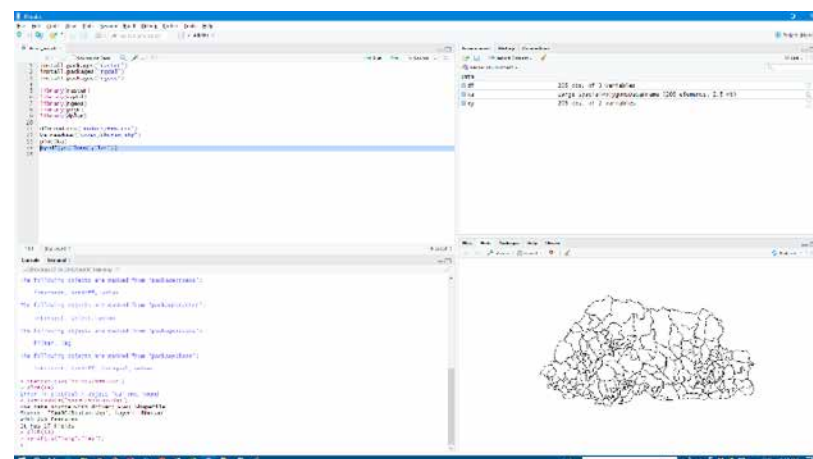
1. Set working directory to the location where source code is stored or Choose directory manually-go to session in the menu bar and click on set working directory and select choose directory.
2. Install required packages raster,rgdal and rgeos and Load the libraries using library() function
3. Read the Bhutan geocoordinates file (Columns should be long, lat, value)



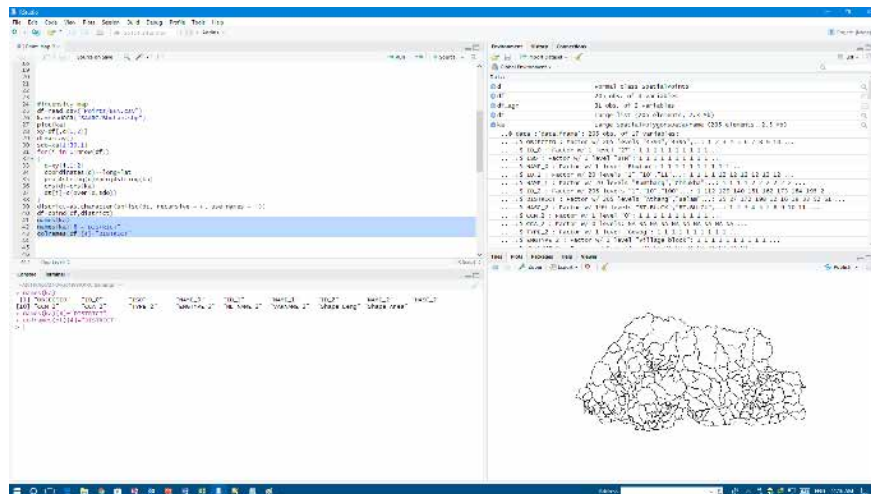
4. Plot the map of read shapefile



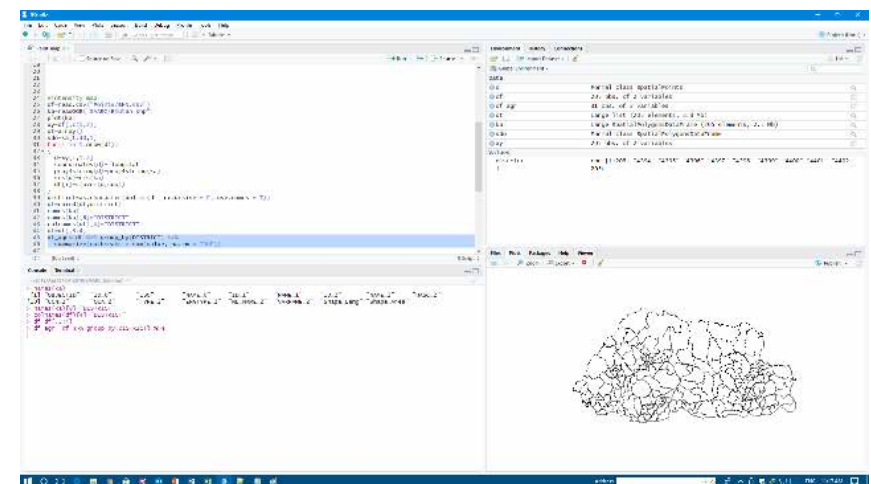
5. Extract geocoordinates and store in xy variable. Plot points on the map.



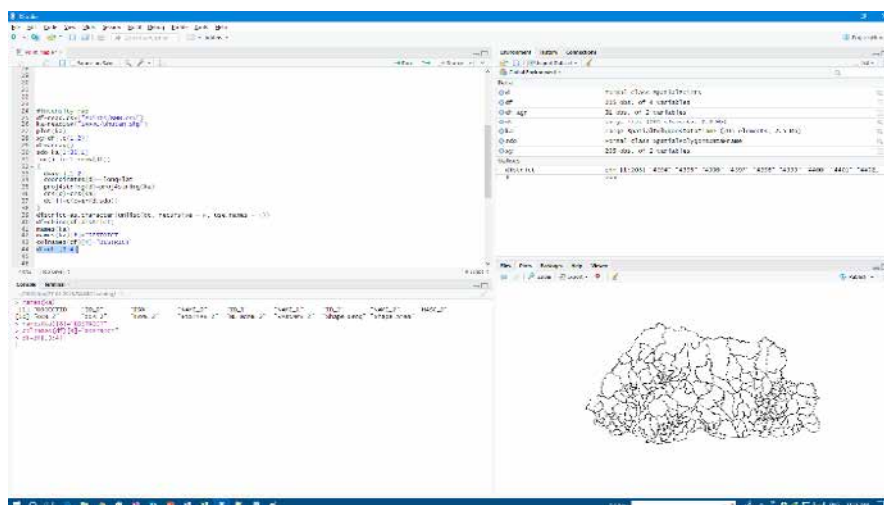
- Change the column name (**NAME_2**) containing district names in the shapefile to **DISTRICT** and binded **district** column to **DISTRICT** to join the shapefile and geocoordinates data frame (**df**) for plotting.



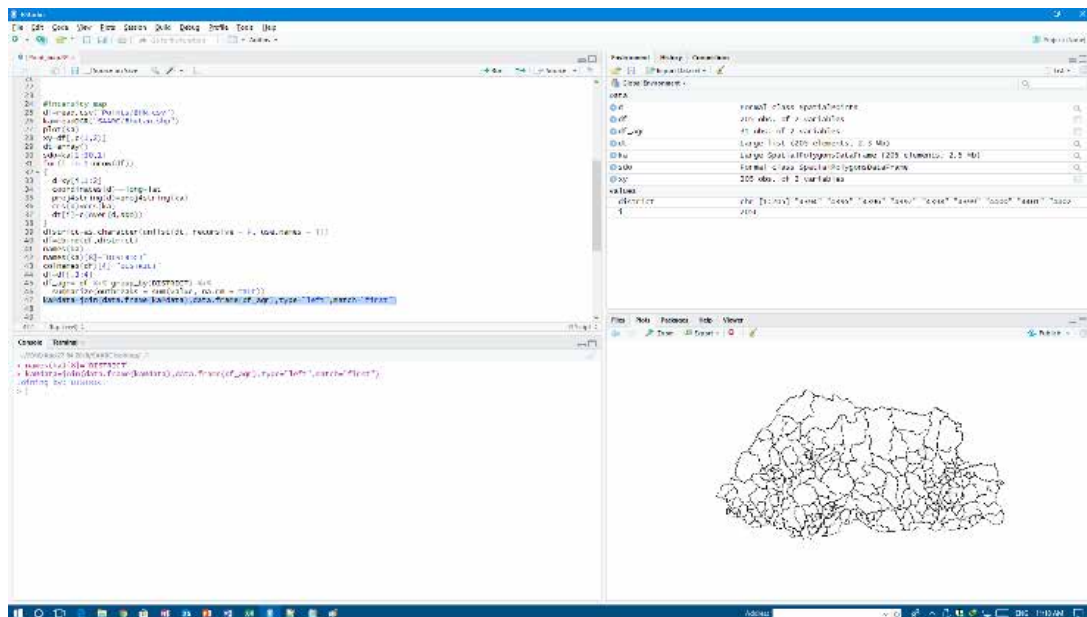
- Aggregate the **value** column using sum function for districts.



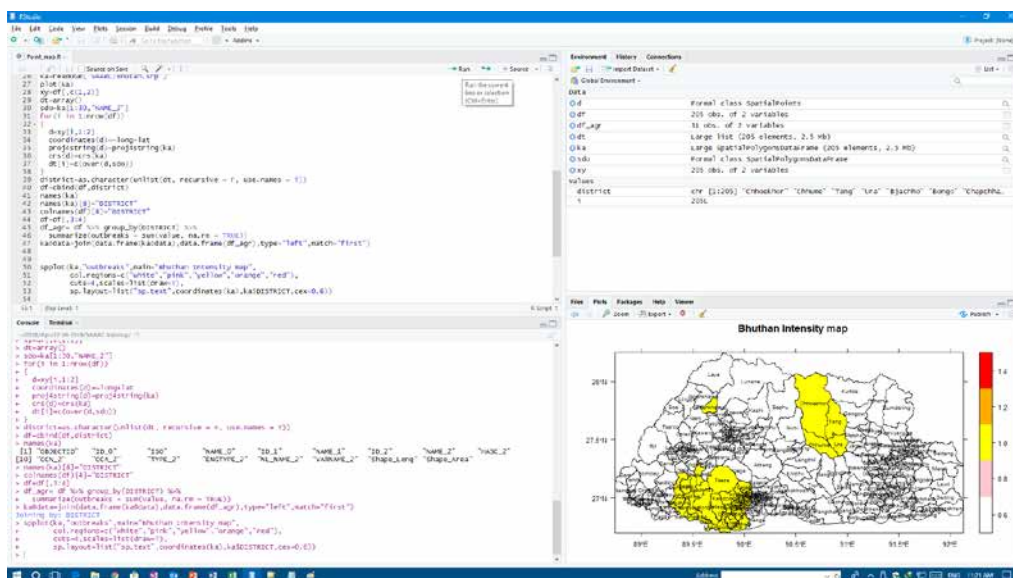
- Keep the district and value column in the data frame(**df**)



13. Join the shapefile data and aggregated data frame(**df_agr**) by using **DISTRICT** column.



14. Plot the map using the `spplot`.



VI. Annexure

A. R programming code required for generating map using User interface

i. Point map

```
install.packages("rgdal","shiny","shinyjs")
library(rgdal)
library(shiny)
library(shinyjs)
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
runApp(
  list(
    ui = fluidPage(
      # Application title
      titlePanel("Point Map"),
      useShinyjs(),
      # Sidebar with controls to provide a caption, select a dataset,
      # and specify the number of observations to view. Note that
      # changes made to the caption in the textInput control are
      # updated in the output area immediately as you type
      sidebarLayout(
        sidebarPanel(width=6,position = "right",
          fileInput("file1", "Choose CSV File (header names should be long,lat,value)",
            multiple = TRUE,
            accept = c("text/csv", "text/comma-separated-values,text/plain", ".csv")),
          selectInput('state_name','Choose Country',
            c("India","Afghanistan","Sri Lanka","Nepal","Bhutan","Pakistan","Bangladesh")),
          uiOutput('states'),
          actionButton("idSubmit", "SUBMIT")
        ),
        #' Panel to display image plot
        mainPanel(plotOutput("plot1",width = "1080px", height = "600px"))
      )
    ),
    server = function(input, output, session) {
      #' Generate state dropdown based on input.
      observeEvent(input$idSubmit, {
        df=read.csv(input$file1$datapath,header = T)
        st=input$state_name
        ka=readOGR(paste0("SAARC/",st,".shp"))
        xy=df[,c("long","lat")]
        coordinates(xy)=~long+lat
```

```

output$plot1 <- renderPlot({
  spplot(ka,"NAME_2",col.regions="white",scales=list(draw=T),colorkey=F,sp.layout = list("sp.
  points",xy,col="red"))
})
})
}),launch.browser = T)

```

ii. Intensity Map

```

library(rgdal)
library(raster)
library(plyr)
library(dplyr)
library(shiny)
library(shinyjs)
setwd(dirname(rstudioapi::getActiveDocumentContext())$path))
runApp(
  list(
    ui = fluidPage(
      # Application title
      titlePanel("Intenstiy Map"),
      useShinyjs(),
      # Sidebar with controls to provide a caption, select a dataset,
      # and specify the number of observations to view. Note that
      # changes made to the caption in the textInput control are
      # updated in the output area immediately as you type
      sidebarLayout(
        sidebarPanel(width=6,position = "right",

          fileInput("file1", "Choose CSV File (header names should be lon,lat,value)",
            multiple = TRUE,
            accept = c("text/csv",
              "text/comma-separated-values,text/plain",
              ".csv")),
          selectInput('state_name',"Choose Country",
            c("India","Afghanistan","Sri Lanka","Nepal","Bhutan","Pakistan","Bangladesh")),
          uiOutput('states'),
          actionButton("idSubmit", "SUBMIT")),
        #+ Panel to display image plot
        mainPanel(plotOutput("plot1",width = "1080px", height = "600px"))
      )
    )
  )

```

```

),
server = function(input, output, session) {
observeEvent(input$submit, {
  st=input$state_name
  df=read.csv(input$file1$datapath,header = T)
  ka=readOGR(paste0("SAARC/",st,".shp"))
  xy1=df[,c("long","lat")]
  dt=array()
  sdo=ka[, "NAME_2"]
  i=1
  for(i in 1:nrow(df))
  {
    d=xy1[i,1:2]
    coordinates(d)=~long+lat
    proj4string(d)=proj4string(ka)
    crs(d)=crs(ka)
    dt[i]=c(over(d,sdo))
  }
  district=as.character(unlist(dt, recursive = F, use.names = T))
  df=cbind(df,district)
  if(input$state_name=="India")
  {
    names(ka)[1]="DISTRICT"
  } else names(ka)[8]="DISTRICT"
    colnames(df)[4]="DISTRICT"
  df=df[,3:4]
  df_agr= df %>% group_by(DISTRICT) %>%
    summarize(outbreaks = sum(value, na.rm = TRUE))
  ka@data=join(data.frame(ka@data),data.frame(df_agr),type="left",match="first")
  output$plot1 <- renderPlot({
    spplot(ka,"outbreaks",col.regions=c("white","pink","yellow","orange","red"),
      cuts=4,scales=list(draw=T))
  })
})
}),launch.browser = T)

```

B. Codes for generating Point and Intensity Map

i. Point Map

```
install.packages("raster")
install.packages("rgdal")
install.packages("rgeos")
library(raster)
library(rgdal)
library(rgeos)
library(plyr)
library(dplyr)
# Set working Directory#
df=read.csv("Points/BNG.csv") #Choose the input file containing Geo coordinates
ka=readOGR("SAARC/Bangladesh.shp") # Choose the country shape file
plot(ka)
xy=df[,c("long","lat")]
points(xy,col="red")
```

ii. Intensity map

```
install.packages("raster")
install.packages("rgdal")
install.packages("rgeos")
library(raster)
library(rgdal)
library(rgeos)
library(plyr)
library(dplyr)
# Set working Directory#
df=read.csv("Points/BHN.csv") #Choose the input file containing Geo coordinates
ka=readOGR("SAARC/Bhutan.shp") # Choose the country shape file
plot(ka)
xy=df[,c(1,2)]
dt=array()
sdo=ka[1:30,"NAME_2"]
for(i in 1:nrow(df))
{
  d=xy[i,1:2]
  coordinates(d)=~long+lat
  proj4string(d)=proj4string(ka)
  crs(d)=crs(ka)
```

```

dt[i]=c(over(d,sdo))
}
district=as.character(unlist(dt, recursive = F, use.names = T))
df=cbind(df,district)
names(ka)
names(ka)[8]="DISTRICT" #Specify the column number containing the values of district
colnames(df)[4]="DISTRICT"
df=df[,3:4]
df_agr= df %>% group_by(DISTRICT) %>%
  summarize(outbreaks = sum(value, na.rm = TRUE))
ka@data=join(data.frame(ka@data),data.frame(df_agr),type="left",match="first")
spplot(ka,"outbreaks", main="Bhutan Intensity map",
  col.regions=c("white","pink","yellow","orange","red"),
  cuts=4,scales=list(draw=T),
  sp.layout=list("sp.text",coordinates(ka),ka$DISTRICT,cex=0.6))

```

Description of Parameters used in spplot:

main=Title of the map,
 sp.layout to represent points or names on the map, scales to represent the lat and long axes on the map.
 cut divides the range of values for better classification.
 Col.region parameter helps to choose the colours used in map. The colours should be one number more than the number of cuts.
 colors() function Returns the built-in color names.

A Practical approach to calculate herd and animal level sample size

K.P.Suresh, Rashmi Kurli and Dheeraj R

ICAR-National Institute of Veterinary Epidemiology & Disease Informatics, Bengaluru, India

Introduction: In a large population where animals are separated into herds, disease has a strong tendency to cluster. This is because the disease agent or agents (whether infectious, environmental or genetic) which are generally not evenly distributed throughout the population (Rothman, 1990). With rare diseases, this clustering is usually even more pronounced. As a result, a very low proportion of herds may be affected by a particular disease-but within those affected herds, the prevalence of the disease amongst animals may be quite high. If a survey is designed to detect the presence of disease, it fails to take into account the clustering of disease in the population, the results of the survey are likely to be very unreliable. This is because the probability formulae that the surveys are based on assume every unit in the population has the same probability of being affected. Another problem with large-area surveys is the logistics of sampling. Probability formulae assume simple random sampling. Simple random sampling of individual animals from a national herd requires the creation of a sampling frame which may need to list millions of animals (each uniquely identified). Such a sampling frame is usually impossible to construct.

The solution to both these problems is to use a two-stage sampling strategy in which herds form the first stage, and individual animals within selected herds, the second stage. In this way, the sample sizes at each stage can be adjusted to reflect the different disease prevalences (the proportion of herds affected in the first stage, and the proportion of animals affected in the herd at the second stage). Two-stage sampling also means that the construction of sampling frames is much simpler. At the first stage, only a list of all herds in the population is required, and at the second stage, only animals in each of the selected herds is to be included in the list. However, the use of two-stage sampling presents particular problems for sample-size calculation and analysis.

The use of two-stage sampling has evolved to meet surveillance objectives for two reasons. First, list of frames of animals for randomized sample selection do not typically exist at a regional or national level, but the list frames of herds can be developed and maintained more readily. Secondly, the theory and applications of within-herd sampling with imperfect diagnostic tests is well developed. The within-herd sampling research has guided the approach to sampling to classify the herd's disease or infection status.

The herd-level sensitivity (HSe or HSESNS) and Specificity (HSp or HSEPC) depend on the individual animal test characteristics, sample size, within herd prevalence etc., HSe and HSp are test characteristics which can be applied at the herd level in a manner equivalent to animal-level Sensitivity (Se) and Specificity(Sp) at the within -herd level. HSe and HSp usually are based on detecting infection if it is present above a fixed level, that is the level is determined according to the epidemiology of the disease or specific national or international rules.

Methodology

The sample size calculated this way takes into account the sensitivity (Se) of the diagnostic method (the lower the Se the larger the sample size). Test specificity is not considered in this calculation. The lack of Sp of a diagnostic test produces false positive results and increases the probability of a Type II error (that is, considering a population as affected by an event when it is actually free of it).

1. First stage Sample size Determination: Herd size

The sample size required for detection of an event if it is present in a population from two-stages sampling is determined by calculating independently the number of herds from which the individuals will be sampled and the number of individuals per herd to include in the sample.

Method 1: Based on the assumption of perfect classifications of each herd tested and of each animal, as either positive or negative. First stage sample size might be estimated using the simple formula based on the normal approximation to the binomial distribution (Snedecor and Cochran, 1989)

$HN = \left(\frac{Z^2}{L}\right) HTP (1 - HTP)$, where HN is the sample size (number of herds tested), HTP is the estimated Herd prevalence and L (usually 5 to 10%) is Tolerance around prevalence for the varying level of confidence Z (90%CI, 95%CI or 99%CI). This is based on assumption of approximately infinite population and perfect test.

Method 2: Survey design allowing for imperfections in the test.

This formula is based on Herd level sensitivity and Herd level specificity and also the assumptions of normal approximation to binomial distribution.

$$HN = \left(\frac{Z}{L}\right)^2 \left[\frac{(HSENS(HTP) + (1 - HSPEC)(1 - HTP)) \times (1 - HSENS(HTP) - (1 - HSPEC)(1 - HTP))}{(HSENS + HSPEC - 1)^2} \right]$$

where HSENS and HSEPC are the herd level sensitivity and specificity of test, HTP is herd level prevalence. L (usually 5 to 10%) is Tolerance around prevalence for the varying level of confidence Z (90%CI, 95%CI or 99%CI).

2. Second-stage sample size determination: Number of animals per herd

$$n_i = \left[1 - (1 - CL)^{\frac{1}{e}} \right] \times \left(N_i - \frac{e - 1}{2} \right)$$

CL : The level of confidence, The confidence that the user wants to have in the results.
Acceptable values: 90%, 95% or 99%.

e : The number of detectable individuals with the event in the population. This value is the product of population size (N) by detectable prevalence. Detectable prevalence is the result of the product of expected prevalence (p) by Sensitivity (Se) of the diagnostic device or method

$$e = N \times p \times Se$$

N_i : Number of animals per herd

Valid estimates of herd-level prevalence can be obtained from population surveys using cluster sampling. Herds are selected at random and a diagnostic test is applied to randomly selected animals from these selected herds. Based on the results of individual-animal tests, each herd is assessed as either positive or not positive (thus providing a herd-level test so that the herd-level prevalence can be estimated). The difficulty with this approach lies in that most tests have imperfect animal-level sensitivity (HSENS) and specificity (HSPEC), which means that the categorisation of the herd as either positive or negative (i.e. herd tests) is also imperfect.

Results

Sample size for 2-stage survey with fixed herd/flock sensitivity can be calculated using the web based calculator "http://epitools.ausvet.com.au/content.php?page=2StageFreedomSS_2"

Calculate sample sizes for 2-stage surveys for demonstrating disease freedom, for specified target herd/flock sensitivity and system sensitivity. This analysis calculates the number of herds and the number of animals within each herd to be tested to provide specified herd and system sensitivities (probability of detecting disease) for the given animal and herd-level design prevalence and test sensitivity. Test specificity is assumed to be 100% (or follow-up testing of any positive will be undertaken to confirm or exclude disease).

Numbers of herds to test are calculated using the hyper geometric approximation if the number of herds in the population is specified as well as using the binomial formula assuming unknown (large) number of herds in the population. If the population size is not specified only the binomial results are presented.

Numbers of animals to test in each herd are calculated for a range of herd sizes using the hyper geometric approximation and for unknown (large) herd sizes using the binomial calculation.

Design prevalence (specified level of disease to be detected) must be specified at both animal and herd levels. Design prevalence can be specified as either:

- a proportion of the population infected; or
- a specific (integer) number of herds infected (for herd-prevalence only and only if the number of herds in the population is specified).

Inputs required include:

- animal-level design prevalence (as a proportion only);
- herd-level design prevalence and whether this is specified as a proportion or number of herds;
- the estimated test sensitivity;
- the target herd or flock sensitivity (SeH or HSENS) which is the probability of detecting disease if it is present in a herd at the specified animal-level design prevalence;
- the target system sensitivity (SSe) which is the probability of detecting disease if it is present in the population at the specified animal and herd level design prevalences;
- The number of herds in the population (optional).

Outputs from the analysis include:

- The total numbers of herds to be sampled;
- The maximum total sample size (if animal-level prevalence is specified as a proportion);
- The numbers of animals to test in herd, for a range of herd sizes to achieve the specified value for SeH; and
- The numbers of animals to test in herd and the corresponding numbers of herds to test, for a range of herd sizes and SeH values.

Hypothetical Example: For Animal level design prevalence of 10% , herd level design prevalence of 20%, Test sensitivity of 90%, Target herd sensitivity of 50% and target system sensitivity of 95% confidence interval, the results are presented as follows

Table 1 : Inputs

Animal-level design prevalence	10%
Herd-level design prevalence	20%
Test sensitivity	0.9
Test specificity	1
Target herd/flock sensitivity (SeH)	0.5
Target system sensitivity (SSe)	0.95
No. herds in population	Unknown

Table 2 : Number of herds to be sampled and number of herds or villages to be sampled

	Number of herds to sample	Maximum number of samples
No. herds in population unknown	29	232

Table 3 : Numbers of animals to be sampled for different herd sizes for SeH = 0.5

	Number of animals to sample
Herd size = 10	6
Herd size = 20	7
Herd size = 30	7
Herd size = 40	8
Herd size = 50	8
Herd size = 100	8
Herd size = 200	8
Herd size = 500	8
Herd size = 1000	8
Herd size = 5000	8
Herd size = 10000	8
Herd size = unknown	8

Table 4 : Sample Size requirement at different prevalence rates at fixed sensitivity and specificity of tests

Population Prevalence (%)	Maximum Number of Samples	Number of villages to samples	Sampling at Herd Size											
			10	20	30	40	50	100	200	500	1000	5000	10000	NA
1	670176	3744	9	18	27	36	45	89	123	153	166	176	178	179
2	166519	1871	9	18	27	36	45	62	74	83	86	89	90	89
5	26180	748	9	18	19	25	24	31	34	35	36	36	36	35
10	6714	373	9	13	14	15	16	17	18	18	18	18	18	18
15	2988	249	7	10	10	11	11	12	12	12	12	12	12	12
20	1674	186	7	8	8	9	9	9	9	9	9	9	9	9
25	1043	149	5	7	7	7	7	7	8	8	8	8	8	7
30	744	124	5	6	6	6	6	6	6	6	6	6	6	6
35	530	106	4	5	5	5	5	5	6	6	6	6	6	5
40	372	93	4	5	5	5	5	5	5	5	5	5	5	4
45	328	82	4	4	4	4	4	4	4	4	4	4	4	4
50	222	74	4	4	4	4	4	4	4	4	4	4	4	3

Conclusion:

The herd-level sensitivity and specificity needs to be decided by the investigator. For a given screening test, the herd-level sensitivity and specificity can be set to a range of values by varying the sample size. While sensitivity plays a role in the sample size, herd-level specificity is more important, due to the much greater proportion of false positive results than false negative when prevalence is very low. If specificity is increased, the number of herds that needs to be sampled decreases, but the number of animals that must be sampled from each herd to achieve this level of herd test specificity increases. The choices of herd-level sensitivity and (especially) specificity therefore determine the balance between the required number of herds and the number of animals per herd.

References:

1. Cannon, R.M., 2001. Sense and sensitivity – designing surveys based on an imperfect test. *Prev. Vet. Med.* 49, 141–163.
2. Cannon, R.M., Roe, R.T., 1982. *Livestock Disease Surveys – A Field Manual for Veterinarians*. Canberra.
3. Collins, M.T., Sockett, D.C., Goodger, W.J., Conrad, T.A., Thomas, C.B., Carr, D.J., 1994. Herd prevalence and geographic-distribution of and risk- factors for, bovine paratuberculosis in Wisconsin. *J. Am. Vet. Med. Assoc.* 204, 636–641.
4. Snedecor, G.W., Cochran, W.G., 1989. *Statistical Methods*, 6th ed. Iowa State University Press, Ames, Iowa, p.121.

BIOINFORMATICS APPLICATIONS IN ANIMAL DISEASE PATHOGENS

S. Shivakumara Swamy

Institute of Bioinformatics and Applied Biotechnology, Bengaluru, India

Bacterial, viral, and parasitic diseases pose a major threat to animal welfare and production worldwide. Due to their zoonotic potential, some of these diseases are also of public health concern. The terms “genomics” and “genomic methods” describe “the molecular and bioinformatics techniques that employ all or part of the genome to answer a question about an organism or a group of organisms” (Buckley, 2004). Bioinformatics is indispensable in genomic (*i.e.*, nucleic acid and protein sequences) research; the acquisition of data, creation of databases, and analysis of data require computers and tools/software. Genomics has immense applications in the quest to understand nature, and comparative genomics is an indispensable tool for pathogen characterization. The Centers for Disease Control and Prevention (CDC) opined that “advances in science and technology aimed at identifying the complete genetic makeup of microorganisms are ushering in a new era for controlling infectious threats. By using genetic sequencing to examine infectious pathogens, these technologies are on the verge of revolutionizing our ability to diagnose infectious diseases, investigate and control outbreaks, understand transmission patterns, develop and target vaccines, and determine antimicrobial resistance—all with increased timeliness and accuracy and decreased costs” (CDC, 2013).

The first complete genome sequence from a free-living organism was that of *Haemophilus influenzae* strain Rd KW20 (Fleischmann *et al.*, 1995). This pioneering work at The Institute for Genomic Research (TIGR) introduced and popularized the concept of whole genome random sequencing by the ‘shotgun’ approach. Since the completion of the first bacterial genome sequence, thousands of bacterial and archaeal genomes have been sequenced and annotated using novel tools and techniques. The genomes of several bacterial pathogens of veterinary importance have also been sequenced, and whole genome comparisons have provided new insights into the physiology and evolution of these organisms (Siddaramappa, 2016). Although these developments have revolutionized veterinary microbiology, there are many challenges/questions that need to be addressed. My lecture will primarily focus on two questions:

- (i) are we at a stage where comparison and/or normalization of genomic data is possible on a global scale?
- (ii) what needs to be done in the area of disease diagnosis to realize the potential of genomic data?

References :

1. Buckley, 2004: The genomics of disease-causing organisms: mapping a strategy for discovery and defense, American Society For Microbiology. (available at <http://academy.asm.org/index.php/browse-all-reports/515-the-genomics-of-disease-causing-organism-mapping-a-strategy-for-discovery-and-defense>)
2. CDC, 2013: A new landscape for combating infectious disease, Centers for Disease Control and Prevention. (available at <http://stacks.cdc.gov/view/cdc/13524>)
3. Fleischmann *et al.*, 1995: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. (PMID: 7542800)
4. Siddaramappa, 2016: *Histophilus somni* genomics and genetics. (PMID: 26728065)

Risk Factor identification for incidence of Brucellosis and its informatics

Rajeswari Shome

ICAR-National Institute of Veterinary Epidemiology & Disease Informatics, Bengaluru, India

Brucellosis is one of the major bacterial zoonoses, affecting domestic animals and humans in many developing countries (McDermott et al., 2013) and recognized as one of the seven neglected, under-detected and under-reported zoonosis (World Health Organization, 2011). Brucellosis has been classified as under multiple species diseases, infections and infestations (OIE, 2016). The economic impact of brucellosis is due to productive and reproductive failures in livestock, loss of man days and morbidity in humans and international obstacle to trade and export of animals and their products (Mcdermott and Arimi, 2002). Brucellosis has been eradicated from several countries in Northern and Central Europe, Canada, Japan, Australia and New Zealand (OIE, 2016). Although continuous progress has been made in brucellosis control in many parts of the world still it remains a major public health hazard of great economic importance causing an ever-increasing concern in many countries including India.

The presence of brucellosis in India was first established early in the previous century and since then it has been reported from almost all states (Renukaradhya et al., 2002). Increasing brucellosis prevalence in the country is attributed to constraints that forbid culling of positive animals due to religious and economic impact leading to desperate sale of positive animals, infected animal per se act as a source of infection, purchase of unscreened infected animals for replacement or upgradation or using semen from unscreened bulls for artificial insemination are contributin to the spread of the infection (Smits and Kadri, 2005). Hence, knowledge of seroprevalenec and spatio-temporal distribution of the disease and risk associated with disease is of paramount importance to strengthen the disease control program.

Principles of Risk Analysis

Risk analysis comprises three components: risk assessment, risk management and risk communication.

Risk assessment

In this component, the risks of an event occurring or of taking a particular course of action are first identified and described. The likelihood of these risks occurring is then estimated, their potential consequences evaluated and the assessment of the risk modified accordingly. For example, an exotic disease with a high risk of entry to a country but only a low risk of establishment or minimal potential socio-economic consequences would only obtain a low overall score on a risk assessment.

Risks can be assessed in a quantified, semi-quantified or qualitative way. It is inherently extremely difficult to quantify or actually put probability numbers to risks in many biological systems because of the lack of historical precedents and serious gaps in available biological data.

The risks can be described as “extreme”, “high”, “medium” or “low”, or by a simple scoring system, for example, 1-5 for the level of risk and 1-5 for the level of potential consequences.

Risk management

This is the process of identifying, documenting and implementing measures to reduce risks and their consequences. Risks can never be completely eliminated. The aim is to adopt procedures to reduce the level of risk to an acceptable level.

Risk communication

This is the process of exchange of information and opinions on risk between risk analysts and stakeholders. Stakeholders in this context include all those who could be affected by the consequences of risks, that is, everyone from farmers to politicians. It is important that risk assessment and risk management strategies be fully discussed with stakeholders, so that they feel comfortable that no unnecessary risks are being taken and that risk management costs are a worthwhile insurance.

To ensure ownership of decisions, risk analysts and decision-makers should consult with stakeholders throughout the whole process of risk analysis so that the risk management strategies address their concerns, and decisions are well understood and broadly supported.

In epidemiology, surveys are carried out both for estimating descriptive measures of disease frequency, such as incidence and prevalence, and for estimating the effects of risk factors on diseases, by relative risk, odds ratio, and other measures of association

Measures of risk

Risk/ Prevalence (P): The prevalence (or risk) of an diseases/event occurring is calculated as

$$P_{\text{event}} = \frac{\text{Number of events occurring}}{\text{Total number in group}}$$

Where the event could be a disease, condition, risk factor, etc. Risks are not always negative e.g.

We are often interested in comparing risks of an event occurring, e.g. death, between different groups and there are several ways of doing this. The following table will be used to demonstrate the formulae involved.

Category	Event occurs	Event does not occur occurs	Total	Risk of event by group
Exposed or treated	a	b	a+b	$P_{\text{exp}} = \frac{a}{a+b}$
Not Exposed or not treated	c	d	c+ d	$P_{\text{Unexp}} = \frac{c}{c+ d}$

Relative risk:

This measures how much more likely the event is to occur in one group compared to another.

The risk of the event occurring for the exposed population is $\frac{a}{P_{\text{exp}}} a + b$

The risk of the event occurring for the unexposed population is $\frac{c}{P_{\text{Unexp}}} c + b$

Relative risk = This is sometimes called a Risk Ratio. If **RR>1** then the risk of disease for the exposed group is larger than the risk of disease for the unexposed group.

$$RR = \frac{\text{risk of event in exposed group}}{\text{risk of event in un exposed group}} = \frac{P_{\text{exp}}}{P_{\text{unexp}}} = \frac{a (c + d)}{c (a + b)}$$

Odds Ratio (OR).

Another common measure used in medical statistics is the **Odds Ratio (OR)**. First, odds are calculated using odd ratio = $\frac{p}{1-p}$

If the **Odds Ratio** >1 then the odds of disease occurring in the exposed group are larger than the odds of disease in the unexposed group, so exposure to the factor has increased the risk of contracting the disease.

Risk factors associated with brucellosis in animals

The risk factors can be categorized into those associated with characteristics of animal populations, management and brucella structure and biology.

A) Risk factors associated with *Brucella* spp

- ❖ *B. abortus* is the aetiological agent of bovine brucellosis in cattle although also infects other species and important risk to the maintenance of the agent in the animal population with special importance in areas where wildlife and cattle commingled.
- ❖ *B. melitensis* is the main etiological agent of brucellosis in small ruminants, although sheep can be also infected by *B. ovis*. Sporadic cases of brucellosis have been described in sheep and goats as *B. abortus* and *B. suis*.
- ❖ Porcine brucellosis is caused by *Brucella suis* biovars 1, 2 or 3. The disease caused by biovars 1 and 3 is similar, while that caused by biovar 2 differs from 1 and 3 in its host range.

B) Risk factors associated with the host

B1) Age

- Age has been referred to as one of the intrinsic factors associated with brucellosis. Brucellosis has traditionally been considered a disease of adult animals since susceptibility increases after sexual maturity and pregnancy.
- Higher seroprevalence of brucellosis has been observed in older animals, both in cattle over 5 years old and about 2.0 years and above in small ruminants. Similar results have been observed in wild boars and camels.
- *Brucella* spp. presented a tropism to the reproductive tract due to the production of erythritol, a 4-carbon sugar produced in the foetal tissues of ruminants that stimulates the growth of *Brucella* [Petersen et al 2013]. Thus, it may also explain the higher prevalence in adult animals than in young.
- On the other hand, a higher prevalence of brucellosis in adults has also been associated with longer contact with infected animals or with the environment.

B2) Sex

- ❖ Female ruminants presented higher odds of brucellosis infection, this is difficult to explain, it could be associated with the intrinsic biology of the microorganisms and its tropism to the foetal tissues as previously described.
- ❖ Since brucellosis infection in males presented clinical signs such as epididymitis and orchitis, the prevalence in males could be lower than females because they may be culled faster [Coelho et al., 2013].

B3) Species

- ❖ Transmission of brucellosis occurs in ruminants through the excretion of contaminated materials from the female genital tract, which constitutes the main form of transmission to other animals and humans.
- ❖ The prevalence of brucellosis is variable among species and prevalence in farm animals seems to be lower in small ruminants than large ruminants
- ❖ The vaginal excretion of *Brucella* spp. in goats is greater and more prolonged than sheep, lasting for 2-3 months. In sheep, it is generally lower and normally ceases within 3 weeks after birth or abortion.
- ❖ It is also common that excretion occurs through milk or semen [Blasco 2001]. The excretion of *Brucella* in milk is generally intermittent and usually only appears 6 to 12 days after the abortion.

B4) Breed

- ❖ Higher prevalence of brucellosis has been reported in cross-breed cattle than local breeds, although other reports indicated no statistical differences among cattle breeds and small ruminants .
- ❖ In swine, some breeds such as Duroc and Jersey Red crosses may be less susceptible to experimental challenge with *B. suis*, suggesting some genetic resistance [Cameron et al 1942].
- ❖ Previous studies showed that stray dogs demonstrated a greater than three-fold rate of infection versus non-stray dogs [Lovejoy 1976].

C) Risk factors associated with herds

C1) Herd/flock size

- An important risk factor for brucellosis seropositivity is herd size, being higher in large herds and/or flocks. The control of reproductive management is difficult in large flocks, where parturitions on grazing areas are frequent and abortions are a source of pasture contamination.
- In addition, animal movement in large herds is frequent, both for replacement and/or trade, thereby increasing the risk of infection by brucellosis.

- The higher prevalence of brucellosis observed in large flocks may be also associated with the utilization of communal pasture areas, utilization of common paths and/or roads and due to contact with others flocks.
- Cleaning and disinfection procedures of premises and manure removal in large-sized flocks is more difficult than in medium or small flocks because it requires the availability of mechanical equipment and consequently a higher health status of a flock may influence the predisposition to brucellosis infection.
- In small-sized herds, farmers can easily identify sick animals and veterinary and preventive treatments are usually carried out at low financial cost.

C2) Number of species

- ❖ Farming several species in the same herd has been described as a risk factor for brucellosis where several species intermingle with higher chances of being *Brucella* seropositive because of multiple sources of infection.
- ❖ The presence of swine could be a risk for brucellosis transmission to cattle [Cvetnic et al., 2005] and is a public health concern.
- ❖ The practice of mixing cattle, either through grazing or sharing watering points, is a significant risk factor for brucellosis.
- ❖ The presence of dogs has been described as a risk for brucellosis infection in farm animals. However, the higher risk for cattle on farms which also had sheep or goats suggests that some of the cattle infections may have originated from small ruminants since *B. melitensis* biovar 3 was isolated from cow's milk.

D. Risk factors associated with farm management and environment

- ❖ Dairy animals have a much greater chance of not only contracting brucellosis but also of spreading it faster than beef animals because dairy animals are subjected to more stress conditions on farms, leading to a higher susceptibility to brucellosis infection.
- ❖ Introduction of animals from market fairs also presents a higher risk of infection. The majority of infections or reinfection in disease-free herds starts through buying infected animals of unknown status.
- ❖ Larger herds might be expected to be associated with intensive management practices that are typically more difficult to control and allow for closer contact between animals and their environment, which increases the potential for exposure to infectious excretions.
- ❖ Cleaning and disinfection of farm facilities and proper manure removal have been described as a protective factor against brucellosis infection.
- ❖ Insect rodents on dogs could act as a mechanical vector of brucellosis
- ❖ Also the influence of the agro-ecological zone has been also referred as a brucellosis risk factor. High humidity, low temperatures and absence of direct sun light may favour the survival of *Brucella* for several months in the environment.

E. Other factors associated with brucellosis

- ❖ Farmers' older than 55 years had knowledge about brucellosis was a protective factor for brucellosis prevention.
- ❖ Farmers who had previously experienced brucellosis in their herds had a higher probability of having greater knowledge of bovine brucellosis, which is consistent with having experience with the disease.
- ❖ Farmer's lack of awareness about brucellosis, improper handling of aborted materials and the habit of consuming raw milk, among other factors, might
- ❖ contribute to further spread of brucellosis in their livestock and expose the community to a public health hazard.
- ❖ Producer's associations, education and veterinary support have been recognized as protective factors.

F) Brucellosis in wild and marine animals

- ❖ Wild animals have been referred as reservoir of brucellosis and represent an important risk of infection to farm animals, particularly in extensive breeding systems.
- ❖ Several species such as bison, reindeer, caribou or wild boar have been described as potential source of infection of livestock.
- ❖ *B. abortus* and *B. suis* have been isolated worldwide from a great variety of wildlife species and general risk factors, which can be identified in most of the wildlife diseases are wildlife overabundance, movements of wild and domestic animals and fomites.
- ❖ *Brucella* was detected in free-ranging pinnipeds and cetaceans from America, Europe, Japan, New Zealand, the Solomon Islands and the Antarctic, as well as in captive bottlenose dolphin (*Tursiops truncatus*)

G) Brucellosis and zoonotic risk

- ❖ Brucellosis represents an important threat as a work-acquired infection among dairy farmers, butchers, veterinary practitioners, meat inspectors, slaughterhouse personnel or artificial inseminators who do not take adequate biosafety precautions while performing their jobs
- ❖ In addition, brucellosis vaccines such as Rev-1 and RB51 are live dried living vaccines. Thus, needlestick accidents during their preparation or administration could also be a risk factor for human infection.
- ❖ Foodborne brucellosis is an important biological hazard associated with dairy products. However, the presence of *Brucella* spp. in marine animals indicates that fish-borne brucellosis could be a future hazard to be considered.

Conclusions

Brucella spp. is responsible for a contagious disease that results in reproductive failure and has an important economic impact, not only in animal health but also in public health

because of its zoonotic characteristics. To achieve the control and eradication of brucellosis, the identification of all potential risks is necessary. Given the important role of domestic and wild animals as potential sources of *Brucella* infection, further risk assessment will require more complete and reliable data on the infection prevalence. Several risk factors have been described for brucellosis infection, although the herd or flock size, species and age have been cited as the most important. Brucellosis has traditionally been associated with farm animals, however, risks of brucellosis associated with wildlife and marine mammals could be a new threat and further epidemiological studies are necessary. In addition to animal sanitary measures, complementary measures such as good farm practices, biosecurity, training and education are necessary to control this old disease that is still of concern today.

References :

1. World Health Organization (WHO). The control of neglected zoonotic diseases. In Report of the 3rd WHO Conference on the control of neglected zoonotic diseases: 'Community-based interventions for prevention and control of neglected zoonotic diseases', 23-24 November 2010, Geneva. WHO, Geneva. 2011. Available at: www.who.int/neglected_diseases/zoonoses/en/.
2. Renukaradhya GJ, Isloor S and Rajasekhar M. 2002. Epidemiology, zoonotic aspects, vaccination and control/eradication of brucellosis in India. Vet Microbiol 90: 183-195.
3. OIE Terrestrial Manual 2016. Brucellosis Chapter 2.1.5. Version adopted by the World Assembly of Delegates of the OIE in May 2016.
4. McDermott JJ, Arimi SM. Brucellosis in Sub-Saharan Africa: epidemiology, control and impact. Vet Microbiol. 2002; 90: 111-134.
5. McDermott J, Grace D, Zinsstag J. Economics of brucellosis impact and controlling low-income countries. Rev Sci Tech. 2013; 32: 249-261.
6. Cvetnic Z, Spicic S, Curic S, Jukic B, Lojkic M, Albert D, Thiébaud M, Garin-Bastuji B. Isolation of *Brucella suis* biovar 3 from horses in Croatia. Veterinary Record-English Edition. 2005;156:584-585.
7. Lovejoy GS, Carver HD, Moseley IK, Hicks M. Serosurvey of dogs for *Brucella canis* infection in Memphis, Tennessee. American Journal of Public Health. 1976;66:175-176.
8. Cameron HS, Hughes EH, Gregory PW. Genetic resistance to brucellosis in swine. Journal of Animal Science. 1942;1:106-110.
9. Blasco JM. Control and eradication programs of brucellosis in small ruminants and cattle. Implementation of Control and Eradication Programs in Animals. Zaragoza; Curso de Epidemiologia; 2001.
10. Coelho AM, Díez JG, Coelho AC. Brucellosis en pequeños rumiantes: efecto de la aplicación de un programa especial de vacunación en masa con REV-1. REDVET. Revista Electrónica de Veterinaria. 2013;14:1-16.
11. Petersen E, Rajashekara G, Sanakkayala N, Eskra L, Harms J, Splitter G. Erythritol triggers expression of virulence traits in *Brucella melitensis*. Microbes and Infection. 2013;15:440-449.

Participants and Organizing team



Dr. Md. Nure Alam Siddiky
Senior Program Officer
SAARC Agriculture Centre
BARC Complex, Farmgate
Dhaka-1215, Bangladesh
Cell: +88 01716475486
Email: nasiddiky.saarc@gmail.com



Dr. Md. Rezaul Karim
Scientific Officer
Bangladesh Livestock Research Institute
Savar-1341, Dhaka
Bangladesh
Cell: +88 01717320308
Email: rezavetmicro@yahoo.com



Dr. B.P. Shankar
Scientist-II
South Regional Disease Diagnostic
Laboratory
Bengaluru, India
Cell: +91 9844287557
Email: drshankarbp@gmail.com



Dr. M D N Jayaweera
Director/Animal Health
Dept. of Animal Production & Health
P.O Box 13, Getambe, Peradeniya
Sri LankaCell: +94 71 8033463
Email: mdnjayaweera@yahoo.com



Dr. Monsur Ahmed
Veterinary Surgeon
Department of Livestock Services
Dhaka, Bangladesh
Cell: +88 01717880715
Email: monsurcvu09@gmail.com



Dr. Sonam Pelden

Veterinary Officer
Dzongkhag Veterinary Officer,
Department of Livestock
Zhemgang, Bhutan
Cell: +975-17634786
Email: spvet87@gmail.com



Dr. Rajib Das

Veterinary Surgeon
Department of Livestock Services
Dhaka, Bangladesh
Cell: +88 01710464472
Email: dasrajib@gmail.com



Dr. Sudam Bag

Assistant Director
National Institute of Animal Health
Baghpat, U.P. India
Cell: +91 7060813090
Email: sudambag.niah@gov.in



Dr. Sulav Shrestha

Department of Livestock Services
Nepal
Cell: +977 9845151324
Email: sulavvet@gmail.com



Dr. Sonam Chopel

Senior Veterinary Officer
Dzongkhag Veterinary Hospital
Department of Livestock
Dagana, Bhutan
Cell: +975-77887910
Email: schophel@dagana.gov.bt



r. SurajSubedi

Department of Livestock Services
Nepal
Cell: +977 9840067696
Email: suraj_subedi@ymail.com



Dr. Muhammad Riazul Islam

Assistant Director
Rural Development Academy
Bogra, Bangladesh
Cell: +88 01716012709
Email: mrislam82@gmail.com



Dr. AnirbanGuha

Livestock Officer
Department of Animal Husbandry,
Dairying and Fisheries
Ministry of Agriculture and FarmersWelfare
New Delhi, India
Cell: +91 983641948
Email: dranirban.dadf@gmail.com



Dr. Parimal Roy

Director and
Course Coordinator-SAARC Training Programme
ICAR-NIVEDI, Bengaluru-64
India



Dr S S Patil

Principal Scientist and
Course Coordinator-SAARC Training Programme
ICAR-NIVEDI, Bengaluru-64
India
Cell: +91 9448759965



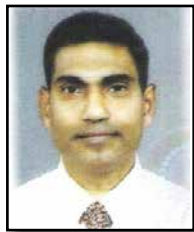
Dr K.P Suresh

Principal Scientist and
Course Coordinator-SAARC Training Programme
ICAR-NIVEDI, Bengaluru-64
India



Dr. Jagadish Hiremath

Scientist and
Course Coordinator-SAARC Training Programme
ICAR-NIVEDI, Bengaluru-64
India



Dr. L W B Epakanda

Veterinary Surgeon
Animal Health Division
Dept. of Animal Production & Health
P O Box 13, Getambe, Peradeniya
Sri Lanka
Cell: +94 77 1111675
Email: lakshith63@yahoo.com



Dr Mohd. Mudassar Chanda

Scientist and
Course Coordinator-SAARC Training Programme
ICAR-NIVEDI, Bengaluru-64
India



**ICAR-National Institute of Veterinary Epidemiology
and Disease Informatics (NIVEDI),**

Post Box No.6450, Ramagondanahalli, Yelahanka, Bengaluru-560064, Karnataka, India.

Ph: +91-80-23093110, 23093111, Fax: +91-80-23093222

Website: www.nivedi.res.in, Email: director.nivedi@icar.gov.in