

DISCLAIMER

This resource is a Field Epidemiology Manual in PDF format.

It aims to support field epidemiologists on their field or desk assignments. This content was created by a community of epidemiologists between 2010 and 2018 in a Wiki format.

ECDC had a community support role in this activity and **takes no responsibility for the accuracy or completeness of the content.**



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

FEM Wiki home








Last modified at 8/17/2021 11:44 AM by Rodrigo Filipe

FEM Wiki is an open information sharing platform for public health experts, hosted and funded by ECDC. The content of FEM Wiki is provided by users of the platform and does not necessarily represent opinion of ECDC. By contributing content to FEMWIKI, users agree to the conditions as described in the [legal notice](#).

The FEM Wiki site is a manual on Field Epidemiology publicly available in Wiki format, organised by concept in a book structure. You can browse the taxonomy using the menu on the left.

Any user can read the wiki articles which were approved. If you want to enter any content you must login on the right upper corner and ensure you have subscribed the FEM Wiki application in the [ECDC subscriptions page](#).

Use the bookmarks on this PDF to navigate to specific articles.
Below is an illustration of the navigation previously available in FEM Wiki.

- ▲  FEM Wiki
 - ▲  Navigation
 - ▷  Assessing the burden of disease and risk assessment
 - ▷  General Communication
 - ▷  Infection control and hospital hygiene
 - ▷  Introduction to Public Health and basic concepts
 - ▷  Statistical Concepts

- ▲ ☐ Assessing the burden of disease and risk assessment
 - ▲ ☐ Descriptive data analysis
 - ☐ Analysis by person characteristics
 - ▷ ☐ Analysis by place characteristics
 - ▷ ☐ Analysis by time characteristics
 - ☐ Spatial Analysis (Geographical Information Systems)
 - ▲ ☐ Epidemic Intelligence
 - ▷ ☐ Rapid Risk Assessment
 - ▷ ☐ Surveillance - Principles
 - ▲ ☐ Field Epidemiology
 - ☐ Combining Studies: Meta-Analysis
 - ☐ Key definitions in infectious diseases epidemiology
 - ▷ ☐ Measurement in Field Epidemiology
 - ▷ ☐ Methods in Field Epidemiology
 - ▷ ☐ Outbreak Investigations
 - ☐ Formal Risk Assessment
 - ▲ ☐ Health Economics
 - ☐ Cost-effectiveness
 - ☐ Cost-minimisation
 - ☐ Cost-utility
 - ☐ Measuring the Quality of Healthcare
 - ▲ ☐ Screening
 - ☐ Evaluation of Screening
 - ☐ Measures of Accuracy in Screening
 - ☐ Principles of Screening

- ▲ ◻ General Communication
 - ◻ Aide memoire for oral presentations and visual aids
- ▲ ◻ Health communication
 - ◻ Crisis communication
 - ◻ Health advocacy
 - ◻ Health education
 - ◻ Health literacy
 - ◻ Outbreak communication
 - ◻ Risk communication
 - ◻ Social marketing
- ▲ ◻ Writing a Study Protocol
 - ▷ ◻ Analysing a public health problem
 - ◻ Checklist for study protocols
 - ◻ Data collection instrument
 - ▷ ◻ Questionnaire Design
- ▲ ◻ Writing for Publications
 - ◻ Checklist for posters
 - ◻ Scientific paper review form
- ◻ Writing for Stakeholders



Infection control and hospital hygiene

- Antimicrobial stewardship
- Healthcare-associated infection prevention and control as a part of patient safety programme
- Healthcare-associated infection prevention and control programme
 - Cooperative learning as active learning in adult
 - Education and training of staff
- Healthcare-associated infections and risk assessment
 - Burden of HAIs
 - ◦ Main pathogens and resistance
 - ◦ Types of HAIs
- Infection control interventions
 - Cohorting patients and/or staff
 - Contact precautions
 - Developing infection control interventions: isolation
 - Environment as a source of healthcare-associated infections
 - ◦ Hand hygiene
 - Hospital cleaning and decontamination
 - Immunization of health care workers
 - Isolation of patients with Clostridium difficile infection (CDI)
 - Methods for assessing environmental cleanliness
 - Microbial typing
 - Molecular epidemiology
 - Pre-emptive isolation
 - Reprocessing of flexible endoscopes
 - Standard precautions
 - ◦ The role of the clinical microbiology laboratory in infection prevention and control
- Surveillance and investigation of healthcare-associated infections
 - ◦ Case definitions of healthcare infections
 - ◦ European surveillance of healthcare-associated infections
 - ◦ Surveillance of healthcare-associated infections - definition
 - ◦ Surveillance of healthcare-associated infections - other approaches

▲ (+) Introduction to Public Health and basic concepts

▲ (+) Public Health Informatics

- (+) Apps for epidemiologists
- ▷ (+) Data Capture Systems
- ▷ (+) Health Informatics Standards ? general introduction
- (+) Software for epidemiologists
- (+) Weblinks for epidemiologists

▲ (+) Public Health Interventions

- (+) Contact tracing
- (+) Host
- ▷ (+) Prevention
- (+) Reservoir for infectious agents
- (+) Source of infection
- (+) Transmission routes

▲ (+) Public Health Law

- (+) Brief history of International Communicable Disease Law
- ▷ (+) EU Public Health Legislation
- ▷ (+) International Health Regulations 2005 edition

▲ (+) Public Health Microbiology

- ▷ (+) Applied Immunology
- (+) Practical Aspects of Specimen Collection and Shipment
- ▷ (+) The Role of Microbiology
- (+) Types of Microorganisms
- (+) Public Health Programs

- ▲ ◻ Statistical Concepts
- ▲ ◻ Models
 - ▲ ◻ Logistic Regression
 - ◻ Estimating Odds Ratios in the presence of interaction
 - ◻ Fitting logistic regression models
 - ◻ Interpreting model coefficients
 - ◻ Linear models
 - ◻ Model building strategies
 - ◻ The logistic model
 - ▲ ◻ Stratified Analysis
 - ◻ The Mantel Haenszel Method
 - ◻ Power and Sample Size
 - ◻ Presenting Statistics
- ▲ ◻ Sampling
 - ◻ Concepts in sampling
- ▲ ◻ Statistical Tests
 - ▲ ◻ Significance and Confidence
 - ◻ Confidence Intervals
 - ◻ Significant probability to be different from the expected
 - ◻ The idea of Statistical Inference



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Updated
Pages

User
manual
for
editors

Legal
notice

Help and
Support

About

User
manual
for
approvers

User manual for readers

Getting Started (non-authenticated user)

The FEM Wiki site is organised into: **Documents** (formally articles) and **Discussions**. There is also a section with **Help & Support** (you are here now) the **Legal notice** and '**About**' FEM Wiki.

Search

Refinement panel

Search results can be filtered by result type and modification date using the refinement panel.

Once you login to FEM Wiki, more filters will become available on the refinement panel.

Wild card

One of the best ways to search for something in FEM Wiki is by using the wild card search. That is when you don't know exact keywords, only part of the text you are looking for (i.e. first few letters). In the example below, I am searching for the keyword "vector" based on first few letters. As you can see, I am getting the results I expect, even though I didn't use the entire word.

Insert Image here!

For wildcard to work in FEM Wiki...

You have to start with the first few letters of the word. In other words, in a word "vector", you can't search for text "ect", it has to be "vec"

The wildcard character in FEM Wiki is `"**"`. You have to put the asterisk (wildcard character) after the first few letters, not before. For example, `vec*`, not `*vec`

You can use wildcard search with both Options (global search and library-level search)

Quotes

Use double quotes `""` to find exact phrases if you are sure about the phrases. Example: `"intervention epidemiology"`

Commands

You can use search commands (Boolean operators) to narrow or expand the search results. Note that all SharePoint search commands need to be writing in capitals.

Insert Image here!

Navigation in FEM Wiki (Browse using the taxonomy)

- Browsing is recommended to explore the content.
- Recommended browser is Google Chrome because browsing function is optimized.
- Browsing in is simple - Browsing out is tricky
- Search + browsing is possible.
- Viewing incoming links
 - You can see which pages link to the current page by clicking Incoming Links at the top of the page. This will show you any pages which link to this page.

Read / interpret content

There are some things you should be aware of when you are reading / interpreting the content in FEM Wiki.

- Non- authenticated users see only last author and last update date of approved pages.
- Authenticated users can see page history of approved pages and their unapproved pages.
- Reviewers can see all pages and the approve button for unapproved pages.

Create an account (including Privacy statement)

You will need an account if you want to do more than just read FEM Wiki. Registering is easy: click the "Sign in" link at the top-

right of any page and fill in your details. If you use your real name then it will be easier to approve the content you submit for publication.

Filling in your profile

Once you have registered, you will have a profile page that you can always get to by clicking your name at the top of the window.

Here you can display your:

- Interests or areas of speciality.
- Affiliation, e.g. your institution or company.
- Contact details, including links to profiles on other social networks.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Assessing the burden of disease and risk assessment

Last modified at 3/22/2017 8:26 PM by Vladimir Prikazsky

Effective disease prevention and control depends on several factors that all have to be present and work together in the community. It all starts with the ability to **detect threats** to the health of the population. A threat can be seen as an undesirable situation that has *not yet* occurred, but that *may* happen unless protective measures are taken. The ability to detect a disease threat implies that we already have basic knowledge about the 'normal occurrence (or **burden**)' of this disease in the population.....

Assessing the burden of disease requires a public health workforce with the competence to collect, analyse and interpret health data from "your" population plus the infrastructure in the health care system that allows access to relevant data. Methods used in [Field Epidemiology](#) play a central part in assessing the burden of disease.

To detect health threats requires (in addition to the above) continuous monitoring of burden of disease information of 'your own' and surrounding populations, trends in risk behaviour, characteristics of pathogens (e.g. development of antimicrobial resistance) plus competent staff responsible for continuous collection, analysis and interpretation of information. The process aimed at detection of health threats is sometimes referred to as [epidemic intelligence](#).

Once health threats have been detected and validated, information needs to be shared with "those who need to know" in the health system. This usually requires translation of specific epidemiology jargon in a format that can be used by policy and decision makers in order to decide on [interventions](#) (preventive and control measures).

This part of this FEMWiki addresses methods that can be used to assess the health status of the population and detection and assessment of health threats. Methods for [Surveillance](#), Risk Assessment and [Outbreak Investigations](#) will be described in this section.

[Interventions](#) (public health measures, policy making and decision taking) is a topic described in another part of the FEMWIKI. [Communication](#) is yet another topic.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

[General Communication](#)

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

General Communication

Last modified at 4/17/2015 6:50 PM by Arnold Bosman

For purposes of this FEMWIKI content on methods and principles in disease prevention and control, we will consider two specific aspects of communication; science communication (1) and health communication (2).

Science communication includes

- public communication presenting science-related topics to non-experts (including [writing study protocols for stakeholders](#))
- communication between scientists (e.g. oral presentation such as at conferences or [written presentation through scientific journals](#))
- communication between non-scientists on science-related topics.
- science exhibitions, journalism, policy or media production

Science communication can aim to generate support for scientific research or study, or to inform decision making, including political and ethical thinking.

[Health communication](#) is the study and practice of communicating promotional health information, such as in public health campaigns, health education, and between doctor and patient. The purpose of disseminating health information is to influence personal health choices by improving health literacy.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

[Infection control and hospital hygiene](#)

Introduction to Public Health and basic concepts

Statistical Concepts

Infection control and hospital hygiene

Last modified at 10/12/2016 12:52 PM by Vladimir Prikazsky

The Infection Control/Hospital Hygiene (IC/HH) Wiki is an online platform which aims to store, share and optimise knowledge on infection prevention and control and hospital hygiene in a flexible and collaborative way.

It's ambition is to become:

- a tool for supporting training in the field of IC/HH;
- a repository for IC/HH training materials;
- an opportunity to compare multiple points of view and experiences concerning contemporary IC/HH issues to stimulate future research and complement the evidence base;
- a place to lodge discussions concerning the focus of IC/HH field training, competencies and other aspects of IC/HH practice;
- a place to share new developments in IC/HH methods.

Following chapters of the IC/HH wiki were agreed by EU Member State experts on infection control training:

- [1. Healthcare-associated infections and risk assessment](#)
- [2. Healthcare-associated infection prevention and control programme](#)
- [3. Healthcare-associated infection prevention and control as a part of patient safety programme](#)
- [4. Surveillance and investigation of healthcare-associated infections](#)
- [5. Infection control interventions](#)
- [6. Antimicrobial stewardship](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

[Introduction to Public Health and basic concepts](#)

Statistical Concepts

Introduction to Public Health and basic concepts

Last modified at 7/27/2014 5:10 PM by Arnold Bosman

Definition of public health

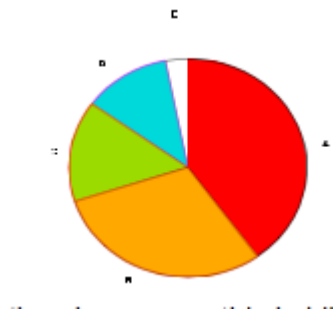
"... the science and art of preventing disease, prolonging life and promoting health through the organized efforts and informed choices of society, organisations, public and private, communities and individuals." Though this definition is over one century old, it remains valid, despite all societal changes. Public health is a field where many disciplines and many professional sectors collaborate.

Relevant disciplines and sectors within public health:

- [epidemiology](#)
- [microbiology](#)
- [biostatistics](#)
- health services
- environmental health
- community health
- behavioural health
-
- [informatics](#)
- [public health interventions](#)
- public policy

Mortality statistics and health priorities

Which diseases are important for public health? Are they those that most people die from? In that case we should concentrate on preventing/curing the main causes of death. Or....? There is always 'a most common cause of death' in the population. Take a look at this pie-chart that represents the proportional distribution of mortality in an hypothetical population for 5 causes of death :



In the above case this is 'disease A in red'. Yet imagine what happens when we successfully eliminate that disease..... then there will be another disease that will be the biggest:



This is easy to imagine when we consider that no man is immortal: "mortality in the EU is 100%". This creates a fundamental question in public health: *Which would we consider the best distribution of causes of death? or: How do we want to die?*

Perhaps in all of human history, we are among the first generations that actually may have an active choice about the distribution of causes of death in society. In the EU, half of those who die are over 75-80 years old. Their causes of death will greatly influence mortality statistics. What role will this play for research, prevention, resources?

Measuring the health of the population.

How can we measure 'the health of the population'?

One indicator is the average life expectancy. Surely a population where people die before they are 40 can be considered less 'healthy' than one where most people reach 80? Perhaps, but the 'average age' is very much influenced by child mortality: if half of the children that are born die at birth and the rest lives exactly until 100 years, then the average life expectancy is 50 years.

And it appears that child mortality (under 5 years old) is also a very good indicator of the population health.



(Image from the Gapminder)

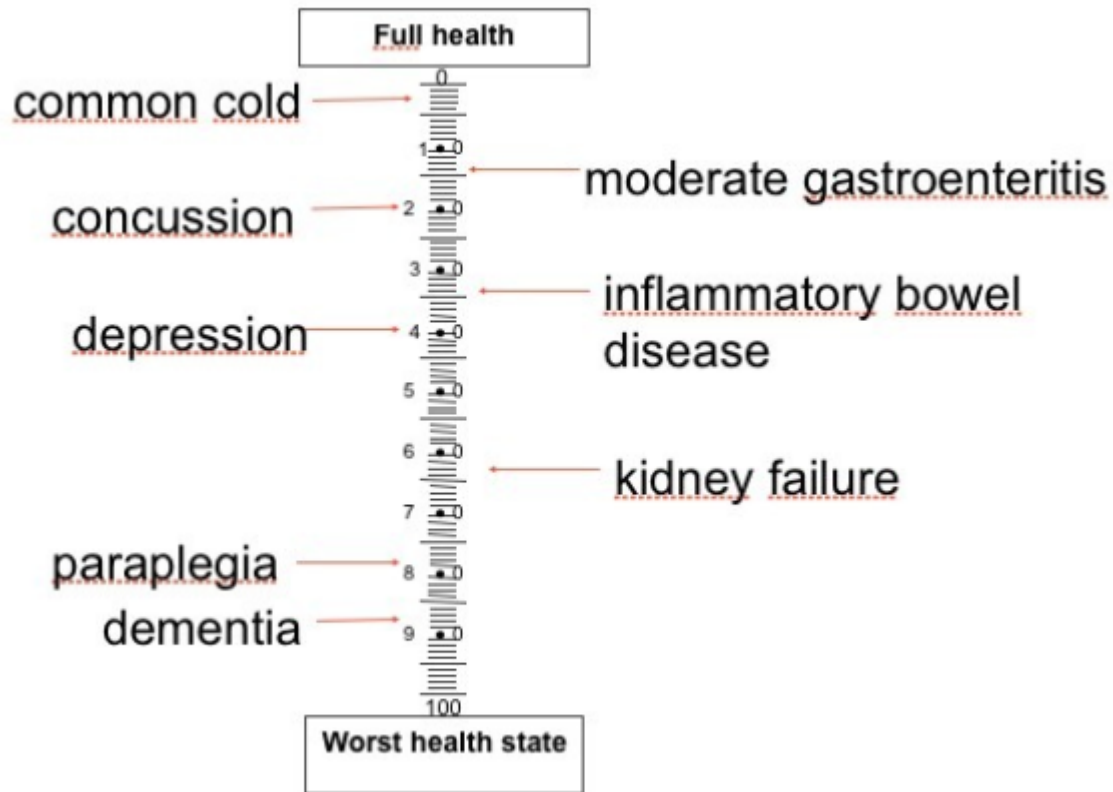
However, if we all spend the years we live while suffering from chronic diseases, could we still call that 'more healthy'?

If we want to correct for years spent with diseases, the indicator of DALY (disability adjusted life years) is often used.

DALY's are based on two things: (1) Life expectancy and (2) Number of years with less than full health.

But how do you measure disability?

On a scale from 0 to 100, where 0=death and 100 is 'total health', how much are various ill healths? One way of finding out is to interview people, and ask them to indicate on that scale how they would feel if they had a common cold, or diarrhea, or when they lost an arm or a leg, or become paralyzed....



If you would ask many people, then the 'average scale' could look like the one above. It also appears that the resulting scales differ a lot between countries. By attaching a DALY value for each disease, and measuring how common this disease is in the population the actual burden of this disease can be assessed. A useful tool for public health (and sometimes quite surprising).

Some important Definitions

When reading general texts in public health, you will encounter certain epidemiological indicators often. Here is an overview of the most common ones, and their definition. Advanced readers may use the hyperlinks to read more detailed descriptions about these indicators.

Incidence (or **cumulative incidence**): the number of new cases per population. Usually per 100.000. Outbreaks usually show a higher incidence; the incidence is then expressed per thousand or per hundred (%). In outbreaks, cumulative incidence is often called '[attack rate](#)'.

= number of new cases of the disease per population and per time unit (most often per year and per 100,000 people)

"The incidence rate of chlamydia in the EU is 175 per 100,000 per year"

"The incidence of vCJD in the EU is 1 or 2 cases per year."

Prevalence = proportion of the population that have the disease right now

"The prevalence of high blood pressure in US adults is around 30%"

"Tuberculosis is eliminated from a country when prevalence is less than 1 per million." **Morbidity** = almost the same as prevalence

Lethality = proportion of people who get a disease that will die from it (also called '**case fatality**')

Mortality = number of deaths per population (most often per 100.000)

Proportional mortality = of all people who died, what proportion have died from this disease

Mortality of Infectious Diseases in the European Union is quite low, except pneumonia (several % of all deaths) and influenza (0 – several thousands of deaths). This picture is very different from the rest of the world: globally, around 62 million people die each year, 25% of them from infectious diseases.

Globally, relevant infectious diseases causing significant mortality are:

Pneumonia	4.2
AIDS	3.0
Diarrhoea	2.0
Tuberculosis	1.7
Malaria	1.4
Measles	0.2

Followed by neonatal tetanus, pertussis, meningitis, syphilis and hepatitis. Even though they seem an important public health problem globally, so why are they important to the EU? In the next paragraph on the origins of public health, we try to find the answer, which lies in the success of prevention.

Origins of Public Health:

Though history shows many descriptions of preventive advice from various medical scholars, and social response against serious **outbreaks** included quarantine, the concept of 'population health' probably took off with the sanitation movement in the second half of the 19th Century. This led to promotion of clean water, sewage works, better ventilation, more light in the houses and work places and in general cleanliness, personal and in home.

In addition, improvements such as Safety at work, Vaccines, Food control were developed: it is clear that most of these interventions were very much concentrated on infectious disease. Perhaps we can consider this as 'classic public health':

- Mainly controlling risk factors that people could not be expected to control themselves
- Promoting better education, better housing, etc.

Aimed at very large population groups

- Often linked to socio-liberal political movements

‘Modern public health’ seems much more targeted at individual human behavior; something each of us should be able to influence:

- smoking
- eating
- drinking
- exercise
- sexual health
- tanning

Both the classic and the modern approach have in common that public health involves almost all areas of society and attempts to:

- Identify the main determinants of ill health
- science-based action to eliminate or prevent them

In our current EU Law,

Maastricht Treaty, 1992, Article 129:

"The Community shall contribute towards ensuring a high level of human health protection .. Community action shall be directed towards the prevention of diseases... Health protection requirements shall form a constituent part of the Community's other policies."

Societal changes that influence the infectious disease panorama

- increased international traffic (goods, people)
- ecological changes (global warming/extreme weather)
- new production methods (for food, mainly)
- microbial adaptation (= resistance)
- changing contact patterns (mega-cities, sexual contacts)
- societal breakdown (conflicts, recession)

With all the above, it may be clear that in the EU we have been successful in pushing infectious diseases back as a significant cause of mortality. However, we also see that around the world, there is still many infectious diseases that are ‘alive and kicking’. When we look at the factors behind our success, we notice that our ever changing society may open many future doors for infectious diseases to come back in.

Final remarks:

- The present 'victory' of the rich countries over the infectious diseases is fragile
- The way we live is as decisive for infections as characteristics of microbe or host
- There will be new ones.....

References:

- Winslow, Charles-Edward Amory. "The Untilled Fields of Public Health", *Science*, 1920; 51(1306): 23–33.
- www.gapminder.org: for a fact based world view. Map on child mortality and country income, 2011 data
- [the Maastricht Treaty on the European Union \(1992\)](#)
- Lecture by Johan Giesecke in "ECDC Public Health 101 course", October 2013



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

[Statistical Concepts](#)

Statistical Concepts

Last modified at 4/12/2016 7:25 AM by Vladimir Prikazsky

In this part we describe concepts such as confidence interval, sampling, sample size and study power. The chapter is incomplete and deserves additional content. You are very welcome to add content, so that this part of the FEMWIKI contains key methods and concepts in statistics that epidemiologists should understand in order to interpret results of data analysis.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Field Epidemiology

Last modified at 6/18/2014 3:33 PM by Arnold Bosman

The subject of epidemiology is "about the investigation of causes of health related events in the populations" (Morabia, 2004), and is as a scientific discipline relatively recent. Epidemiology can be exercised in a variety of health contexts such as clinical health care and research. It is considered a cornerstone discipline in Public Health, aimed to generate an evidence base for policy and decision making for healthy populations.

Though the scientific discipline of epidemiology is relatively new, knowledge about factors that influence epidemics dates much further back in medical history. Hippocrates described already his intuition of the association between environmental factors, human behaviors and disease, even when most people in his time (in the 4th century b.C) believed in supra-natural causes of disease. ([See also the discussion forum on this](#))

Applied Epidemiologists, or "Field Epidemiologists", use science as the basis for intervention programmes designed to improve public health (White, 2001). Such application may then also be called '*intervention epidemiology*' and even the term '*consequential epidemiology*' has been phrased. This branch of epidemiology has its origin in the post graduate public health residency programme of the [Centers for Disease Control and Prevention](#) (CDC, Atlanta, USA): the '[Epidemic Intelligence Service \(EIS\)](#)', that started in 1951. Since then, many countries have set up Field Epidemiology Training Programmes (FETP) and in 1995 the [European Programme for Intervention Epidemiology Training \(EPIET\)](#) started to train the first cohort of fellows in the 2 year full time curriculum, funded by the European Commission.

Field epidemiology aims to apply scientific methods in day to day public health field conditions in order to generate new knowledge and evidence for decision making. The context is often complex and difficult to control, which challenges study design and interpretation of study results. However, often in Public Health we lack the opportunity to perform controlled trials and we are faced with the need to design observational studies as best as we can. Field epidemiologists use epidemiology as a tool to design, evaluate or improve interventions to protect the health of a population.

For example, think of a sudden increase of mumps among teenagers in a country with high vaccine coverage. A key

question to answer quite early after detecting such event is: is the mumps outbreak caused by a vaccine failure or a failure to vaccinate? Obviously a vaccine failure would require a different set of public health actions than a failure to vaccinate. Field Epidemiology includes the ability to rapidly design and execute an appropriate epidemiological study to generate reliable answers to key questions, timely enough to allow control measures to be effective. And for such a study the epidemiologist will need to get out of the office and into the field or onto the streets. For this reason, field epidemiology has yet another synonym: *shoe-leather epidemiology*, as opposed to *armchair epidemiology*.

Traditionally, core activities in field epidemiology are , design, operation or evaluation of communicable disease **surveillance systems** and **field research** to study risk factors for or distribution of communicable diseases. Since communicable diseases may rapidly spread in populations they often create the urgent need to get answers about risk factors, risk groups and effective ways to intervene. Time pressure, media attention and anxiety among public and decision makers create a classical context for the field epidemiologist to work in. They are often required to create ad hoc teams to help them with the investigations, and often they will have to instruct and train new team members to perform the tasks that are required for the investigation

So it makes sense that they need to have excellent and to have competences in **teaching**. It will therefore not be surprising that the traditional FETP requires the following training achievements:

- Perform an epidemiological [outbreak investigation](#)
- Design, evaluate or coordinate a [surveillance system](#)
- Design and perform an [epidemiological field study](#) (research)
- Scientific communication
- Teaching

References:

1. Alfredo Morabia (Editor). A History of Epidemiologic Methods and Concepts. 2004. ISBN 3-7643-6818-7
2. Hippocrates, "On air, people and places", (460-377 BC)
3. Mark E. White, Sharon M. McDonnell, Denise H. Werker, Victor M. Cardenas, Stephen B. Thacker. Partnerships in International Applied Epidemiology Training and Service 1975-2001. American Journal of Epidemiology, Vol. 154, No. 11, 2001
4. Moren A, Drucker J, Rowland M, Van Loock F. European Programme for Intervention Epidemiology Training (EPIET). Rev Epidemiol Sante Publique, 1998, 46(6):p. 533-4
5. Greg, Michael B. (Editor). Field Epidemiology. Oxford University Press, new York, 1996

Synonyms used

Field epidemiology - Intervention epidemiology - Applied epidemiology - shoe leather epidemiology - consequential epidemiology



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

Introduction to Public Health and basic concepts

General Communication

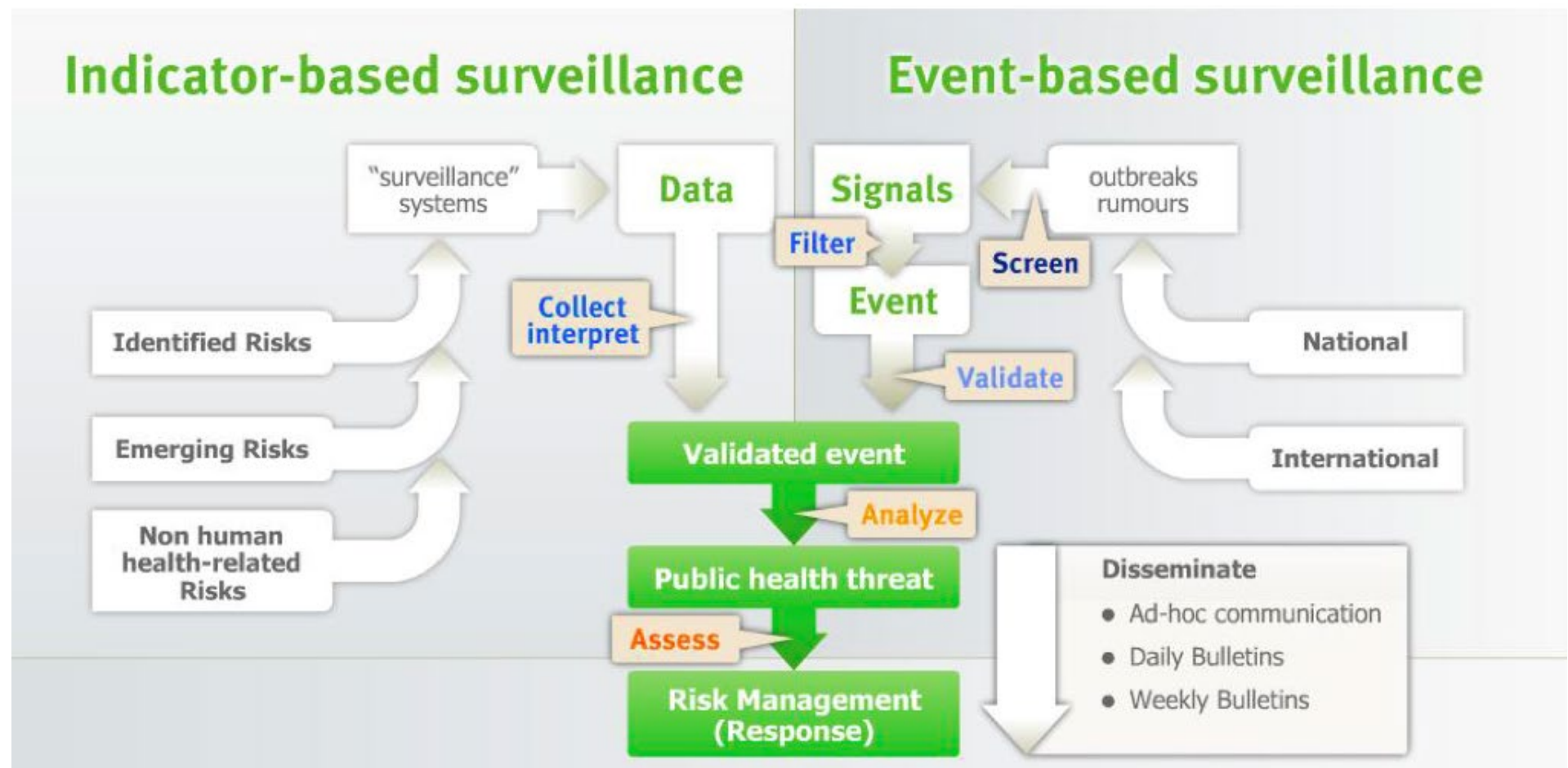
Statistical Concepts

Infection control and hospital hygiene

Epidemic Intelligence

Last modified at 4/3/2019 4:23 PM by Rodrigo Filipe

Epidemic intelligence integrates indicator-based and event-based components. 'Indicator-based surveillance' refers to structured data collected through routine [surveillance systems](#). The '[event-based surveillance](#)' refers to unstructured data gathered from formal and informal sources, such as the media and scientific publications. The purpose of both these components of epidemic intelligence is to quickly identify any event which might become a public concern. The epidemic intelligence covers [risk assessment](#) and risk monitoring.



The objective of epidemic intelligence is to produce timely, validated and actionable intelligence on events related to communicable diseases or of unknown origin that are of interest for public health and health authorities. The process can be divided into early detection of new threats and events, and monitoring the threats that have already been identified, including potential threats.

Early detection comprises six elements:

1. *Screening* news, official reports or notes and rumours relevant from a European perspective in order to distinguish the meaningful information signals by applying specified criteria.
2. *Filtering* the events to identify potential public health events of European interest.
3. *Validating* the events that originate from unofficial sources, by cross-checking with official and/or reliable media sources to ensure that the event detected is real and fully understood.
4. A validated event will then be *analysed* to capture the full information available about the event, including epidemiological data, facts related to exposures and contextual information.
5. Based on the analysis, an *assessment* is made to estimate the risk associated with the event.
6. Finally, *communication* and *documentation* of the identified threats are an integral part of the epidemic intelligence, throughout the five steps above.

Monitoring identified threats refers to the active follow-up of all relevant information directly related to the concerned threat. This iterative process continues until the threat is considered to have subsided or until all appropriate public health measures have been implemented.

Of course in a rapidly evolving situation professional judgement should be exercised and it may be appropriate under severe time constraints to skip some of the above-mentioned steps in order to quickly share information. However, if epidemic intelligence can be gathered systematically as described above, the outcome is a better informed decision and more effective action.

Reference:

[Tutorial ECDC on Epidemic Intelligence](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

[Introduction to Public Health and basic concepts](#)

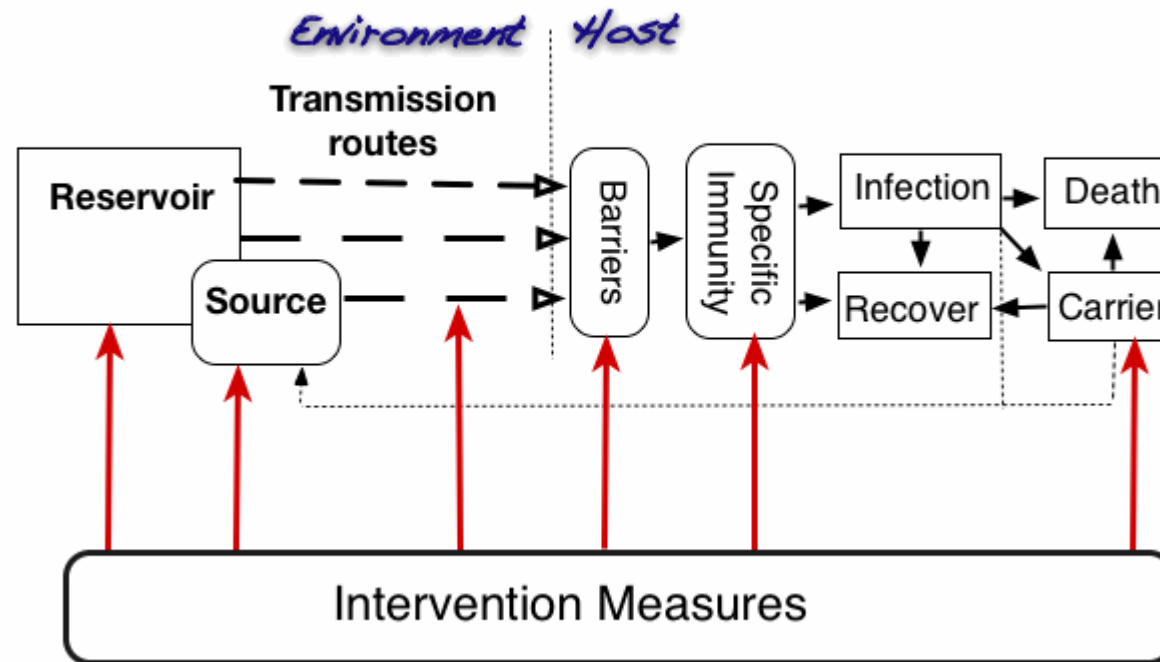
Statistical Concepts

Public Health Interventions

Last modified at 12/7/2016 11:20 AM by Vladimir Prikazsky

Interventions in communicable disease control are by preference guided by evidence, coming from existing scientific knowledge and newly generated by [assessment methods for burden of disease](#) and [field epidemiology](#).

Interventions are aimed to prevent or interrupt transmission of an infectious agent to [hosts](#) and may be proactive (*prevention measures*) or reactive (*control response measures*). Each of those measures can be targeted at critical control points' in the **chain of transmission** of infectious agents (see figure).



For example, measures targeting *reservoirs* may include rodent control / extermination in case of leptospirosis or culling of poultry flocks in case of avian influenza. A large portion of control measures include measures targeted at specific *sources*, for example remove botulism contaminated olives from the market or searching for an infectious tuberculosis patient in order to treat the infection.

Examples of measures targeted at include *vector* control, behavioural education (e.g. safe sex, safe cooking, promotion of hand hygiene) and treatment of drinking water. Interventions targeting *barriers* around the *host* could include provision of personal protective equipment, treatment and covering of skin lesions. Among the most well known intervention measures are immunisations (vaccinations and prophylaxis through immunoglobulins) to enhance the *specific immunity* against micro organisms.

Finally, treatment and / or quarantine of *infectious patients* or *carriers* ensure that risk of person to person transmission is reduced in certain diseases.

Question:

What is the difference between preventive measures, response control measures and intervention measures?

Reflection:

In various textbooks and articles, these terms are used in different ways. In the FEMWiki, we prefer the following distinction:

Preventive interventions in communicable disease control:

1. **Primary prevention:** Intervention in the population, targeting healthy population (i.e. not infected) at risk in order to avoid infection (e.g. immunisations)
2. **Secondary prevention:** intervention in the population infected, in order to mitigate symptomatic disease or avoid complications (e.g. [screening](#))
3. **Tertiary prevention:** intervention targeting the population with disease symptoms aimed at limiting the impaired functions in daily life or society, due to the disease (e.g. rehabilitation regimes for paralytic polio patients)

In addition to these 3 levels of preventive interventions, a more generic level is often used: [primordial prevention](#), which includes generic measures such as improvements in civil engineering (clean drinking water, sewage systems etc).

Communicable disease **control response measures** that do not include primary, secondary or tertiary prevention are for example treatment or quarantine of carriers, culling of infected poultry flocks or removing certain food items from the market.

However it is important to be aware that these classifications of prevention are under debate in public health, since they do not offer a complete framework for disease prevention and control, and several alternatives have been suggested.

References

1. Barbara Starfield. Public Health and Primary Care: A Framework for Proposed Linkages. Am J Public Health, 1996 Oct.;86(10):1365–1369.
2. B Starfield, J Hyde, J Gervas, I Heath. The concept of prevention: a good idea gone astray? J Epidemiol Community Health 2008;62:580–583. doi:10.1136/jech.2007.071027



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Surveillance – Principles

Last modified at 9/28/2015 1:00 PM by Arnold Bosman

Summary

The objective of the lecture is to understand: the important role of surveillance in public health (and clinical) practice; how to link the design of a surveillance system to the objectives that it serves; what types of data might be used for surveillance; approaches to data collection, analysis and reporting; and the need for quality and governance standards, as well as clear operating procedures. The lecture includes brief descriptions of examples of systems that exemplify many of the principles outlined, and notes common pitfalls in the design and operation of surveillance systems.

Surveillance is first and foremost a process for producing information that will trigger, inform or be used to evaluate defined public health (or clinical) action. If there is no clear and immediate link between the information output of an activity and existing or planned public health action then it is unlikely that the activity is surveillance.

Public health action that is informed by good quality surveillance is likely to be both more effective and more efficient than action that is undertaken in the absence of surveillance. A good surveillance system should provide timely, accurate and relevant interpreted information while at the same time minimising the burden placed on data providers.

[The Role of Surveillance](#)

[The Components of Surveillance](#)

[Objectives of Surveillance – Inputs and Outputs](#)

[Surveillance or Research?](#)

[Criteria for Surveillance](#)

Sources and Types of Data

Surveillance System Design and Operation

Quality, Governance and Operating Procedures

Analysis, Interpretation and Dissemination

Opportunities and Challenges

Learning Objectives

After reading this chapter, you will be better able to:

- understand the role of surveillance in public health practice, and how this differs from research
- describe the component activities of surveillance
- define the objectives for a surveillance system and use these to determine (or evaluate) the inputs and outputs of a surveillance system
- understand the differences between surveillance and research
- understand the criteria for undertaking surveillance
- be aware of potential sources of data for surveillance
- understand how the characteristics of data and the practicalities of collecting those data need to be taken into account in the design and operation of a surveillance system
- understand the need for quality criteria, data governance and clear operating procedures for a surveillance system
- understand the importance of, and be aware of a range of approaches to, analysis, interpretation and dissemination of information as part of the surveillance process
- be aware of the challenges and opportunities posed by emerging threats and emerging technologies



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Outbreak Investigations

Last modified at 10/2/2015 6:40 AM by chwilliams

Outbreak investigations are often seen as key examples of [field epidemiology](#).

Definition of an outbreak

What exactly is an outbreak? Different definitions of an outbreak can be found in literature. Here we will define an outbreak simply as ***'the occurrence of more cases than expected in a particular population, in a specific geographical area and over a specified period of time'***.

Difference between an outbreak and an epidemic

The word 'epidemic' is often used interchangeably with 'outbreak' in professional communication. In the media, the word epidemic usually has a more threatening ring to it, which is why most communication experts tend to avoid using it. Outbreak investigators may decide to use the word 'epidemic' or not depending on whether they want to attract or deflect attention. In September 2014, president Obama made a careful distinction between the two terms to **attract** attention to the emergency that Ebola constituted. (1)

From the identification of a cluster to the establishment of the existence of an outbreak

To establish the existence of an outbreak, we first of all need to understand what is meant by ['cases'](#). This needs to be defined. Usually the first signal of an outbreak can come from a telephone call or report from the health care system about a **cluster** of cases. We may or may not yet know if such a **cluster** is 'more than expected', so a systematic approach is required.

At this stage it is important to understand the distinction between **a cluster** of cases (2 or more cases that are related by sharing similar characteristics in time and/or place and or personal characteristics) and an **outbreak** (more cases than expected). For example 5 cases of respiratory illness occurring in the same week can be considered 'a cluster in time', yet this may be the usual number of such cases that one would expect in that week of the year.

Similarly cases can be clustered in place (same village / neighbourhood etc) or according to personal characteristics (e.g. cases sharing the same age-category). In each of these examples of clusters, a key question will be: 'is this number of cases more than we would expect?'. If the answer is 'yes' then the cluster can be considered an 'outbreak'.

"More cases than expected", implies that we need to have knowledge of the 'normal' number of cases (or baseline). This knowledge may come from surveillance or surveys. The increase in the number of cases is best documented as a population-based incidence rate. Investigators may want to examine possible artifact in the numerator (e.g., batch reporting of old cases or of prevalent cases) or in the denominator (e.g., population movements, mass gathering).

Steps of an outbreak investigation

The various lists of various steps

Investigating an outbreak requires a systematic **approach** that is summarized as a number of steps. Unfortunately, various groups have generated various lists where steps differ in sequence and number. One of the most classical lists (2) includes **Ten steps** summarized as below:

1. **Establish the existence of an outbreak.** This may be achieved as described above by calculating rates, comparing the rate with the baseline and excluding artifacts in the numerator or the denominator.
2. **Confirm the diagnosis.** This may be achieved through (a) shortlisting a number of possible diagnoses on the basis of the frequency of signs and symptoms and (b) confirming the diagnosis, most often with support from the laboratory.
3. **Establish a case definition** (that needs to have time, place and person elements and that may have different levels of sensitivity and specificity, including possible, probable and confirmed cases) and **count cases** following a case search strategy that can be passive, stimulated passive or active, but that must be always homogeneous in the area considered.
4. **Orient the data in terms of time, place and person** through . This will lead to an epidemic curve (time), a map (place) and rates by age and sex (person).
5. **Determine who is at risk** of becoming ill (population at risk)
6. **Develop a hypothesis that explains the specific exposures that caused disease and test this hypothesis by appropriate statistical methods** (through **analytical studies**). An article by **Werber and Bernard published in Eurosurveillance** (3) describes the development of a toolbox consisting to increase the use of analytical studies in the investigation of outbreaks of food borne diseases. In general, analytical epidemiology may use case-control investigations (more adapted if the attack rate is low, under 5-15%) or cohort investigations (more adapted if the attack rate is not too low, above 5-15%)
7. **Compare the hypothesis with the established facts.**
8. **Plan a more systematic study** (environmental, microbiological etc)
9. Prepare a written report (**outbreak reporting**, to **communicate findings** to those who need to know)
10. **Execute control and prevention measures** (Recommend options for **interventions** based on the findings)

Some lists have referred to **13 steps** (4), adding additional steps such as logistical aspects (e.g., Prepare for field work) while another prepared for foodborne outbreaks was more conceptual with only **7 steps** (1. Detecting a possible outbreak, 2. Defining and finding cases, 3. Generating hypotheses about likely sources, 4. Testing the hypotheses, 5. Finding the point of contamination, 6. Controlling the outbreak and 7. Deciding an outbreak is over).

Since 2012, the EPIET and EUPHEM fellowships have used [an adaptation of the original 10 steps](#) for teaching and supportive supervision. These 10 steps are very similar from the one above, with minor adjustments derived from [an analysis of the common errors in outbreak investigations](#) (5). This adaptation has the advantage of disentangling (a) case definition (step 3) from case search (step 4) and (b) generation of hypothesis (step 5) from hypothesis testing (step 6). Case definition and case search on one side and hypothesis formulating and testing on the other side are quite different processes that can suffer from specific pitfalls and benefit from specific guidance (Hence the benefit in the split). In addition, the 10 steps adapted in such a way places the 'middle' of an investigation between hypothesis generating and testing. This reflects the pivotal thought process that needs to take place at that critical phase of the outbreak investigation when the outbreak investigation team may have to write a [mini-protocol](#).

How to understand the lists of steps of an outbreak investigation?

The lists of steps for outbreak investigations must not be taken to literally. First, they are ordered in a sort of logical sequence that does not necessarily match the temporal sequence. For example, some outbreak investigations may start with the enforcement of control measures (e.g., implementation of infection control in health care facilities to prevent secondary spread). Second, they summarize a number of steps that should take place for most investigations. However, some this may vary from investigation to investigation. Overall, they can be thought of as a list of 'things one wants to consider' while investigating an outbreak.

The outbreak team

Key to the investigation and control of an outbreak is the constitution of a [team](#).

References:

1. Remarks by the President on the Ebola Outbreak. <https://www.whitehouse.gov/the-press-office/2014/09/16/remarks-president-ebola-outbreak> (Accessed 11 May 2015)
2. Gregg, Michael B. (Editor). Field Epidemiology. Oxford University Press, New York, 1996
3. Werber D, Bernard H. Reducing the barriers against analytical epidemiological studies in investigations of local foodborne disease outbreaks in Germany – a starter kit for local health authorities. Euro Surveill. 2014;19(8):pii=20714.
4. US CDC. Principles of Epidemiology in Public Health Practice, Third Edition [An Introduction to Applied Epidemiology and Biostatistics](#). 2006, updated in 2011. Lesson 6; Section 2: Steps of an Outbreak Investigation.
5. Murhekar M, Moolenaar R, Hutin Y, Broome C. Investigating outbreaks: practical guidance in the Indian scenario. Natl Med J India. 2009. :252-6.
6. Heymann, David L. (Editor). Control of Communicable Diseases Manual, 19th Edition.

EPIET Lectures:

Epidemiological Outbreak Investigation

Operational Aspects of Outbreak Investigations



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

[General Communication](#)

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

General Communication

Last modified at 4/17/2015 6:50 PM by Arnold Bosman

For purposes of this FEMWIKI content on methods and principles in disease prevention and control, we will consider two specific aspects of communication; science communication (1) and health communication (2).

Science communication includes

- public communication presenting science-related topics to non-experts (including [writing study protocols for stakeholders](#))
- communication between scientists (e.g. oral presentation such as at conferences or [written presentation through scientific journals](#))
- communication between non-scientists on science-related topics.
- science exhibitions, journalism, policy or media production

Science communication can aim to generate support for scientific research or study, or to inform decision making, including political and ethical thinking.

[Health communication](#) is the study and practice of communicating promotional health information, such as in public health campaigns, health education, and between doctor and patient. The purpose of disseminating health information is to influence personal health choices by improving health literacy.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

[General Communication](#)

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Writing a Study Protocol

Last modified at 11/30/2015 4:21 PM by Arnold Bosman

Once you have decided that you want to perform a specific study after having analysed a particular public health problem, you will usually take the next step by writing a study protocol.

The Study Protocol is a document that describes every step of the study and answers relevant questions about topics such as the **public health problem** that will be addressed, the **study questions** that need to be answered, what **objectives** the study will achieve, how much **power** the study will have and what the **impact** of the findings will be **on public health**.

The purpose of writing a study protocol can include the objective to acquire funding or ethical approval for the study, to lay down the rules to all study partners or just to ensure that all study team members are on the same page in terms of expectations and contribution.

A possible useful **outline of your study protocol** could be with the following chapters:

1. [Presentation](#)
2. [Background and justifications](#)
3. [Objectives](#)
4. [Methods](#)
5. [Ethical considerations](#)
6. [Project management](#)
7. [Timetable](#)
8. [Resources](#)
9. [References](#)
10. [Appendices](#)

Presentation

The presentation includes the **Formal Title** of the study, with a **short, accurate and concise summary** of what it is

about. In this section all the **investigators** are presented as well as the **main centres** that collaborate in the study.

The **Steering Committee** (or scientific board), if present, will be mentioned here, as well as a **summary of the protocol**.

Background and justifications

This section includes the specific statement of the public health problem and the justification for this study. This will also describe what specific gaps in existing scientific knowledge this study will fill.

The importance of the subject area has to become evident after reading this part of the document, including the magnitude and frequency of the disease that is under investigation.

The description of the gaps in existing knowledge about this topic require a comprehensive and rigorous literature review, which will often benefit from the support of an experienced librarian or documentalist.

The principal questions to be addressed have to be described here as well as an indication how the expected results will contribute to existing knowledge. Describe how you plan the dissemination of results and more importantly, [how policy makers or decision takers can use the results](#).

Objectives

The objectives should answer the study question and need to be worded "[S.M.A.R.T.](#)". If there are more than 1 objectives, then it may be useful to distinguish between *Principal objectives* (that **must** be achieved and that dictate the design and methods of the study) and *Secondary objectives* (which may be of interest, but not essential).

The objectives then need to be translated to hypotheses in terms that allow statistical testing.

An example:

Non S.M.A.R.T objective: "To identify risk factors for HCV infection"

[SMART](#) objectives:

Principal objective: "To determine if sharing a haemodialysis machine with a HCV infected patient is a risk factor for HCV infection"

Secondary objective: "To identify failures in procedures designed to prevent cross-infection via haemodialysis machines"

Hypothesis: "The incidence of HCV infection in haemodialysis patients is higher in patients sharing machines with HCV infected patients than in patients not sharing machines with HCV infected patients"

Methods

The methods section describes in specific detail the procedures to achieve the study objectives: what will exactly be done and how? The information in this section will be used by decision makers and financing organisations to judge the validity of your proposal.

It includes the description of the [Study Design](#) ([cohort](#), [case control](#), [cross-sectional](#)...) plus a brief justification for the design choice.

Furthermore the Study Population is specified here:

- definitions (population, exposures, outcomes)
- selection
- criteria for inclusion and exclusion
- mechanisms of recruitment
- accessibility, follow-up, representativeness

Sampling design

Description of the [sampling frame](#) (district, household, persons,...), method (random, cluster, stratified,...), randomisation procedures, replacement procedures (in case of refusal) are described here.

Sample size and power calculations need to be specified here, which are based on the principal objective. This section will then also indicate the feasibility of performing a study of this size.

Example definitions:

Exposure is defined as: "consumption of custard slices in June or July 1991"

Case definition: "a person living in South-West Wales with a laboratory confirmed infection due to S. Enteritidis in June and July 1991"

Case finding takes place through: Public Health Laboratory; weekly notifications

Controls definition: "persons living in SW Wales in same neighborhood as cases"

Control finding: random selection of people using telephone directory

Data analysis plan:

This part of the methods is structured in terms of objectives. For each of the hypotheses tested it includes dummy tables (clarifying the comparison of groups, which risk factors or protective measures are compared). If appropriate the assessment of dose-response relationship for key exposures and assessment of possible confounding factors / effect modifiers.

The plan includes a description of the statistical tests used and what type of adjustment is made to the analysis (e.g. [stratification](#), multivariable analysis)

Define the indicators you will need to reach objectives and specifically what data you will need to collect.

Data collection

Describe how data will be collected (interview, observation, record review) and by whom (interviewers? in that case describe selection, training and level of supervision). What will be the tools for data collection ([questionnaires](#), recording materials). In case questionnaires are used: will they be self or interviewer administered, face-to-face or telephone interviews?

To check appropriateness of your data collection instrument use the [practical guide](#).

Describe if data collection will be blinded or not and the procedures for taking samples.

Data handling:

How will coding of the data take place (e.g. anonymisation)? During data collection or afterwards? And by whom?
What will be the procedures for data processing (what software and hardware used, method of data entry). Will data entry be done during the data collection or afterwards? Will it be single or double entry?
What will be the steps for validation and data cleaning?

Pilot studies and pre-testing:

Some investigators say that "No study should be done without a pre-test". Such a test helps to assess the feasibility of sampling, check if data collection plans are realistic. It tests the measurement methods and questionnaires.

Describe how you will perform the test.

Limitations:

A relevant part of the methods section in your study protocol is the identification of potential sources of biases (e.g. selection bias, information bias, misclassification bias, interviewer bias).

How to deal with them? Will you consider possibilities for correcting? How will these limitations affect the results?

Ethical considerations

In this section, all ethical considerations of performing the study are to be discussed. An obvious start is to ensure that the participants have a good understanding of what participating to the study will mean for them personally: will they have to supply confidential information? If yes, how will these data be protected? How long will the data be stored? Will their names and personal details remain attached to the data?

Will participants be expected to be subject to laboratory investigations? If so, will they be informed about their individual results?

All these questions need to be clarified in this section, and explained how participants are informed about these procedures, and how they will be asked to give their "Informed consent".

If required, this information needs to be translated in the local (lay) language.

Most countries require population studies to be approved by an Ethics committee: in that case a section such as this will need to address all their possible questions.

Project management

Use this section to describe in detail the participating institutes and persons, in order to acknowledge in advance their contribution and role. Describe clearly enough the responsibilities and tasks of each partner.

Who will have access to what data? How is data ownership documented and agreed upon. Use also this paragraph to describe in advance the distribution of roles for authorship. The Vancouver criteria will be helpful here (see reference section below)

Timetable

All operational and practical issues can be described here, including planning- and organisation of the study, when milestones will be finished such as questionnaire design, recruitment of participants, purchases of materials. If permissions are required (e.g. to include certain staff in your project team, to spend certain funds or to contact the study population) then here you need to include a timetable with deadlines for each of these permissions. When funding need to be obtained, indicate here the deadlines for having approval for those funds.

Pilot study
time to do adjustments

Final study
data collection
analysis
presentation of results and write up

Resources

Extent of this section depends on target audience
Specify
available sources
requested sources
Keep budget
reasonable
detailed
well justified

References

Limit number of references to key articles
Follow recommended style
Vancouver

www.library.soton.ac.uk/infoskills/vancouver.shtml
www.transfusion.ca/new/bulletin/vancouver-style.html

Appendices

Methodological appendices
List of definitions
Questionnaires
Introductory letters to study participants
Informed consent forms

See also the following EPIET Lectures:

[Writing a Study Protocol](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

[General Communication](#)

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Writing for Publications

Last modified at 7/24/2015 5:04 PM by CeRC

This is a hub for advices and checklists to help readers improving their manuscripts - in self review or peer review mode.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

[General Communication](#)

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Health communication

Last modified at 4/17/2015 8:25 AM by Arnold Bosman

The health communication knowledge is based on several complementary concepts:

[Health literacy](#)

Health literacy can be defined as the capacity that an individual has to access and effectively use health-related information.

[Health education](#)

Health education aims to influence a person's knowledge, attitudes and behaviours connected to health in a positive way.

[Social marketing](#)

Using social marketing tools to conduct public health improvement programs can help to clarify goals and improve success.

[Risk communication](#)

Risk communication is a sustained communication process with a diverse audience about the likely outcomes of health and behavioural attitudes.

[Crisis communication](#)

While risk communication is ongoing, crisis communication is a reactive communication effort in the face of an unforeseen event.

[Health advocacy](#)

Advocacy is one strategy to raise awareness and promote health and access to quality health care at the individual and community levels.

[Outbreak communication](#)

An effective outbreak communication can help to bring an outbreak under control as quickly as possible, with as little social disruption as possible



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

[Infection control and hospital hygiene](#)

Introduction to Public Health and basic concepts

Statistical Concepts

Healthcare-associated infections and risk assessment

Last modified at 4/21/2016 7:28 AM by Vladimir Prikazsky

[Burden of HAIs](#)

[Types of HAIs](#)

[Main pathogens and resistance](#)

Healthcare settings



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Healthcare-associated infection prevention and cont...

Last modified at 4/21/2016 7:36 AM by Vladimir Prikazsky

European organization

Programme components

Advocating and managing infection prevention and control programme

Legal and ethical aspects

Education and training of staff

Research in prevention and control of HAIs



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Healthcare-associated infection prevention and cont...

Last modified at 4/21/2016 8:03 AM by Vladimir Prikazsky

1. Patient safety including clinical and corporate governance
2. Risk management
3. Quality improvement



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Surveillance and investigation of healthcare-associat...

Last modified at 4/21/2016 7:41 AM by Vladimir Prikazsky

Surveillance of HAIs

Detection and outbreak investigation

Multidrug resistant organisms (MDRO)

Reporting



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Infection control interventions

Last modified at 4/21/2016 7:46 AM by Vladimir Prikazsky

1. Developing infection control procedures
2. Isolation
3. Standard precautions
4. Transmission based precautions
5. Cleaning, decontamination, disinfection and sterilisation
6. Environment safety
7. Staff health
8. Use of laboratory data in infection control



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Antimicrobial stewardship

Last modified at 10/12/2016 2:27 PM by Vladimir Prikazsky

1. Introduction.

Increasing rates of resistance to antimicrobials among hospital pathogens is a worldwide problem that has been recognized for more than 20 years. The spread within the last 10 years of Extended Spectrum Beta Lactamase producing Enterobacteriaceae (ESBL-PE), and recently Carbapenemase producing Enterobacteriaceae (CPE) and the role of antimicrobial consumption (ref) in the increasing of this phenomenon, underline the emergency of implementation of antimicrobial stewardship program before and during outbreak episodes concomitantly with infection control measures. Patterns of consumption of different classes of antibiotics classes have been closely correlated with the emergence of bacteria resistant to those classes [1,2]. Consequently, many publications have suggested the importance of antimicrobial stewardship to avoid [3] and to control [4,5] the emergence of antibiotic resistance.

The main purpose of antimicrobial stewardship programs is to improve how antibiotics are used, in order to optimize clinical outcomes while minimizing unintended consequences of antimicrobial use, including toxicity, the selection of pathogenic organisms (such as *Clostridium difficile*), and the emergence of resistance [6]. Since at least a third of antibiotic use is usually inappropriate or unnecessary in most hospital settings, antibiotic stewardship programs lead to reductions in total antibiotic consumption.

2. Antimicrobial stewardship programs: definitions and objectives.

Several terms have been used to describe antimicrobial stewardship programs (ASPs) such as antibiotic policies, antibiotic management programs or antibiotic control policies. They all refer to an effort by the healthcare institution (as a whole) to optimize antimicrobial use among hospitalized patients in order to improve patient outcomes and to reduce antimicrobial resistance.

The main objectives of ASPs could be summarised as:

- optimizing antimicrobial use for treatment and prophylaxis of infections among hospitalized patients in order to improve clinical outcomes, ensure cost-effective therapy and reduce adverse effects associated with antimicrobial use [7];
- prevention and control of antimicrobial resistance by reducing the use of antibiotics;
- avoiding the occurrence of difficult-to-treat infections and reducing the incidence of multidrug-resistant microorganisms.

An advisable step before development of any program is to first attempt to define the most important issues that exist with respect to antimicrobial use within a given healthcare institution. Once institution-specific problems have been identified, it is important to evaluate potential causes and solutions. As part of this, any existing antibiotic recommendations and policies should be reviewed.

3. Antimicrobials stewardship components.

- Structure: the structure of institutional antimicrobial stewardship programs has been defined in different guidelines and publications [5,8]. It should be multidisciplinary, typically with a core team consisting of an infectious disease physician or clinical microbiologist, and a clinical pharmacist with training infectious diseases. It is important to obtain the support from the hospital administration. All of these individuals should be full-time employees of the institution in which the stewardship program resides. The administration should give core team members the authority to enforce stewardship tactics [9]. Close collaboration with the microbiology department, an information system specialist, an infection control practitioner and hospital epidemiologist is also recommended. The role of staff nurses can be important, but is currently less defined: nurses are often antibiotic first responders, central communicators, coordinators of care, as well as 24-hour monitors of patient status, safety, and response to antibiotic therapy[10].
- Components: many different strategies have been employed in ASP interventions. These are often introduced simultaneously, as multifaceted interventions, and no single type of intervention appears to be much more effective than others. The ASP should foster appropriate antimicrobial use and include monitoring of resistance, in collaboration with an effective infective control program. Here we detail examples of commonly used interventions.
- Audit and feedback: this is one of the two core ASP strategies recommended by the Infectious Diseases Society of America (IDSA). It has been shown to reduce the inappropriate use of antimicrobials [11]. In a one-step prospective method, targeted antibiotics are directly audited by an ID physician or clinical microbiologist during clinical rounds with immediate feedback provided to the responsible team. In a two-step review method, all cases are initially reviewed by a pharmacist or nurse member of the ASP team, and then selected cases meeting criteria for further review are discussed with an ID physician or clinical microbiologist, who will then provide recommendations for changing or discontinuing antibiotics. Since acceptance of recommendations is voluntary, the clinical teams responsible for patient care do not perceive a loss of prescribing autonomy.
- Formulary restriction or preapproval for specific antibiotics: restriction strategies that target one or several classes or antibiotic have been shown to contribute to the control of outbreaks of many specific resistant bacteria. Preapproval requires the prescriber to indicate the appropriate rationale for the selection of a particular agent, either electronically or on paper, before use of specific antibiotics is permitted. Clinicians may inappropriately circumvent this type of restriction by listing an unconfirmed diagnosis or a differential diagnosis that meets the required criteria for use [12].
- Guidelines with or without feedback: many before/after studies [13] conducted in hospitals suggest that improvements in appropriateness and reductions in antibiotic consumption can occur in response to the implementation of local and regional antibiotic guidelines for specific infections.
- Many other strategies are employed in ASPs, including education of prescribers, implementation of clinical

pathways and use of computerised clinical decision support systems. Education efforts in isolation appear to be marginally effective and tend not to have a sustained effect without repetition. Computer assisted surveillance, and clinical decision support systems have shown promising improvements in antibiotic prescribing resulting in more appropriate dosing and fewer adverse drug events [14].

3.1. In case of an outbreak, when should we implement an ASP?

Many publications suggest that ASPs can help to control the spread of resistant microorganisms. The most convincing evidence of an effect on antimicrobial resistance rates was provided by studies aimed at reducing the incidence of *Clostridium difficile* associated disease [15]. However, use of specific antibiotic classes seems to be correlated with a higher incidence of certain microorganisms, suggesting that close monitoring of their specific consumption could help to contain outbreaks [2,16].

4. Evaluating antimicrobial stewardship programs.

From an infection control point of view, the most relevant goal to be assessed may be the ecological effects of the ASP. In this context, the main objectives of the ASP may be to reduce antimicrobial collateral damage, such as *Clostridium difficile* associated diarrhea, or avoiding multi drug resistance microorganisms.

A major consideration when measuring resistance is choosing which types of specimens to include. Four main options are possible:

- surveillance cultures that detect colonization and are the only surveillance cultures that will identify asymptomatic carriers;
- cultures taken during routine care of the patient;
- microbiologically and clinically documented infections;
- site specific cultures

Measurements of antibiotic use are an essential component of ASPs, and provide data for assessing the impact of ASP interventions. The most commonly used metric for measuring aggregated antibiotic use is the defined daily dose (DDD) proposed by the World Health Organization, expressed as DDD per 1000 patient-days. This measure allows comparisons between institutions. However, it underestimates the real consumption in some populations, such as children and patients with renal failure. Others methods that can be used include days of therapy for each antibiotic administered, and total length of therapy.

Studies [17] have shown that early reassessment of antibiotic therapy after 24-48 hours is an important step towards appropriate use of antibiotics. This may be focused on appropriateness of antibiotics used according to local clinical guidelines or available microbiological results. Rates of early switching to oral antibiotic therapies can also be used to evaluate ASPs.

All of these data may be collected on a hospital-wide basis as part of regular point prevalence studies within the ASP.

References:

1. Neuhauser MM, Weinstein RA, Rydman R, Danziger LH, Karam G, Quinn JP. Antibiotic resistance among gram-

- negative bacilli in US intensive care units: implications for fluoroquinolone use. JAMA. 2003;289:885-8.
2. Lepper PM, Grusa E, Reichl H, Högel J, Trautmann M. Consumption of imipenem correlates with beta-lactam resistance in *Pseudomonas aeruginosa*. Antimicrob Agents Chemother. 2002;46:2920-5.
 3. Paterson DL. The role of antimicrobial management programs in optimizing antibiotic prescribing within hospitals. Clin Infect Dis. 2006;42 Suppl2:S90-5.
 4. Anderson DJ, Kaye KS. Controlling antimicrobial resistance in the hospital. Infect Dis Clin North Am. 2009;23: 847-64.
 5. Gould IM. Antibiotic policies to control hospital-acquired infection. J Antimicrob Chemother. 2008;61:763-5.
 6. Dellit TH, Owens RC, McGowan JE Jr, Gerding DN, Weinstein RA, Burke JP, et al. Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America guidelines for developing an institutional program to enhance antimicrobial stewardship. Clin Infect Dis 2007; 44:159-177.
 7. MacDougall C, Polk RE. Antimicrobial stewardship programs in healthcare systems. Clin Microbiol Rev 2005; 18:638-656.
 8. Lesprit P, Brun-Buisson C. Hospital antibiotic stewardship. Curr Opin Infect Dis. 2008;21:344-9.
 9. Drew RH. Antimicrobial stewardship programs: how to start and steer a successful program. J Manag Care Pharm. 2009;15:S18-23.
 10. Olans RN, Olans RD, DeMaria A Jr. The Critical Role of the Staff Nurse in Antimicrobial Stewardship- Unrecognized, but Already There. Clin Infect Dis. 2016; 62:84-9.
 11. Chung GW, Wu JE, Yeo CL, Chan D, Hsu LY. Antimicrobial stewardship: a review of prospective audit and feedback systems and an objective evaluation of outcomes. Virulence. 2013;4:151-7.
 12. Reed EE, Stevenson KB, West JE, Bauer KA, Goff DA. Impact of formulary restriction with prior authorization by an antimicrobial stewardship program. Virulence. 2013;4:158-62.
 13. Zahar JR, Rioux C, Girou E, Hulin A, Sauve C, Bernier-Combes A et al. Inappropriate prescribing of aminoglycosides: risk factors and impact of an antibiotic control team. J Antimicrob Chemother. 2006; 58 :651-6.
 14. Evans RS, Pestotnik SL, Classen DC, Clemmer TP, Weaver LK, Orme JF Jr, et al. A computer-assisted management program for antibiotics and other anti-infective agents. N Engl J Med. 1998;338 :232-8.
 15. Valiquette L, Cossette B, Garant MP, Diab H, Pépin J. Impact of a reduction in the use of high-risk antibiotics on the course of an epidemic of *Clostridium difficile*-associated disease caused by the hypervirulent NAP1/027 strain. Clin Infect Dis. 2007;45:S112-21.
 16. Kaier K, Frank U, Hagist C, Conrad A, Meyer E. The impact of antimicrobial drug consumption and alcohol-based hand rub use on the emergence and spread of extended-spectrum beta-lactamase-producing strains: a time-series analysis. J Antimicrob Chemother. 2009;63:609-14.
 17. Lesprit P, Landelle C, Brun-Buisson C. Clinical impact of unsolicited post-prescription antibiotic review in surgical and medical wards: a randomized controlled trial. Clin Microbiol Infect. 2013;19: E91-7.

Original contribution from Jean Ralph Zahar (*Unité d'Hygiène Hospitalière, CHU Angers*)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

[Introduction to Public Health and basic concepts](#)

Statistical Concepts

Public Health Microbiology

Last modified at 7/26/2014 5:11 PM by Arnold Bosman

'Public health microbiology' is 'a crosscutting area that spans the fields of human, animal, food, water and environmental microbiology, with a focus on human health and disease. It requires laboratory scientists with the ability to work effectively across disciplines, particularly epidemiology and clinical medicine'.

Public health microbiology aims to interpret diagnostics at the population level, rather than at the level of the individual patient.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

[Statistical Concepts](#)

Statistical Concepts

Last modified at 4/12/2016 7:25 AM by Vladimir Prikazsky

In this part we describe concepts such as confidence interval, sampling, sample size and study power. The chapter is incomplete and deserves additional content. You are very welcome to add content, so that this part of the FEMWIKI contains key methods and concepts in statistics that epidemiologists should understand in order to interpret results of data analysis.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Health Economics

Last modified at 7/26/2014 8:10 PM by Arnold Bosman

This article is a result of new content structure of the FEM Wiki.

You are invited to contribute suitable content such as definitions, scope, examples and other related material. Feel free to link to other external resources.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

[Introduction to Public Health and basic concepts](#)

Statistical Concepts

Public Health Informatics

Last modified at 7/26/2014 5:08 PM by Arnold Bosman

The domain of *medical informatics* has evolved fast since the early 90ies. Until that moment there had been many developments in information systems to support the infrastructure of medicine. Gradually the awareness emerged that medical doctors and other health care professionals (including managers) needed support in education, decision making, communication and other professional activities. As the focus shifted to information management of the health care professionals, the discipline of 'medical informatics' took further shape.

"Medical informatics is the field that concerns itself with the cognitive, information processing, and communication tasks of medical practice, education, and research, including the information science and the technology to support these tasks."

The field is very much interdisciplinary, with branches of high applied activities and also involved in fundamental research. Medical informatics currently is a distinct academic entity in most countries, with a strong network between EU countries in particular. And the focus is on *medical practice*.

The domain of **public health informatics** seems not yet so clearly defined in the EU. Most of the time, IT in public health (e.g. supporting new surveillance systems), follow [standards](#) defined within medical informatics (e.g. standards such as LOINC, SNOMED, HL7).

Where medical informatics comes from a predominantly patient oriented focus, it follows that choices in IT architecture, standards, protocols etc have been made from that perspective (e.g. high requirements for confidentiality, data protection, and accuracy of information of individual diagnostics and diagnosis).

Key requirements from the public health perspective focus on populations rather than individuals and may include timely data access (which can conflict with decisions for data shielding in medical informatics systems) and representativeness for (sub)populations (e.g. high focus on getting continuous and unbiased samples of information on subpopulations; accuracy of information on individual diagnosis would be much less important). These key

requirements could (and probably should) lead to appropriate choices for IT architecture that may be different from those in medical informatics.

Therefore there is a need for building and maintaining a critical mass of public health information experts, that are well versed in core activities of public health (e.g. [surveillance](#), [outbreak investigations](#), [field epidemiology](#), public health microbiology, screening) and who are able to use that knowledge in developing IT infrastructure that serves the needs of public health professionals and that is well integrated in health care systems.

"Public health informatics is the systematic application of information and computer science and technology to public health practice, research, and learning."

The USCDC has developed a 2 year fellowship in public health informatics, with the goal to provide training and experience in applying computer and information science and technology to real public health problems. In the EU, the public health informatics is not yet coordinated at the Community level in either development of professional standards or training.

References

1. Greenes RA, Shortliffe EH. Medical informatics. An emerging academic discipline and institutional priority. JAMA. 1990 Feb 23;263(8):1114-20.
2. [William A Yasnoff, J Marc Overhage, Betsy L Humphreys, Martin LaVenture. A National Agenda for Public Health Informatics: Summarized Recommendations from the 2001 AMIA Spring Congress. JAMIA 2001;8:535-545 doi:10.1136/jamia.2001.0080535](#)
3. [William A. Yasnoff, Patrick W. O'Carroll, Denise Koo, Robert W. Linkins, and Edwin M. Kilbourne. Public Health Informatics: Improving and Transforming Public Health in the Information Age. J Public Health Management Practice, 2000, 6\(6\), 67-75](#)
4. [Public Health Informatics Fellowship Programme PHIFP, Centers for Disease Control, Atlanta, Georgia, USA](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Attack rates and case fatality

Last modified at 5/28/2014 12:12 AM by Arnold Bosman

Attack rates and case fatality rates are an example of how the epidemiological jargon may be confusing.

Attack rates

1. Attack rates are actually **risks** (or "incidence proportion" or "cumulative incidence")
2. Attack rates are often expressed as a percentage.

An attack rate is not an **incidence rate**. It is actually a risk (also called incidence proportions), and the time contribution of each individual is not included in the denominator.

The denominator consists of the number of people present at the beginning of the outbreak, disregarding those who will leave, develop illness, or die. This means that the cases (numerator) are also included in the denominator: it is therefore a true proportion.

In outbreaks of short duration, attack rate is a term frequently used instead of risk or incidence proportion. In a foodborne outbreak, we will often refer to "food specific attack rates". In this circumstance, the denominator will consist of the number of people who ate a specific food, while the numerator will consist of the number of people who ate that food and became ill.

Case fatality, rates and ratios: all the same?

No!

These are distinct different concepts, though in many epidemiological manuscripts (and even text books) you will find that case fatality, case fatality rate and case fatality ratio are used as synonyms. However, they are not.

Case fatality

Case fatality is the concept used to express the proportion of cases of a certain disease that actually dies due to the

consequences of that disease. Since it is a proportion, it is usually expressed as %, or per 1000. The case fatality can be seen as a cumulative incidence. It is relevant to keep in mind that the death has to be due to the consequences of the disease, since otherwise each disease would have a case fatality of 100% (since all people die eventually). It is a true proportion, since the denominator includes all cases, even those who died (the numerator).

Example of case fatality: around 1850, the case fatality of cholera (for which then there was no effective treatment) was up to 40%. This means that out of each 100 cases of cholera, 40 would eventually die due to the disease, usually within 2 weeks after onset. In comparison, the case fatality of tuberculosis in those times was almost 100% within the first 2 years after diagnosis, since there was no cure for tuberculosis either.

Case fatality rate (CFR)

The CFR is a case fatality expressed over time. It is therefore a true rate, since time is included in the denominator. It can be expressed as number of deaths among cases per 100 or 1000 person-years. Depending on the disease, it may also be expressed per person-weeks or person months.

As a rate, it reflects the dynamic of the fatality over time, among cases.

To stay with the same example as above, around 1850, most cases of cholera had either recovered after 2 weeks, or had died. Once recovered from the disease, a person is no longer a case. That means that the person time of that person may no longer contribute to the denominator. If we assume that of the 100 cases of cholera, 40 die due to the disease after 2 weeks and the rest (60) recover from the disease after the same amount of time, then the CFR for cholera is in that situation 40 per 200 person-weeks (=1 per 5 person-weeks = 4 per 5 person-months = 10 per 1 person-year).

Likewise, of the 100 newly diagnosed tuberculosis patients, 50 would die in the first year and 50 would die in the second year. That comes down to a case fatality of 1 per 2 person years for the first year after diagnosis.

Here we can clearly see the major difference between case fatality and CFR: tuberculosis is clearly the 'greater killer' compared to cholera (because the case fatality is 100%, and of cholera 'only' 40%), however M.tuberculosis kills its victims much slower than Vibrio cholerae does.

Case fatality ratio

This is simply the comparison of two case fatalities, expressed as a ratio. So the cholera:tuberculosis case fatality ratio is 40:100 (or 4:10). Usually we put the greater killer first, so the TB:cholera case fatality ratio is 2.5:1. In this sense, it is a comparison between 2 populations, similar as we do with odd ratio, risk ratio, sex ratio etc.

The Case Fatality Ratio could also be used to assess the impact of an intervention. For examples, if untreated cholera has a case fatality of 40% and when treatment is given in time, the fatality could be below 1%. This leads to a Case Fatality Ratio of 40 or more when comparing untreated and treated groups of cholera patients.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Incidence rate

Last modified at 9/14/2011 2:06 PM by CeRC

Incidence rate (synonym: incidence density rate) is expressed as number of events per person-time

In a population that we may study over a predefined period of time, not every member of the population will be at risk of developing the disease for the same amount of time during the study period. Some individuals will develop disease soon and no longer be at risk of disease, some will die, some will be lost to follow-up; some will enter the population half way through the study (birth, immigration), etc. The time contributed by each person is sometimes called “time at risk” (of an event occurring). As a consequence the population contributing to time in the follow up is also called “population at risk

Denominator

In order to measure the incidence rate of a disease in a population we first need a denominator. The denominator is a measure of the time spent by each individual in the population at risk of developing illness during the study period. We then need to sum up all of the time at risk for each individual person to obtain a time denominator. The time in the denominator includes every instant during which an individual is at risk of developing the disease [1]. All time units in the denominator are equivalent regardless of whether they reflect the time contribution of the same person or of different persons. This way 10 people that have been observed for exactly one year will contribute the same amount of time than 20 persons that have been observed for 6 months. This is why the time at risk is frequently called person-time (e.g. person-years, person-months).

Numerator

[1]. Thus being part of a population at risk is a dynamic process.

The incidence rate measures the occurrence of disease onsets in a population per unit of time of follow-up. Because of its similarity to population density, in an area, over time, it is sometimes called “incidence density rate”.

$$\text{Incidence rate} = \frac{N}{\text{Time}} = \frac{\text{Number of individuals developing disease}}{\text{Total time experienced by population at risk}}$$

The figure illustrates the computation of time contribution of 10 persons of a hypothetical population to the denominator of an incidence rate.

Figure. Graphical example of occurrence of disease according to time at risk of developing disease in a hypothetical population of 10 people (D*, Disease onset).

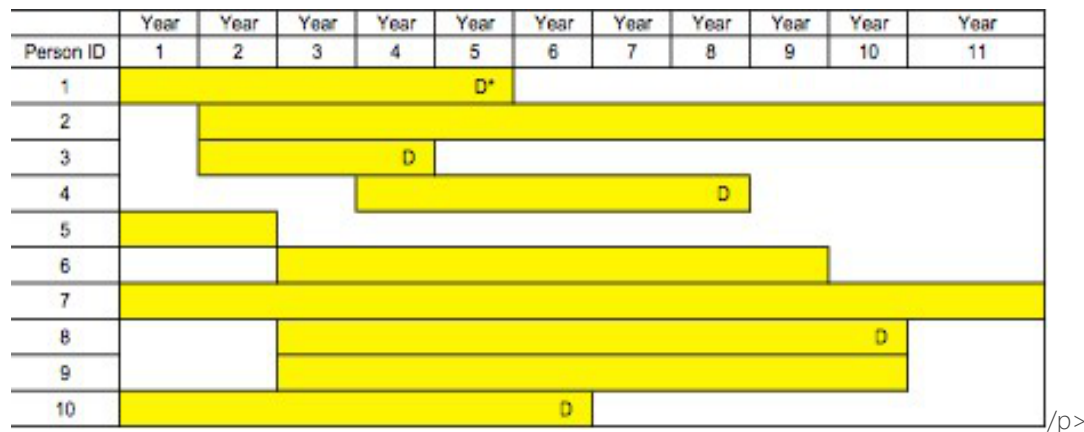


Table. Summary table of number of years at risk and disease onset in the same hypothetical population of 10 people.

Person ID	Number of years at risk	Disease onset
1	5	1
2	10	0
3	3	1
4	5	1
5	2	0
6	7	0
7	11	0
8	8	1
9	8	0
10	6	1
Total	65	5

$$\text{Incidence rate} = \frac{\text{Number of individuals developing disease}}{\text{Total time experienced by population at risk}} = \frac{5 \text{ disease onsets}}{65 \text{ person years}}$$

Five disease onsets occurred during a total follow up time of 65 years. This is equivalent to an incidence rate of 7.7 onsets per 100 years of follow up of individuals being at risk of developing disease.

Incidence rate: $(5/65 \text{ persons years}) = 0.07 \times 100 = 7.7 \text{ per 100 persons-years}$

Alternatively the incidence rate can be written as follows: $7.7 \times 100 \text{ years}^{-1}$

It is very common to multiply the rate per units of 100, 1,000 or 10,000 in order to make comparisons among studies and interpretation easier.

An incidence rate will range from 0 to infinity according to the unit of time used to express the person-time incidence. Among 100 people no more than 100 deaths can occur. But those deaths can occur in 1000 person-years (if on average all 100 die after 10 years), 100 person-years (if on average all 100 die after 1 year) or even 1 person-year (if each of the 100 persons dies on average after 3.65 days). There is therefore no upper limit to an incidence rate. The numerical value of an incidence rate is not by itself interpretable because it depends upon the unit of time chosen. This unit should be chosen in order to make sense. For example 14 deaths per 10 person-year means that a certain number of people (at least 14) were followed for periods of times (quite short) with the total of which equals 10 years. This rate is better expressed in months or days.

Incidence rate	= 14	deaths per 10 person-years
	= 14	deaths per (10 x 12) 120 person-months
	= 12.3	deaths per 100 person-months
	= 14	deaths per (10 x 365) 3650 person-days
	= 38,3	deaths per 10000 person-days

In an incidence rate the only units involved are time units which appear in the denominator.

Whereas risk (incidence proportion) can be interpreted as a probability, the incidence rate cannot.

References

1. Rothman KJ; Epidemiology: an introduction. Oxford University Press 2002, p.28-33.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Prevalence

Last modified at 10/31/2010 4:28 PM by CeRC

Prevalence (synonym: prevalence proportion) is expressed as a percentage.

The prevalence measures the proportion of individuals in a population with a specific disease (or a specific characteristic) at a certain point of time (T_0). The term prevalence pool is sometimes used to identify the subgroup of the population with the disease. Some individuals exit the prevalence pool by recovering or dying, some enter by developing disease. However prevalence is measured at a specific point in time (T_0).

$$\text{Prevalence} = \frac{\text{Number of persons with disease at } T_0}{\text{Population present at } T_0}$$

The prevalence will range from 0 to 1 or 0% to 100% if expressed as a percentage. Since time is not measured in the denominator the prevalence is not a [rate](#).

The prevalence reflects both disease incidence and disease duration. The higher the incidence and the longer the duration of a disease, the larger the amount of people with the disease at a specific point in time. If incidence rates and duration are stable over time, the prevalence and [incidence rate](#) are related in the following way [\[1\]](#):

$$\frac{P}{1 - P} = I * \bar{D}$$

in which P is the prevalence, I the incidence rate and D the average disease duration.

If the prevalence is low, $1 - P$ tends to 1 and $P / (1 - P)$ is almost equal to P.

Then the formula can be simplified as:

$$P = I * \bar{D}$$

The prevalence proportion is used to measure disease burden in a population. It applies more to administrative areas of public health than into the cause of disease. It is also used to describe characteristics or conditions *other than diseases* (vaccine coverage, prevalence of smokers, prevalence of blood groups, etc.).

References

1. Rothman KJ; Epidemiology: an introduction. Oxford University Press 2002, p.40-44.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

The Components of Surveillance

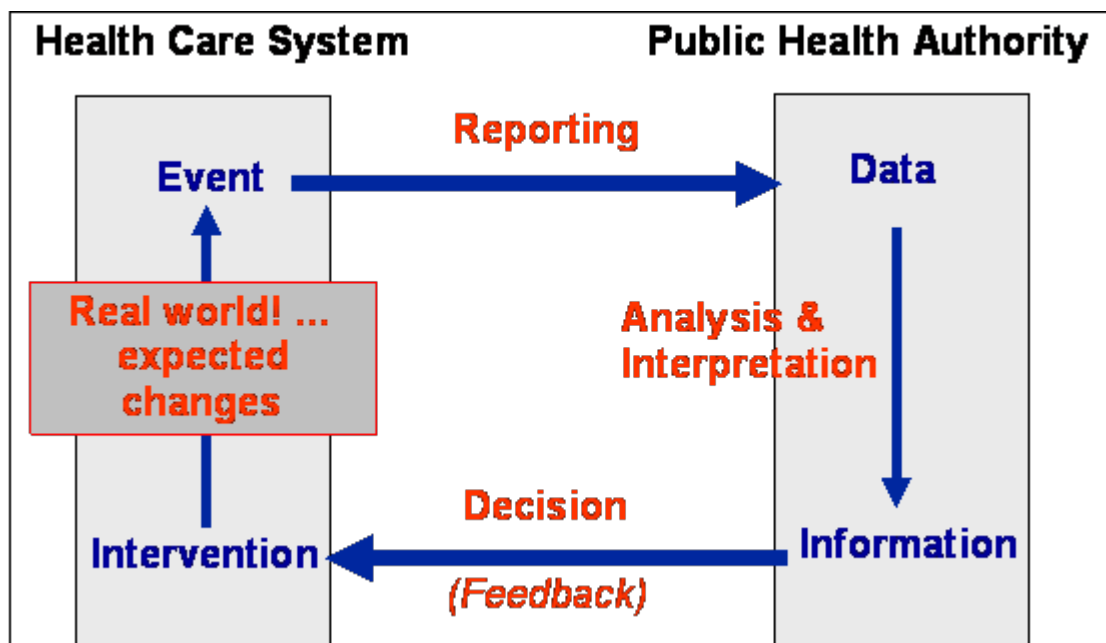
Last modified at 4/17/2015 10:27 PM by Arnold Bosman

In 1965, the Director General of the World Health Organization (WHO) established the epidemiological surveillance unit in WHO's Division of Communicable Diseases. In 1968, the 21st World Health Assembly affirmed the three main features of surveillance: a) the systematic collection of pertinent data, b) the orderly consolidation and evaluation of these data, and c) the prompt dissemination of results to those who need to know-particularly those in position to take action [1]. In addition, "epidemiologic surveillance" was said to imply "...the responsibility of following up to see that effective action has been taken." This addition emphasises the cyclical nature of the surveillance process, as outlined in figure 1.

The classical model of surveillance thus includes three major processes:

- Capture and collation of data
- Analysis and interpretation of data (to generate information)
- Dissemination of information

This is often shown as a cyclical process, with a fourth process of public health response (intervention), which may result in changes that will then be evaluated by the collection, analysis and interpretation of data. This surveillance cycle is depicted in figure 1.



References

1. National and global surveillance of communicable disease. Report of the technical discussion at the Twenty-First World Health Assembly. A21/Technical Discussions/5. Geneva, Switzerland: World Health Organization, May 1968.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Types of Study

Last modified at 7/19/2018 11:08 AM by Jakob Schumacher

Epidemiology is about studying the occurrence of health events and their determinants in populations and there are different ways to classify the types of epidemiological studies. One way is to differentiate between observational studies and experimental studies. In experimental studies, the researchers control almost every aspect of the study, including who is exposed and who is not exposed. When studying infectious diseases in human populations, it is unacceptable to deliberately decide on exposure of humans, so most often observational studies are used to study the exposure that has happened accidentally in real life (which is sometimes referred to as 'the experiment of nature'). Among the observational studies we can distinguish *descriptive studies* and *analytical studies*.

Descriptive studies aim to *quantify* and *qualify* public health problems (**what** goes on **where**, **when**, among **whom**) while analytical studies aim to *explain* the mechanisms in which public health problems emerge, propagate and sustain themselves in populations (asking '**how**' and '**why**'). Both classes of studies require a careful design, in order to ensure that the results accurately and realistically reflect the situation in the population. The challenge in any of those study designs is to minimize bias (which leads to a misrepresentation of the real situation) and to define the objects of measurement (disease, health even or determinants such as behaviour) in a precise enough manner.

Other Classification

Studies can also be classified into descriptive, exploratory, inferential, predictive, causal and mechanistic studies. Descriptive studies aim to describe a dataset. Explorative analysis aims to find relationships between several variables. Inferential analysis refers to using a small sample of data to infer something on a bigger population. Predictive analysis uses data on some objects to predict values for another object. Causal studies aim to identify causes preferably via randomized controlled trials. Mechanistic analysis aims to understand the exact changes in variables that lead to changes in other variables.

Importance of descriptive studies

Descriptive epidemiological studies are an important source of evidence for setting priorities in public health.

Authorities that have the ability to effectively and efficiently describe the health status of the population, will also be able to set priorities for example according to the magnitude, impact or burden of disease. This may also include economic consequences of disease.

Similarly it will be relevant to public health authorities to know how determinants of health are developing in 'their' population; do people still decide to vaccinate? Is the use of recreational drugs increasing? Does our needle exchange programme for drug users still cover the needs? By measuring and monitoring determinants, authorities can be informed about risks of health events within the population even before they occur.

Importance of analytical studies

As it is not enough to know what is going on, but also to understand how and why, analytical studies are a powerful tool for generating evidence for policies for disease prevention and control. As most authorities aim to interfere with the daily life of citizens as little as possible and only when it is really necessary, interventions designed to prevent and control diseases are ideally very specific and highly effective. This usually requires a solid knowledge of the relationship between health events and their determinants. Analytical studies aim to contribute to that body of knowledge.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Event-based Surveillance

Last modified at 4/26/2015 6:48 PM by Arnold Bosman

This article is a result of new content structure of the FEM Wiki.

You are invited to contribute suitable content such as definitions, scope, examples and other related material. Feel free to link to other external resources.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Rapid Risk Assessment

Last modified at 4/17/2015 7:17 PM by Arnold Bosman

Once an incident has been verified as being of potential public health concern, a rapid risk assessment is undertaken (usually within 24 to 48 hours) to evaluate the risk to human health. The risk to a population from a communicable disease is dependent on the likelihood of transmission in the population (probability) and the severity of disease (impact). The probability of an incident developing, and the impact if it does, are based on both the nature of the infectious agent and details of the incident. This may be further influenced by context or the broad environment in which the incident occurs, including political, public, media interest, perception of threat, and the acceptance of risk, which may vary between countries and cultures.

Rapid risk assessment is a core part of public health response and thus widely undertaken by public health professionals. Formal systems which are used to grade evidence and recommendations, such as the systematic methods used in evidence-based medicine, rely on published research evidence, and studies are graded according to design and susceptibility to bias. However, as time and evidence are limited, a rapid risk assessment may need to rely at least in part on specialist expert knowledge and these formal systems are not directly applicable.

There are only a limited number of examples of a more systematic and transparent approach to rapid risk assessment in the literature including:

- a qualitative method for assessing the risk from emerging infections in the UK (Morgan et al. 2009) using algorithms to consider the probability of an infection occurring in the UK population, its potential impact, and identifying gaps in knowledge or data;
- a prioritisation approach to rank emerging zoonoses posing the greatest threat in the Netherlands, based on seven criteria (including probability of introduction, likelihood of transmission, economic damage, morbidity and mortality) to aid decision-making (<http://www.rivm.nl/bibliotheek/rapporten/330214002.html>);
- a dynamic risk assessment model developed in the UK to assess the risk from an outbreak or incident, consisting of five attributes (severity, spread, confidence in the diagnosis, ease of intervention and the wider context in which events are occurring) rated over a 0 to 4 scale. During an outbreak, the dynamic risk assessment of each event occurring is used to inform management action at that

However, the same principles of transparency, explicitness, and reproducibility also apply to a rapid risk assessment. For the rapid risk assessment of most infectious disease threats, observational data is often the only available and obtainable source of information. Expert knowledge is also important if there is lack of time and evidence. In such cases it is important to 'unpack' and make explicit the expert knowledge and distinguish between knowledge based on good research, and experience and opinion-based knowledge. Serious attempts should be made to assess the quality of the evidence, based on the source, design and quality of each study or piece of information.

Uncertainties should be identified, clearly documented and communicated and the assessment updated in light of new evidence over time.

A rapid risk assessment includes the approach and tools required at each stage of the process: stage 0 is the preparation stage; stage 1 is the collection of event information; stage 2 is the literature search and systematic collection of information about the aetiological agent; stage 3 focuses on the extraction of evidence; stage 4 conducts an appraisal of the evidence; and stage 5 estimates the risk. Transparency and sharing of information is essential at every stage. This document incorporates a step-by-step guide through each stage with examples and checklists of the resources and evidence required.

Advance preparation and planning saves time and is vital to ensure that potential threats are identified, assessed, and managed effectively. Ideally the following should be in place: evidence-based protocols and guidance for responding to incidents, protocols for identifying sources of key information for rapid risk assessment, strategies for literature searches, and lists of relevant contacts including named experts.

Rapid risk assessments of potential communicable disease threats can be complex and challenging as they must be produced within a short time period when information is often limited and circumstances can evolve rapidly. The rapid risk assessment methodology described in this document enables the structured identification of key information using systematic appraisal of the best scientific evidence and/or specialist expert knowledge available at the time in order to provide a clear estimate of the scale of the health risk. This is important in not only communicating the potential magnitude of the risk in a systematic and transparent way, but allows documentation of evidence and gaps in knowledge at the time when the assessment is made.

References:

This text is the entire executive summary of a technical document published by ECDC:

[Operational guidance on rapid risk assessment methodology. ECDC Technical Document. August 2011.](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Assessing the burden of disease and risk assessment

Last modified at 3/22/2017 8:26 PM by Vladimir Prikazsky

Effective disease prevention and control depends on several factors that all have to be present and work together in the community. It all starts with the ability to **detect threats** to the health of the population. A threat can be seen as an undesirable situation that has *not yet* occurred, but that *may* happen unless protective measures are taken. The ability to detect a disease threat implies that we already have basic knowledge about the 'normal occurrence (or **burden**)' of this disease in the population.....

Assessing the burden of disease requires a public health workforce with the competence to collect, analyse and interpret health data from "your" population plus the infrastructure in the health care system that allows access to relevant data. Methods used in [Field Epidemiology](#) play a central part in assessing the burden of disease.

To detect health threats requires (in addition to the above) continuous monitoring of burden of disease information of 'your own' and surrounding populations, trends in risk behaviour, characteristics of pathogens (e.g. development of antimicrobial resistance) plus competent staff responsible for continuous collection, analysis and interpretation of information. The process aimed at detection of health threats is sometimes referred to as [epidemic intelligence](#).

Once health threats have been detected and validated, information needs to be shared with "those who need to know" in the health system. This usually requires translation of specific epidemiology jargon in a format that can be used by policy and decision makers in order to decide on [interventions](#) (preventive and control measures).

This part of this FEMWiki addresses methods that can be used to assess the health status of the population and detection and assessment of health threats. Methods for [Surveillance](#), Risk Assessment and [Outbreak Investigations](#) will be described in this section.

[Interventions](#) (public health measures, policy making and decision taking) is a topic described in another part of the FEMWIKI. [Communication](#) is yet another topic.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

[Introduction to Public Health and basic concepts](#)

Statistical Concepts

Host

Last modified at 10/11/2012 5:06 PM by ecdc

A person or other living animal, including birds and arthropods, that offers subsistence or lodging to an infectious agent under natural conditions. A transport host is a carrier in which the organism remains alive but does not undergo development.

A primary host is where a parasite reaches maturity or passes its sexual stage. A secondary host is where a parasite is in a larval or asexual stage.

References:

1. David L. Heymann (editor). Control of Communicable Diseases Manual. APHA, 2008



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

[Introduction to Public Health and basic concepts](#)

Statistical Concepts

Reservoir for infectious agents

Last modified at 5/17/2014 11:33 AM by Arnold Bosman

The **Reservoir for Infectious Agents** is the principal habitat where a specific infectious agent lives and multiplies. The reservoir is necessary for the infectious agent either to survive, or to multiply in sufficient amount to be transmitted to a susceptible host. Examples may include **primates** (including human beings), the reservoir of pathogens such as hepatitis A virus, hepatitis B virus, Polio virus (all 3 types), Bordetella pertussis, *Corynebacterium diphtheria*, etc.

Other micro organisms have larger animal reservoirs, e.g. Salmonella species can be found in almost every animal. The **environment** contains a large number of reservoirs: **soil**, the reservoir for *Clostridium tetani* or **water**, the reservoir for *Legionella pneumophila*.

In a number of articles the concept of '[source](#)' and 'reservoir' are used as synonyms, though strictly speaking they are not. In this FEMWIKI we consider a source as the starting point of a transmission route; it usually can be found at a specific time in a specific place (in other words: it often has 'an address'). Sources can be part of a reservoir. For example: warm water systems (generic) are known to be reservoirs for legionella and the shower in room 911 of Hotel X was found the source of a number of legionella infections.

It is important to know the reservoir of pathogens, as this may offer [opportunities for control](#). For example, a disease like smallpox (variola major) could be eradicated from this planet, in part because humans were the main reservoir. By immunizing the majority of the reservoir population, and by rigorously keeping infectious patients isolated and immunizing contacts, the smallpox virus could no longer survive in nature. This is one of public health's great achievements and currently similar attempts are underway to do the same with poliovirus.

References:

1. David L. Heymann (editor). Control of Communicable Diseases Manual. APHA, 2008



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

[Introduction to Public Health and basic concepts](#)

Statistical Concepts

Source of infection

Last modified at 5/22/2014 3:04 PM by Arnold Bosman

We refer to the *source* of infection as the **origin** from which a [host](#) acquires the infection, either *endogenous* (i.e. originating from a person's own commensal microbial flora) or *exogenous* (i.e. an individual, animal or object that in the external environment of the host). Usually the *source* can be identified as an individual, animal or object in a specific place, and at a specific time.

Thus, a **person can be a source of infection**; either for him/her self (endogenous) or to other people (directly through personal contact, or indirectly, e.g. by contaminating food or beverages).

In addition to people, **also animals can be sources** of infection

Objects may be sources of infection; food, water, air-conditioning systems, showers, medical instruments, recreational waters, door knobs, cotton handkerchiefs etc. Most man-made products that may be sources of infection are required to be produced while limiting the risk of contamination.

In most [outbreak investigations](#), the principal objective is to identify the source of the infection. Interestingly enough this sometimes leads to semantic problems: an identified 'source' (e.g. a chocolate cake) is usually contaminated by some other source (e.g. the baker of the cake, or the eggs used). Tracing back such a '[chain of transmission](#)' usually leads back to the reservoir. In a number of articles the concept of 'source' and '[reservoir](#)' are used as synonyms, though strictly speaking they are not.

Inanimate sources of infection are sometimes referred to as '*vehicle of infection*' (e.g. the chocolate cake) or '*fomites*' (e.g. the cotton handkerchief). Inanimate sources (vehicles, fomites) are part of the [indirect transmission route](#)..

Source of infection should be distinguished from **source of contamination** (e.g. overflow of a septic tank, contaminating a water supply).

[References:](#)

1. David L. Heymann (editor). Control of Communicable Diseases Manual. APHA, 2008



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Vector Borne

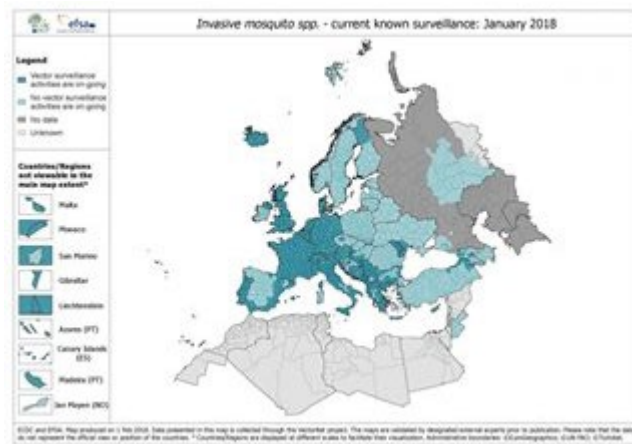
Last modified at 11/11/2018 2:39 PM by Arnold Bosman

What are vectors and vectorborne diseases?

Vector-borne diseases are infections transmitted by the bite of infected arthropod species, such as mosquitoes, ticks, triatomine bugs, sandflies, and blackflies (1). Arthropod vectors are cold-blooded (ectothermic) and thus especially sensitive to climatic factors. Weather influences survival and reproduction rates of vectors (2), in turn influencing habitat suitability, distribution and abundance; intensity and temporal pattern of vector activity (particularly biting rates) throughout the year; and rates of development, survival and reproduction of pathogens within vectors. However, ECDC states that climate is only one of many factors influencing vector distribution, such as habitat destruction, land use, pesticide application, and host density. Vector-borne diseases are widespread in Europe and are the best studied diseases associated with climate change (3).

Disease vectors in Europe

Together with the network of vectorborne disease experts, ECDC monitors occurrence of disease vectors in the European Union. Geographical surveillance maps for [mosquitoes](#), [ticks](#), and [phlebotomine sandflies](#) are continuously updated and published online.



Example of ECDC vector surveillance maps for mosquitoes, January 2018.

Studying the epidemiology of infectious diseases that are transmitted to humans via vectors, we need to know more about the relationship between the [reservoir](#) of the infectious agent and the vector that is needed to transmit the disease. The fact that a vector is required between the reservoir and the [host](#) will make the spread of the disease within a population more complex to study and predict.

The basis of vector borne disease epidemiology is the triangle between pathogen, vector and [hosts](#). As with other type of infectious diseases, the pathogens (virus, parasites, bacteria) cause disease, yet they depend on the vector to be transmitted to the hosts. The natural (or primary) host of a vector-borne disease is part of the reservoir that maintains the pathogen in natural cycles of infection and [transmission](#) by vectors to other susceptible natural hosts. For example, in West Nile Virus, the primary transmission cycle takes place among various bird species and a number of mosquito species. In most birds in Europe, Africa and Asia, fatal outcome is rare when infected with West Nile Virus, in contrast to birds in the Americas (especially the family of crows). In this particular example of West Nile Virus, humans and horses are incidental (dead - end) hosts; this means that they do not contribute to the further spread of the disease.

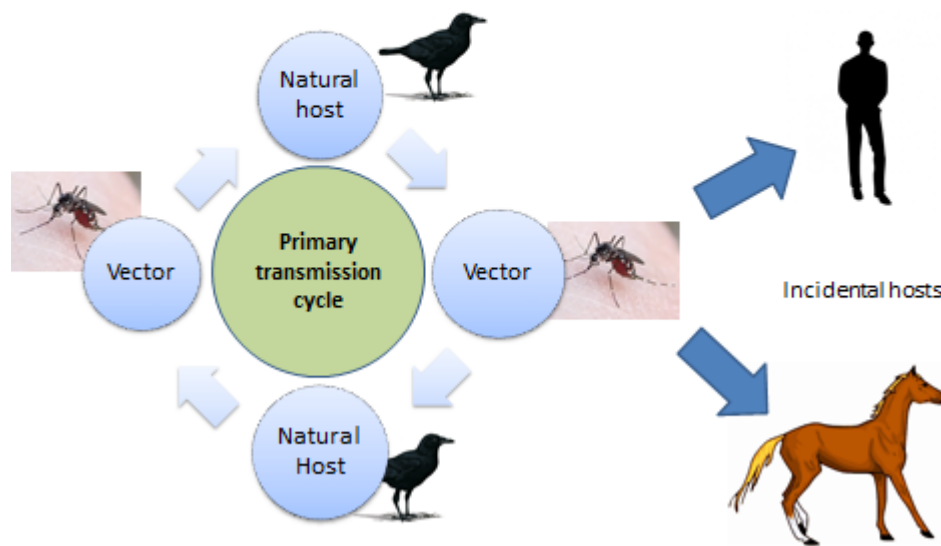


Fig 1. Example of vector borne disease transmission: the West Nile Virus cycle

Modes of transmission of vector borne diseases

Vertical [transmission](#) from vector (e.g. mosquito) to progeny may occur via transovarial passage of the infectious agent.

Horizontal transmission occurs when infected mosquitoes transfer the agent to vertebrate hosts. This can be mechanical (e.g. when the agent is transferred by the vector via the mouth parts, without multiplication in the vector) or biological (where the agent multiplies in the vector).

Surveillance of vector borne diseases

Surveillance can target the vector:

- measuring abundance and spread
- testing vectors for infection (if tests exist for such investigation)
- calculate vector infection rates

Surveillance can target hosts:

- Sentinel animals, via periodic bleeding and testing for infection
- Via notification of animal or human infections / disease.

References

1. Confalonieri U, Menne B, Akhtar R, Ebi KL, Hauengue M, Kovats RS, Revich B, Woodward A. Human Health. In: Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Parry ML, Canziani OF, Palutikof JP, van der Linden PJ, Hansson CE (eds). Cambridge University Press, Cambridge, U.K., 2007: 391-431
2. Rogers DJ, Randolph SE. Climate change and vector-borne diseases. *Adv Parasitol.* 2006;62:345-81.
3. Semenza JC, Menne B. Climate Change and Infectious Diseases in Europe. *Lancet ID.* 2009;9:365-75.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Primary prevention

Last modified at 11/4/2014 11:24 PM by Vladimir Prikazsky

Primary prevention takes place before a person develops the health outcome (usually disease) and in fact aims to prevent the health outcome. Success of primary prevention can be shown through indicators such as incidence and prevalence of disease going down. Primary prevention usually targets the [reservoir](#) or [source](#) of infection, or aims to block the [transmission](#) route. Yet it could also strengthen the [immune system](#), such as through immunisation.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Secondary prevention

Last modified at 7/2/2013 10:03 AM by Vladimir Prikazsky

This article is a result of new content structure of the FEM Wiki.

You are invited to contribute suitable content such as definitions, scope, examples and other related material. Feel free to link to other external resources.

Activities that may be considered as secondary prevention:

- Contact tracing - for early treatment
- Screening - Early diagnosis - for prevention of spread, prevention of disease
- Early treatment - for prevention of sequelae
- Postexposure vaccination - prevention of disease onset.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Advantages and disadvantages of cohort and case c...

Last modified at 4/17/2015 7:14 PM by Arnold Bosman

Many text books have described advantages and disadvantages of cohort and case control studies. The following table summarises useful comments.

	Cohort studies	Case control studies
Suited for rare diseases	No	Yes since starting with a set of cases
Suited for rare exposures	Yes since starting with exposure status	No
Allows for studying several exposures	Difficult but examples exists (Framingham study)	Yes
Allows for studying several outcomes	Yes	No
Disease status easy to ascertain	Sometimes difficult	Easier since starting point of the study
Exposure status easier to ascertain	Yes since starting point of the study. Except for retrospective cohorts	Sometimes difficult. Information biases.
Allows computation of risk and rates	Yes	No
Allows computation of	Computation of risk ratio	Estimation of risk ratio, rate ratio

effect	and rate ratio	from odds ratio
Allows studying natural history of disease	Yes Easier to show that cause precedes effect.	More difficult Temporality between cause and effect difficult to establish
Based on existing data sources	Difficult	Yes but access to information sometimes difficult
Easiness to find a reference group	Usually not difficult to identify an unexposed population	No Major potential biases when selecting a control group
Sample size	Large	Small
Cost	Elevated except if retrospective cohorts	Smaller
Time required	Long, sometimes very long except if retrospective cohorts	Shorter
Follow up	Difficult, loss to follow up	No follow up
Logistics	Heavy Many staff, large data sets Long duration	Easier
Concept	Easy to understand	Difficult to understand particularly if case cohort or density case control study
Ethical issues	Major if studying risk factors. Interruption of study if exposure shown to be harmful. Need for intermediate analysis.	None since outcome already happened.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Tertiary prevention

Last modified at 5/25/2015 5:28 PM by Arnold Bosman

This article is a result of new content structure of the FEM Wiki.

You are invited to contribute suitable content such as definitions, scope, examples and other related material. Feel free to link to other external resources.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

[Introduction to Public Health and basic concepts](#)

Statistical Concepts

Prevention

Last modified at 4/7/2014 12:01 AM by Arnold Bosman

Next to [interventions](#), prevention is the other major tool of public health. Prevention strategies are usually based on scientific evidence, such as provided by epidemiology. We distinguish [primary](#), [secondary](#) and [tertiary](#) prevention.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

The Role of Surveillance

Last modified at 3/11/2015 10:52 PM by Arnold Bosman

Surveillance is often described as providing *information for action*. As human beings we can only function effectively by constantly monitoring what is happening around us, using our senses to provide information that will enable us to respond (take action) to external threats and opportunities, or to plan future actions. Although the amount of information we have to process varies according to different situations, in most situations we need to be able to monitor what is happening in our environment on a continuous basis, so that we can act in as immediate a way as possible. There are analogies to this need for continuous monitoring of events or circumstances in order to act effectively in most areas of human activity, such as the need for banks to monitor financial transactions and trends in the economy, and the need for public health bodies to monitor trends in hazards, exposures and health events in order to protect the health of populations and individuals.

While the above analogy may be relatively simplistic, it emphasises the fact that a reliable supply of timely, accurate and relevant information is essential to almost everything we do. The control and prevention of infectious diseases and other environmental threats to health is no different in this respect. The processes by which much of the information that is used to inform public health action is collected, analysed and disseminated to those who need to know are collectively known as surveillance. Without good surveillance, public health action is unlikely to be effective or efficient.

The general definition of surveillance, as given in the Oxford English Dictionary is:

n. *Close observation, especially of a suspected spy or criminal*

ORIGIN C19: from Fr., from sur- 'over' + veiller 'watch'

While this definition makes explicit the observational aspect of surveillance, for many members of the general public the link with crime and espionage has negative associations, while for public health practitioners it does not provide any sense of purpose. Fortunately, however, there are many definitions of surveillance to be found in epidemiological texts, one of the earliest was drawn up by Alex Langmuir, the first chief epidemiologist of the Communicable Disease

Center, now the Centers for Disease Control and Prevention in the United States:

"Continued watchfulness over the distribution and trends of incidence through the systematic collection, consolidation and evaluation of morbidity and mortality reports and other relevant data together with the timely and regular dissemination to those who need to know"

This definition emphasises many of the **key attributes** of surveillance: the need for surveillance processes to be continuous and systematic; the fact that surveillance may, and often needs to, draw on data from many different sources; and the need for regular and timely outputs.

Last, in his Dictionary of Epidemiology, also provides an indication of purpose in his definition of the surveillance of disease:

"The continuing scrutiny of all aspects of occurrence and spread of disease that are pertinent to effective control" [1]

A Historical Perspective

Statistics on morbidity, and more particularly mortality, have been produced in many societies over many hundreds of years, such as in the case of the 'Bills of Mortality' that were produced and published in London. The London Bills of Mortality were devised in the early sixteenth century in London. The information was collected by the Parish Clerk's Company of London, and published weekly. Monthly and Yearly digests were also issued. Over time the detail provided by the bills increased. Initially they contained only burials, but by the 1570s the total number of baptisms was also returned. In 1629 cause of death information was given and in the early eighteenth century the returns began supplying a distribution of the ages at which Londoners died. Although John Graunt published a statistical analysis of the Bills as *Natural and political observations made upon the bills of mortality* in 1662, their production and publication was not directly linked to public health action, and as such they are not generally cited as examples of surveillance (although it is known that Londoners bought copies of the bills and scanned them for signs of impending disease, particularly plague, and may have made their own plans to leave the city at the first sign of rising numbers).

The first exponent of the systematic analysis of official medical statistics for the purposes of monitoring disease and identifying associations between disease and demographic groups is generally held to be William Farr. Farr was the first compiler of scientific abstracts in the **General Register Office** in London, where he set up a system for routinely recording the causes of death in England and Wales. He collected, analysed and interpreted vital statistics, and plotted the rise and fall of epidemics of infectious disease. He published his results in weekly, quarterly and annual reports. For example, for the first time it allowed the mortality rates of different occupations to be compared. In addition, in 1864 Farr was the first to publish work containing material calculated and printed by a machine, Scheutze's Difference Engine, which was a forerunner of the computer.

References

-
1. Last JM. A Dictionary of Epidemiology (Second Edition). Oxford University Press 1988, New York.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Criteria for Surveillance

Last modified at 3/11/2015 10:50 PM by Mike Catchpole

Surveillance should be directly relevant to control needs, and as such the criteria for deciding whether to undertake surveillance of a particular hazard, exposure or event are closely aligned to those for deciding whether to undertake control measures in respect of that hazard, exposure or event. The major criteria for determining the case for surveillance are therefore:

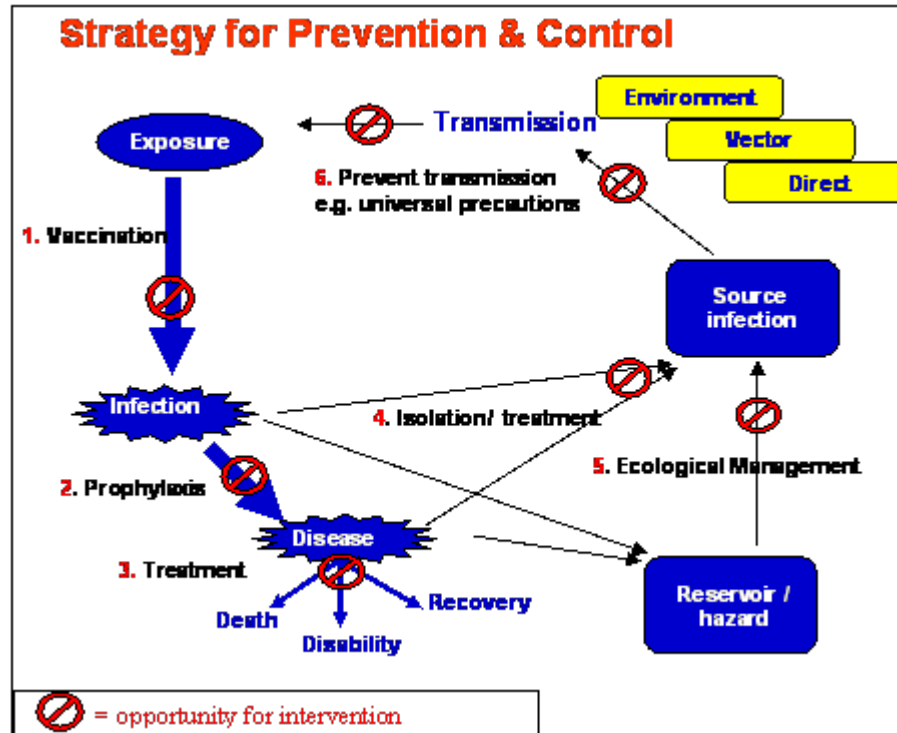
- The public health importance of the proposed subject of surveillance
- The strategy for control or prevention of the hazard, exposure or event
- The information required to inform the interventions that make up the control strategy
- The feasibility and/or cost-benefit of collecting the data required

There are additional considerations in respect of decisions about what type or form of surveillance might be undertaken, and these will be discussed later.

The public health importance of a problem, and the priority for control and prevention, is determined by a variety of factors, the most important of which is usually the overall burden on the population. Other factors, such as the potential for epidemic spread, political or media interest, the availability and cost-benefit ratio of interventions, and national or international disease control targets may also be influential in determining the priority given to the control and surveillance of a public health problem. An example of how priorities can change, and as a result investment in surveillance can be influenced, can be seen in respect of sexually transmitted infections in the UK in the second half of the 20th century. Investment in the control, and surveillance, of sexually transmitted infections in the UK declined during the 1950's and 1960's, as it was perceived that the post war peak in gonorrhoea and syphilis had been successfully controlled. Despite rises in the incidence of these diseases in the 1960's and 1970's it was not until the advent of AIDS in the 1980's that significant investment in improved surveillance, as part of a major new programme of control and prevention, took place [1]. These improved surveillance systems provided much of the information, and identified cases for research, that enabled targeted and effective control measures to be implemented before HIV had

been identified. More recently, the rise in antimicrobial resistance has resulted in significant investment in many countries in surveillance systems for monitoring MRSA [2], antibiotic resistant gonorrhoea [3] and other important resistant bacterial infections.

The decision to undertake surveillance, and more particularly the decision as to what type of surveillance should be undertaken, should take into account the strategy for prevention and control of the disease, hazard or exposure that is to be monitored, and also the types of intervention that are to make up the control programme. Potential points for intervention to control or prevent infectious disease are outlined in figure 1.



The development of a new intervention may introduce the requirement for new surveillance systems to inform and monitor the application of that intervention. Thus the introduction of pneumococcal vaccination in the UK, initially among high risk groups and subsequently as part of the routine childhood immunisation programme, has prompted the development of enhanced surveillance systems for pneumococcal infection, including monitoring of serotypes causing invasive illness and monitoring of vaccine uptake among high risk groups, which has informed policy (in respect of choice of vaccine) and enabled programme delivery to be assessed [4].

New opportunities for the development of surveillance may also arise as a result of changes in the feasibility of collecting data that would inform control and prevention activities. These opportunities might arise because of the introduction of a new health service that is able to provide information that would previously not have been available or prohibitively costly to obtain, or because of the introduction of new technology that significantly reduces the costs,

and hence the cost-benefit, of data collection for surveillance purposes. An example of the former are the surveillance systems that have been developed to make use of data recorded by telephone health advisory services, in the UK [5] and the USA [6]. Examples of surveillance that has been established because of opportunities provided by new technology include primary care surveillance systems recently introduced in the UK [7].

References

1. Catchpole MA. The Role of Epidemiology and Surveillance Systems in the Control of Sexually Transmitted Diseases. *Genitourinary Medicine* 1996; 72: 321
2. Griffiths C, et al. Trends in MRSA in England and Wales: analysis of morbidity and mortality data for 1993-2002. *Health Stat Q* 2004;21:15-22
3. Paine T, Herring A, Turner A, et al. GRASP: a new national sentinel surveillance initiative for monitoring gonococcal antimicrobial resistance in England and Wales. *Sex Transm Infect* 2001; 77:398-401
4. Health Protection Agency. Surveillance of the impact of the pneumococcal conjugate (PCV) vaccination programme for children in England and Wales and protocol for investigation of vaccine failures. (July 2006).
http://www.hpa.org.uk/infections/topics_az/pneumococcal/PneumococcalGuidanceSurveillance.htm
5. Cooper DL, Smith GE, Baker M, Chinemana F, Verlander N, Gerard E, Hollyoak V, Griffiths R. A national symptom surveillance system in the UK using calls to a telephone health advice service. *MMWR* 2004: 53(Suppl);179-183.
6. Steven F. Magruder, J. Henry, M. Snyder. Linked Analysis for Definition of Nurse Advice Line Syndrome Groups, and Comparison to Encounters. *MMWR* 2005: 54(Suppl);93-97
7. Health Protection Agency. Primary Care Surveillance - QResearch (July 2006)
http://www.hpa.org.uk/infections/topics_az/primary_care_surveillance/QResearch.htm



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Sources and Types of Data

Last modified at 9/14/2011 1:57 PM by CeRC

The most common sources of data for surveillance of health events are health service providers, ranging from all health service providers in the case of national statutorily notifiable diseases, to generalist health services (e.g. primary care services), specialised health services (e.g. sexually transmitted infection clinics, maternity services), or diagnostic service providers (e.g. microbiology laboratories). Some sources will provide a broader population perspective than others (e.g. microbiology laboratory reporting of hepatitis B infection is likely to provide information about cases acquired through all modes of transmission, whereas reporting from sexually transmitted infection clinics is likely to exclude a significant proportion of cases that were not acquired through sexual contact). If it is important to be able to produce outputs that are expressed as population-based rates, which allow comparison between populations, particularly at the local population level, then it is necessary to use sources that draw from a well defined population base (e.g. general practitioners in the UK serve defined populations for which age-sex registers exist and can be used for calculating rates).

Other potential sources of surveillance data on disease or health status include disease registries, occupational health records, community services, emergency services, and screening programmes (including not only programmes such as antenatal screening for HIV, syphilis or rubella, and screening of occupational groups such as military recruits, but also screening of blood donations). 'Over the counter' sales of medicines and rehydration solutions have more recently been shown to be potentially useful sources of data for surveillance, providing an early indication of community outbreaks [1]. The increasing uptake of the Internet also raises the possibility of basing surveillance on direct electronic reporting by members of the public.

Surveillance may also focus on exposures and hazards in the environment, in which case potential data sources will include veterinary services, environmental health services, water company quality testing records, and air quality monitoring records.

Expression of surveillance outputs in the form of a population rate or in the form of rates within population subgroups

is required if the burden of morbidity and mortality between different populations and between different groups within a population is to be compared. Defining the population from which cases are drawn by surveillance systems is not always easy. While denominators are readily obtained for surveillance based on reporting by primary care physicians that serve a well defined, and documented, resident population, the same is not the case for surveillance based on cases drawn from emergency departments in hospitals serving cities with large transient populations, including workers who commute in on a daily basis and tourists. Defining the denominator for laboratory-based surveillance systems can also be a problem, particularly for specialist or reference microbiology services, where the population served by the referring clinical services may not be known or may change over time (as the pattern of referral of specimens may change according to contractual or other factors). Appropriate denominator data may also be required where the numerator may be influenced by the opportunity for case identification. For example, in the UK, where hospitals are compared with respect to rates of MRSA infection, the denominator takes into account the average number of patients in the hospital over the time period, and denominator data are also collected on the total number of blood cultures taken during the observation period.

References

1. Debjani Das, K. Metzger, R. Heffernan, S. Balter, D. Weiss, F. Mostashari. Monitoring Over-The-Counter Medication Sales for Early Detection of Disease Outbreaks --- New York City. MMWR 2005; 54(Suppl):41-46



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Surveillance System Design and Operation

Last modified at 9/14/2011 1:58 PM by CeRC

Surveillance system design must take into account not only the objectives and outputs that the system has to deliver, but also the need to be able to operate the system in a consistent and affordable way over a prolonged period of time.

Surveillance systems must be capable of capturing data in a consistent and sustainable way, since fluctuations in reporting may mask or mimic genuine changes in the incidence of the disease, exposure or hazard that is the subject of surveillance. While the attainment of a measure of true disease occurrence within a population through surveillance is probably unrealistic for most diseases and surveillance systems, achievement and maintenance of high levels of ascertainment is still desirable. It is particularly important for the early detection of outbreaks and for surveillance of uncommon diseases of public health importance and/or diseases that are the subject of elimination programmes. It also provides a level of reassurance that the surveillance system is providing a relatively representative picture of the distribution of disease within the target population.

Completeness and consistency of reporting to surveillance systems can be affected by a range of system design and operational factors. Statutory or mandatory reporting has been used as a mechanism for achieving high reporting rates, although there is evidence to suggest that this does not guarantee complete, or even high levels of, reporting. Surveys in the UK have demonstrated very low levels of reporting of some statutorily notifiable infections [1], with little evidence that this can be improved by increasing the payment made to physicians for reporting cases [2]. It is also the case that concerns have been raised that making some diseases notifiable may result in patients being reluctant to seek medical attention, because of fears about being notified, which was a significant factor in the decision not to make HIV or AIDS notifiable in the UK. Another strategy that may be used to achieve high reporting rates is to adopt a sample-based or sentinel approach to surveillance, so that resources can be invested in achieving higher reporting rates from a smaller number of reporters e.g. by providing training and support, such as in the form of specialised software for reporting. For rare diseases it is also possible to consider surveillance based on active reporting, where reporters are actively prompted or reminded to report cases on a regular basis. A good example of this latter

approach is the British Paediatric Surveillance Unit, which sends out reporting cards every month to all paediatricians in the UK, requesting that they report cases of a list of 10-12 conditions that are the subject of active surveillance [3].

Although these approaches can achieve higher rates of reporting, they are not applicable to all diseases or circumstances, and they do not always achieve the desired result. As a general rule, surveillance systems based on capture of data from health services are most likely to be sustainable and achieve acceptable levels of coverage of the target population where the design of the system is coherent with the infrastructure of the healthcare systems within which patients are seen. Where specialist clinics provide the majority of care for a particular disease or group of diseases, such as is the case with sexually transmitted infections in the UK, surveillance based on data reporting from those clinics can often achieve higher quality (with regards to diagnostic validity of reports and compliance with reporting) information for a given cost than would be possible through systems based on universal reporting. National publicly funded health services often have an established culture of central reporting, and as such surveillance based on voluntary (or mandatory) reporting to a national surveillance centre may be more acceptable than in countries with devolved or largely privately funded healthcare. To some extent the increasing use of information technology within healthcare services, and the development of electronic patient records, may overcome some of the barriers to reporting, in that the effort required of clinical staff should become smaller, although this will require the development of standards for electronic data exchange, and clarification of any data protection issues.

The quality and completeness of reporting to surveillance systems is also likely to be better where reporting makes use of data collected for clinical or other operational purposes is captured as a by-product of routine clinical or administrative processes. Once again, the development of electronic patient information systems is likely to make this easier. Other opportunities for capturing data without requiring new or additional effort by clinical staff include making use of laboratory requesting or result reporting data, use of forms completed for the purposes of claiming for payment from health insurance companies, and use of pharmacy records (e.g. for dispensing of vaccines or disease-specific medications).

References

1. Cartwright KA. Meningococcal meningitis. *British Journal of Hospital Medicine* 1987;38(6):516, 521-4
2. McCormick A. The notification of infectious diseases in England and Wales. *Communicable Disease Report. CDR Review* 1993; 3(2):R19-25
3. British Paediatric Surveillance Unit. 19th Annual Report 2004-05. Royal College of Paediatrics and Child Health 2005. London. (http://www.bpsu.inopsu.com/publications/annual_reports/annual-report_2005.pdf)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Quality, Governance and Operating Procedures

Last modified at 9/14/2011 2:02 PM by CeRC

Having determined the sources from which data are to be collected, there are a range of issues that need to be decided upon regarding the data that is to be collected and the principles of operation of the system. These issues may be classified under the following headings:

- Data quality
- Data governance
- Operating procedures

The concept of setting quality standards for data, and measuring data against those standards, is well established for readily measurable aspects of data quality, particularly the validity, timeliness and completeness of data items. Simple quality standards can be set, and measured against, such as maximum acceptable levels of missing items in particular data fields and mean, median or maximum acceptable time between event detection and report to the surveillance system.

Quality criteria for data and information, and the systems that are used to process and deliver those data and information, are frequently expressed in terms of the following dimensions:

Dimension	Type of measure	Quality target
Completeness	Quantitative	Information should be sufficiently complete to be fit for purpose
Timeliness	Quantitative	Information should be available when it is needed
Accuracy	Quantitative	Information should be sufficiently free from error to be fit for purpose

Relevance	Qualitative	Information should be contextually appropriate
Reliability	Qualitative	Provenance, objectivity, believability
Delivery	Qualitative	Information should be formatted to satisfy users' needs

Harder to measure but perhaps more important aspects in the context of overall information quality, such as accuracy and relevance, have sometimes been neglected in the setting of standards.

One approach to addressing the question of how to ensure that case reports are relevant and reliable is to develop case definitions for reporting of surveillance data. This is more likely to be feasible if the case definitions are developed as part of the overall process of developing a new surveillance system, when there is the opportunity to assess the acceptability and resource requirements for the reporting mechanisms, training of staff, and supporting investigations (e.g. to ensure that all suspected cases have appropriate laboratory investigations undertaken) required to ensure compliance with the adoption of the case definitions. As with any case definitions, those used for surveillance need to be defined so that they achieve the desired level of sensitivity and specificity in terms of case ascertainment. Case definitions used for surveillance will often need to be more sensitive (and less specific) for case ascertainment than those used in analytical epidemiological studies, since the purpose of surveillance is frequently to provide an early warning of possible emergence of disease outbreaks or rising trends, which can then be assessed through further epidemiological study.

Retrospective implementation of case definitions, particularly where those definitions are based on specific laboratory investigations or the collection of specific exposure or risk factor data, can pose significant problems in terms of the cost and acceptability to reporters and surveillance system operators. It may, however, be possible to categorise data collected through pre-existing surveillance systems against case definitions that have been developed at a later stage, even if it is not possible in the short to medium term to adapt the systems to report according to a particular level of case definition.

Data collected should be relevant and sufficient to meeting surveillance objectives, and should be restricted to only items that are required to meet the objectives of surveillance. Additional, non-essential data items, which are often collected on a 'nice to know' basis rather than because they are justified in terms of meeting explicit objectives, place additional burdens on data providers and on the supporting information systems and, if the data set being reported is person identifying, may breach data protection restrictions.

Anonymised data should be used where there is no need to be able to identify individuals or there is no other reliable method of achieving record linkage between different data sets. In the case of infectious disease surveillance, it is often necessary to collect person identifying information in order to be able to contact cases rapidly in order to undertake follow up and contact tracing and/or outbreak investigations. When person identifying data are used they should be kept secure and disclosed only on a strict 'need to know' basis, and in accordance with Data Protection laws.

Surveillance systems used to capture, analyse, and disseminate information should be operated to agreed standards. The development and adoption of standard operating protocols and case definitions provides a mechanism for ensuring that surveillance systems operate in a consistent and explicit manner over time and place. The scope of operating protocols should include a statement of purpose, case definitions or definitions of hazards and exposures, laboratory investigation protocols (where appropriate), sources of data (e.g. the type of clinical service from which the data are to be captured, and the 'sampling' approach - universal, random or convenience sample, sentinel), the data items to be collected (including level of person identifier required), the outputs, and the roles and responsibilities of those involved in the surveillance process and the custodian or owner of the system. Publication of these protocols helps to make the purpose and governance arrangements for surveillance systems explicit to data providers, data subjects and the recipients of surveillance outputs.

Many of these suggested components of an operating protocol have been covered earlier in this chapter. One component that requires mention is that of guidance on what and when to report. To some extent this guidance can be provided through the publication of case definitions for reporting, but such definitions do not exist for many surveillance systems, particularly those that cover a wide range of infections (such as laboratory reporting schemes that capture data on all organisms identified by reporting laboratories). Even when case definitions do exist, guidance may be required as to whether cases should only be reported when they meet certain criteria (e.g. those for a confirmed case) or when all exposure and risk factor data are available, or whether preliminary reporting should be made on the basis of suspected case identification and/or when only partial exposure or risk factor data are available (in which case, clear mechanisms need to be defined for how more detailed information should be reported at a later date). Protocols should also cover the issue of how frequently data should be reported, and through what mechanism, and, where electronic reporting systems are used, what form and structure the data should be reported in.

Surveillance systems should also be subject to regular audit against their objectives and periodic evaluation.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Analysis, Interpretation and Dissemination

Last modified at 9/14/2011 2:01 PM by CeRC

The information provided by surveillance systems is typically in the form of descriptive statistics on the frequency and distribution of cases, and temporal trends in these. The information on the distribution of cases may include categorisation by geography, demographic characteristics, occupational and other risk factors. Surveillance systems, particularly for infectious disease and other environmental threats to health, may collect, analyse and disseminate information on hazards (e.g. sources of environmental contamination), exposures (e.g. occupational exposures to blood borne viruses), control or prevention measures (e.g. vaccinations given), as well as information on cases of disease or injury. The objective of providing such information is to enable the recipients to [1]:

- Detect and enumerate hazards, exposures and cases in need of public health intervention
- Detect outbreaks (i.e., identifying an increase in frequency of disease above the background occurrence of the disease)
- Determine the distribution and spread of illness
- Portray the natural history of a health condition
- Estimate the impact of a hazard, exposure, disease or injury
- Evaluate prevention and control measures
- Facilitate planning
- Generate hypotheses that will stimulate or can be addressed through research

In order to inform action that will result in improved control or prevention, surveillance must provide information that is not only timely and accurate, but also that is interpreted and presented in a format and through channels that are appropriate to those who have responsibility for taking action. The range of target audiences is large, but can be categorised broadly as follows:

- Public
- Clinicians/Microbiologists/Control of Infection staff
- Public Health professionals
- Environmental Health professionals

- Health service managers
- Health educators/teachers
- Government/politicians

Of these groups, the public are often overlooked as a potential target audience for surveillance outputs, and yet many public health actions that have the greatest potential for improved control and prevention, such as improved food hygiene, safer sexual practice, and uptake of vaccinations require significant public action.

The analysis of surveillance data can range from producing simple tabulations of descriptive statistics by time, place and person, to sophisticated time trend analyses and analyses within geographical information systems.

Although many significant outbreaks are first detected and reported by clinical staff or members of the public, before they are identifiable through surveillance systems (largely because of the delays inherent in surveillance systems), surveillance remains an important mechanism for detecting outbreaks, particularly of uncommon infections and diffuse outbreaks (i.e. outbreaks occurring over wide geographical areas, with relatively small numbers in any one locality).

The need to be able to detect emerging epidemics or outbreaks at an early stage in their evolution is an important element of communicable disease control. As a result a number of analytical techniques have been developed that can be applied to surveillance data to detect possible outbreaks or to assess the statistical significance of an apparent increase in reports. The development of typing schemes, such as serotyping, phage-typing and newer molecular techniques, means that we can now undertake surveillance of several distinct subtypes of an organism where previously it was only possible to detect and monitor the organism at the species level. For many organisms this means that surveillance is now focused on smaller numbers of many different subtypes, which are generally indistinguishable clinically. This provides new opportunities for surveillance, through analysis of the data on different subtypes, to detect outbreaks that might not be detected through the more traditional route of alerting by clinicians. This has been the case for salmonella surveillance for several years now, where the ability to undertake surveillance of many different serotypes and phage-types of salmonella has greatly increased the ability to detect outbreaks that might otherwise have not been detected until considerably later or not at all (see box).

National increase in *Salmonella* Montevideo infections, England and Wales: March to July 2006

Between March and June 2006, the Health Protection Agency (HPA) Centre for Infections (CfI) received 56 *Salmonella* Montevideo isolates from cases of infection in England and Wales. During the same time period in 2005, the CfI received 14 isolates of *Salmonella* Montevideo. Of the cases identified in March - June 2006, 49 were primary cases, of which 37 shared the pulsed field gel electrophoresis (PFGE) profile *S*mvdX07. Cases were distributed widely across the country.

The HPA CfI attempted to contact all cases and detailed food histories were obtained from 15 cases, all of which were confirmed to have the *S*mvdX07 profile. Thirteen (87%) of the cases interviewed reported eating products from one particular manufacturer. The clustering in time of this particular subtype indicated that the cases were part of an outbreak. Two *S.* Montevideo strains isolated from samples

taken immediately before the onset of illness amongst the first human cases from factories of the manufacturer whose products had been eaten by 13 of the 15 interviewed cases were also confirmed as PFGE profile *SmvdX07*. No other common brands, retail outlets, catering chains or single food types were identified as common factors

The frequency of cases of S Montevideo PFGE *SmvdX07* decreased following the voluntary recall of a number of chocolate products, produced by the implicated manufacturer. These were considered as potentially contaminated with S Montevideo PFGE *SmvdX07* after a risk assessment of the results of microbiological sampling and environmental investigations at a number of factory premises.

After carefully considering all the available evidence the Outbreak Control Team concluded that consumption of products made by the manufacturer was the most credible explanation for the outbreak of S Montevideo.

Simple graphs can be used to show trends over time, and to compare those trends between different geographic, demographic or exposure groups. The calculation of rates, based on appropriate denominators, and the graphing of these can similarly show how risks have changed over time and between different groups. Statistical techniques that have been applied to surveillance data, for the purpose of detecting outbreaks or assessing the significance of observed changes in frequency, include the Cusum technique [2], particularly for rare events, the scan statistic [3], and more complex modelling approaches [4].

Interpretation of such analyses may need to take into account issues such as the seasonality of many communicable diseases and the periodicity, which may stretch over several years, shown by several diseases that predominantly occur in childhood. Discontinuities in long term time trends may be the result of interventions, such as the introduction of a new vaccine, but may also arise as the result of changes to factors unrelated to the true incidence of disease, such as the introduction of new diagnostic tests, changes in clinical practice that result in increased case ascertainment (e.g. the introduction of a new screening programme), or changes in coding systems (e.g. changes to the ICD system have resulted in significant discontinuities in trends in deaths attributed to some causes). Reporting delay can be an important factor in some surveillance systems, where there can be significant delays between onset or detection of disease and the date of reporting to the surveillance system. This can be adjusted for if the delay varies little over time, but those interpreting the data must be aware of such delays, since the data could otherwise be incorrectly interpreted as showing a fall in case numbers.

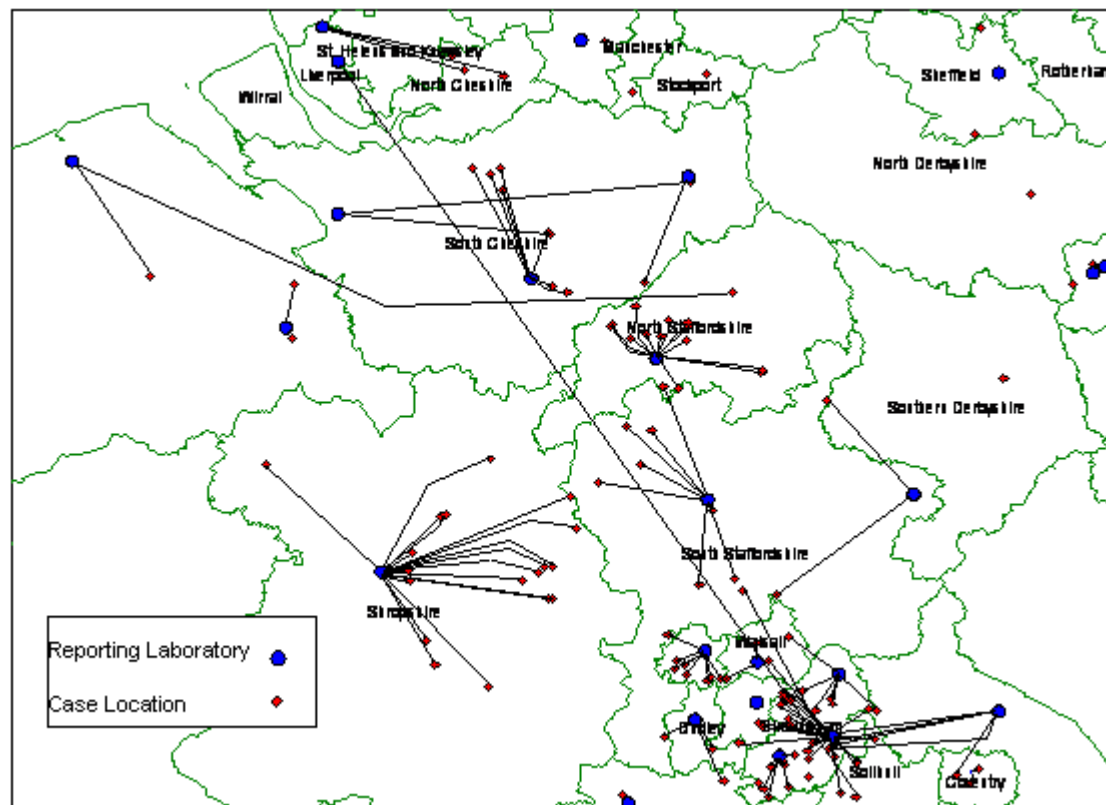
Analysis by person can include tabulation, graph display, or statistical comparisons of counts or rates by age, sex, ethnicity or other risk or exposure factor. This type of analysis can provide pointers to the aetiology or risk factors for acquisition of disease, and is increasingly used to demonstrate and monitor inequalities in morbidity between different population groups. Interpretation of apparent differences between populations or population sub-groups must take into account reporting biases. One of the most common reporting biases seen in surveillance systems is in relation to age, where particular age groups are relatively over or under reported. The very young and the very old are often better represented in surveillance data collected from laboratories, since these age groups are more frequently subject

to laboratory investigation for some common forms of infectious disease, such as respiratory or gastrointestinal infection. In the case of laboratory reporting of rubella infection, in contrast, it is women of child-bearing age who are often over-represented compared to other age and sex groups, because they are more likely to be investigated and reported. In countries with a significant mix of private and public healthcare services, particular population groups may preferentially attend one type of service compared to another, which would give rise to potential bias if surveillance was based on data from one type of service only, or if reporting was consistently better from one type of service. Misclassification and data errors can also have an impact on comparisons between different population groups.

Geographical information systems are increasingly being used to analyse surveillance data. These systems can be used either to increase the visual impact of display of geographical variations (i.e. to produce maps that show how cases are distributed geographically), or can be used for spatial analysis of surveillance data, testing for geographical clustering.

Common problems in the geographical analysis of surveillance data are missing information on the geographic location of cases, and the geocoding of data to the source of the report rather than the likely source of acquisition of infection. For example, an analysis of data from an outbreak of salmonella infection in England and Wales in 2000 shows a considerable difference in the geographic distribution of cases and of the laboratories that submitted reports on their salmonella infections (figure 1).

Figure 1. Geo-spatial analysis: Salmonella Typhimurium DT104 Outbreak, England & Wales, August 2000



Surveillance can only achieve its purpose of providing information for action if the information reaches those who have the responsibility for taking action. Although significant thought and investment is often put into data collection and analysis, when developing surveillance systems, the equally important process of dissemination of the resulting information can sometimes be given less attention.

The production of regular and timely surveillance outputs, and their dissemination in an appropriate format with relevant interpretation, requires significant investment. Development of outputs should be undertaken through close consultation with the target audience for the output, to ensure that they are fit for purpose. Some users of surveillance outputs will only require high level summaries that focus on key messages about overall changes in frequency in distribution, while others may require detailed line listings of cases in order to inform their own operational activities. Some users may wish to be able to manipulate surveillance data in their own systems (e.g. in their local geographical information systems), where they can undertake linkage or ecological analyses against other data that they hold. It is only through regular consultation with the relevant stakeholders that surveillance systems managers can ensure that their outputs continue to meet with recipients' requirements.

Advances in information technology, particularly browser-based web technologies provides the opportunity of making surveillance outputs available, or even pushing them through email or technologies such as RSS, to a large audience as soon as the outputs are ready. This is clearly of benefit in terms of speed and cost of delivery, but such benefits will only be realised if the outputs are relevant and easily understood by the intended audience - if not, they are likely to be overlooked in the face of increasing information overload.

References

1. Teutsch SM, Churchill RE. Principles and practice of public health surveillance. 2nd ed. Oxford, New York: Oxford University Press, 2000.
2. Gallus G, Mandelli C, Marchi M, Radaelli G. On surveillance methods for congenital malformations. *Statistics in Medicine* 1986; 5: 567.
3. Wallenstein S. A test for detection of clustering over time. *Am J Epidemiol* 1980; 3: 367
4. Farrington CP, Beale AD, Andrews NJ, Catchpole MA. A statistical algorithm for the early detection of outbreaks of infectious disease. *J R Statist Soc* 1996; 159: 547-63



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Opportunities and Challenges

Last modified at 3/11/2015 10:55 PM by Arnold Bosman

Many of the issues that face those responsible for the development and management of surveillance systems have been covered elsewhere in this chapter, such as:

- the need to reduce the burden on data providers in the face of increasing pressures to reduce the costs of healthcare
- the need to improve the completeness of coverage and the quality of data collected through surveillance systems
- the need for common data and information system standards that will facilitate the exchange of information between information systems (as healthcare services move towards electronic patient records)
- the need to comply with laws on Data Protection and Human Rights
- the collection of appropriate denominator data

There are also a new range of challenges for surveillance, related to the continuing emergence of new threats, changes in technology, changes in the physical and political environment, and changes in human behaviour. One of the greatest challenges of the last few decades has been the emergence of new diseases, including HIV, Ebola, West Nile virus, BSE, SARS and avian influenza among many dozens of others. Many of these diseases are zoonoses, and it is predicted that the trend in emergence of human disease arising from zoonotic sources is likely to continue [1], particularly as international trade and travel increase, and with the possibility of global warming. There is a need for surveillance systems that can provide better intelligence on environmental hazards and exposures, making use of veterinary and environmental information sources. There is also a need for syndromic surveillance systems that might not only detect emerging zoonoses, but also the occult release of infectious disease agents by bioterrorists.

The threat of bioterrorism, and the need to be able to undertake surveillance at the time of mass gatherings, such as the Olympic Games, is also posing new challenges for surveillance, and in particular the need for 'real time' surveillance systems. Exciting advances have been made in the development of syndromic surveillance systems and in the use of novel data sources such as over the counter medicines sales, as a step towards meeting these requirements,

and it is likely that this is an area that will see further development in the next few years.

The emergence of antimicrobial resistance and the development of more sophisticated disease control programmes also brings with them the need for surveillance that can be used to monitor the outcome of interventions, rather than just the events (cases of infection) that require those interventions.

Advances in technology, both diagnostic and information technology, provide exciting new opportunities for surveillance. The increasing move to electronic patient records and standards for exchange of health data between patient record systems provides significant opportunities, not only for more complete and more rapid capture of information about health events and exposures, but also for syndromic surveillance and for record linkage within surveillance.

Increasing computer processing power, and developments in geographic information systems also provide the opportunity for real time tempero-spatial modelling of emerging epidemics.

The development of molecular diagnostic tests, and of bio-informatics software for manipulating and analysing molecular sequence data, provides significant opportunities for more rapid and more precise data on the characterisation of infections, which could be used for surveillance. The advent of near patient testing also provides new opportunities for surveillance, not only in that it could provide earlier confirmation of aetiology in some clinical settings, but also because confirmation of aetiology may provide a prompt to reporting clinicians to seek important risk factor or exposure data while the patient is still in front of them.

References

1. Report of the WHO/FAO/OIE joint consultation on emerging zoonotic diseases. Food and Agriculture Organization of the United Nations (FAO), World Health Organization (WHO), and World Organisation for Animal Health (OIE), 2004 (http://whqlibdoc.who.int/hq/2004/WHO_CDS_CPE_ZFK_2004.9.pdf)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Case definitions for outbreak assessment

Last modified at 1/6/2012 1:45 PM by Arnold Bosman

From the start until the end of an outbreak investigation, it will remain important that we have an overview of the number of cases that have occurred over time. For this reason we need an exact definition for what we will call a case.

We can share this definition with those involved in finding / reporting the cases and those analysing the data.

We may need a different type of [case definition](#) for initial outbreak appraisal than for the analytical investigation. For example at the start of the investigation, when we want to find all cases of the outbreak to determine the magnitude of the problem, we may want to use a very sensitive case definition, to avoid that we miss cases.

Later, during the analysis phase, we usually want to switch to a very specific case definition, in order to avoid misclassification (hence bias our results). These [practical aspects](#) of case definitions make that they may evolve during investigation.

[Various levels](#) in case definitions are usually the categories *possible*, *probable*, *confirmed*, which may help us to make the case definition more sensitive or more specific.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Confirm the outbreak

Last modified at 10/14/2011 3:18 PM by Arnold Bosman

Confirming the outbreak requires that we have reliable data about the number of cases that we would expect in the same area and within the same time period in which the cases were reported. So can we define what we mean with «more than expected»?

We can't.

Sometimes we even do not have reliable information about the background rate of cases. And even with reliable expected numbers, how much higher does it make an outbreak? One standard error? Two standard errors? Partly it will depend on the level of background incidence. In the end, it remains a subjective choice.

Most of the time it will take judgement in addition to data to confirm an outbreak. A single case of a communicable disease long absent in population can therefore be an outbreak. (Eg : one case of rabies in a country with the 'rabies free status', one indigenous case of polio in the EU). For some diseases it may also depend on the season: 1000 cases per week in the summer may be consider 'an outbreak', while the same number in winter may be considered 'normal'.

Information sources that can help to determine the expected number of cases, such as [surveillance systems](#), hospital registries or surveys.

OK, we have more cases than expected. Now what?

Any increase could mean several things:

1. **An increase in population at risk** (=denominator), for example due to an influx of migrants. If the number of cases doubles, because the population has doubled, then we would have expected that, and in the light of our definition, it would not be an outbreak (however, it could still require public health attention).
2. **Random fluctuations**. If in a city of 1 million inhabitants a rare disease occurs in only 1 patient per year, and suddenly we have 2 cases in a year, then this could still be within the variance of the expected. Again here. it

helps to have the denominator

3. **Registration artefact:** it could be a classification error, or a better diagnostics was used, or a screening programme introduced. Any kind of change to 'the system' that suddenly detects more cases, while the 'true' number of cases in the population has not changed
4. **A true rise in cases** (a 'real outbreak').

When we have no denominator information whatsoever, then we still want to try to rule out causes 1,2 and 3 so that we are more confident that it is 'more cases than expected'

Explanation 1 is difficult to assess without reliable information about the denominator. However, we may still have some information. If we know that migration did not take place, and that birth rate and death rate have not changed significantly, then we can assume that the population has been stable. So without knowing the exact rate, we can consider population changes as an unlikely cause of the increase

Explanation 2 is easier to test, even when we only have absolute case counts, and no rates. Obviously, we first need to rule out differences in the size of the population at risk (see above). Then it helps to have historical data. By observing the fluctuations in the past, we can calculate the expected (usually the mean of the observed from the past) and the variance. This will help us determine how many standard deviations the observed value is from what we expected. To be pragmatic, if the observed value is more than 2 standard deviations from the expected, we consider this explanation 2 unlikely.

Explanation 3 is also independent of population size: we need to know the diagnostic and reporting system very well.

So even in absence of accurate denominator information, we should be able to make a reliable assessment if an increase in cases is due to 'a real outbreak' or due to one of the other three explanations.

In retrospect it is always easy to recognise an outbreak, or an increasing trend. Yet when you are confronted for the first time with an increase, it may be surprising how much doubt on the interpretation is around.

Morabia tells us: [Around 1900, lung cancer was extremely rare. Its incidence seem to grow at a fast pace, but evidence did not convince everyone. It was argued that a better diagnosis and aging population could explain the trends. An editorial in the British Medical Journal in 1942 stated:

"It is doubtful whether the higher incidence of cancer of the lung observed in recent years is real or only apparent."]

It took the same journal 10 years to first comment that *"few trends are more dramatic than the rise during the last 30 years in the notified death rates from cancer of the lung."*

In retrospect we all have 20/20 vision. The question is rather: how much of an open mind do we all have interpreting the present?

References:

1. Alfredo Morabia (editor). A history of epidemiologic methods and concepts. Birkhauser Verlag, 2004. ISBN 3-7643-6618-7



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Describe the outbreak

Last modified at 9/19/2011 9:24 AM by Arnold Bosman

This pages still needs content. Please help us by contributing



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Analytical Study Designs

Last modified at 3/1/2015 9:31 AM by Arnold Bosman

Where descriptive epidemiology describes occurrence of disease (or of its determinants) within a population, the analytical epidemiology aims to gain knowledge on the quality and the amount of influence that determinants have on the occurrence of disease. The usual way to gain this knowledge is by *group comparisons*. Such a comparison starts from one or more hypotheses about how the determinant may influence occurrence of disease.

For example, the hypothesis may be "people who have eaten home preserved green olives in restaurant X in August 2006 have an increased risk of developing botulism than those who have not eaten such olives".

We can test this hypothesis in an analytical epidemiological study where the risk of developing botulism is studied in 2 comparable populations; one group consists of people that have visited restaurant X in August 2006 and who did eat green home preserved olives. The other group consists of guests of restaurant X in August 2006 that have not eaten those olives. In both groups the risk of developing botulism is measured (by counting botulism cases that occurred in each group within 30 days after visiting the restaurant). Then those two risks are compared to see if they are significantly different.

Observational studies

In the above example of a simple analytical epidemiological study, a traditional [cohort study design](#) was chosen.

Another group of traditional study designs that belongs to analytical epidemiology are [case control studies](#). Other less traditional analytical study designs include [case-case studies](#) and [case-cross over design](#). In each of these analytical studies, observations in one group in the population are compared to another group (also called 'reference group').

[Choosing the appropriate reference group](#) is one of the challenging aspects of analytical epidemiology.

The examples above belong to the category of 'observational studies' in analytical epidemiology. In such studies, the investigator observes systematically how exposure and outcome are distributed in the populations, and the comparison of those observations is made.

Experiments

Another category of analytical studies are 'experimental studies', for example in which the investigator is able to randomly assign exposure to individuals from a particular population after which the occurrence of disease is measured in exposed and unexposed groups. Such experiments are called 'randomised controlled trials (RCT)' and are usually considered the gold standard in analytical epidemiology since the amount of [bias](#) is usually very limited. However RCT are not an option if the exposure is known to be very dangerous to humans, in which case it would not be ethical to conduct a RCT. In our example above, it is very clear that a RCT would be completely unacceptable (i.e. deciding randomly which guest should eat green home preserved olives, and then to count botulism cases among exposed and unexposed).

Therefore in [Field Epidemiology](#) we are usually left with observational study designs, to observe the 'experiments that nature has created for us'. This often creates challenges in [finding appropriate comparison groups](#).

EPIET Lectures:

[Case Control and Cohort Studies](#)

[Choice of a reference group](#)

[Alternative study designs](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Definition of an Outbreak Investigation Report

Last modified at 2/25/2012 1:57 PM by Arnold Bosman

What is an outbreak report?

An outbreak report is a document summarising [all the steps](#), main findings and recommendations of the outbreak investigation. It is a public record which has to be objective to reflect the reality of the investigation, clear so that decision makers and the scientific community can understand and use its content, and timely to provide feedback and recommendations on time to those who need them. It is better to have a short outbreak report soon rather than a very long document many months or years after the investigation.

Why write them?

- The outbreak report is a document used to launch a control action, to guide [public health measures](#) and to inform decision makers.
- The outbreak report is also a record of performance, summarising how the investigation has been conducted, its strengths and limitations. It presents the investigation methods and the control measures taken so that others can judge the validity of the results and the appropriateness of the actions carried out.
- Potentially it can be used for legal issues
- Writing the outbreak report represents a good opportunity for the investigator to revise the results, check errors and often while revising the evidence, new questions arise that would need further research.
- The outbreak report is an excellent teaching material from which many others in the future can learn.
- Finally, the outbreak report should be written because it helps preventing future outbreaks and assists in the investigation and control of similar outbreaks

Who writes the outbreak report?

The outbreak control team should write the outbreak report. Each member should contribute to have a comprehensive report including all the steps and aspects of the investigation.

The authors should be named specifying their role in the outbreak control team.

Each participating agency should agree with the document and this is often a challenge, especially if the aim is to have a timely report.

Another issue which needs to be agreed upon before starting writing the report is who owns the report: the participating agencies? the local authorities? the national authorities? the patients? the public?

Who reads them?

The outbreak report is a document that each agency represented in the OCT will use to document its activities. Policy making bodies request it to have evidence for public health action. Moreover, professional colleagues will use it as reference for other similar outbreaks. The public may be also interested in reading the report to evaluate risks and be informed about the subject. Finally, as mentioned previously, if there is a suit against one of the involved parties, the outbreak report can be requested as legal document.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Outbreak investigations 10 steps, 10 pitfalls

Last modified at 10/6/2014 4:15 PM by Vladimir Prikazsky

	Specific recommendations	Pitfalls to avoid
1. Determine the existence of an outbreak	<ul style="list-style-type: none">- Determine whether there is a clustering of cases, a cluster of cases of an outbreak-prone disease or a single case of a disease of international importance.- Review incidence in the past in the area of the outbreak.- Check for recent changes in the surveillance system (numerator).- Check for recent changes in the population size (denominator).	Taking all reported clusters at face value: Reported clusters may be pseudo-outbreaks. Check all reports for background rates, changes in surveillance practices (e.g., increased awareness) and change in the denominator (e.g., population movements).
2. Confirm the diagnosis	<ul style="list-style-type: none">- Make clinical description of a few cases to raise hypotheses in terms of diagnosis.- Collect the right biological specimens the right way to confirm the suspected diagnoses.- Send the biological samples safely and appropriately packaged to the right laboratory.	Failing to obtain a laboratory diagnosis Every effort must be made to obtain a diagnosis as early as possible during the outbreak. This includes obtaining a careful clinical description of the cases and obtaining laboratory confirmation. Ask for assistance with respect to collecting and transporting specimens and identifying the correct laboratory for analysis, if needed.

3. Define a case	<ul style="list-style-type: none"> - Formulate a time, place and person case definition, using generic case definition if applicable (e.g., WHO, CDC, MoH). Multiple levels are possible, including sensitive case definitions (adapted to the descriptive stage) and a specific one (more adapted to the analytical stage). 	Defining cases poorly Cases must be defined with some attention and precision; otherwise, the case count may too large, too small, or inaccurately defined. A good case definition is essential to hypothesis generation. Have precise criteria, and use time, place and person elements. Seek help if needed.
4. Search for cases	<ul style="list-style-type: none"> - Search for cases within the time and space limits of the case definition. - Compile and update a line listing of cases (e.g., on a spreadsheet) For each case, document at least the date of onset, age, sex, the zone of residence and the outcome. 	Conducting a door-to-door case search or a survey upfront Case search does not need to be done through a door-to-door survey all the time. In most cases, you can keep these undertakings for the second part of the investigation (hypothesis testing). For the descriptive initial part, you can (1) search for cases through surveillance and (2) obtain denominator separately. The case search strategy does not need to be 100% exhaustive: it needs to be uniform.
5. Generate hypotheses using descriptive findings	<ul style="list-style-type: none"> - Describe the outbreak over time through an epidemic curve. - Draw a spot map, and if possible, a map with incidence / 1000 population by area of residence. - Calculate population-based incidence by age and sex groups. - Conduct hypothesis-generating interviews with case-patients to try to find out what is common to all case-patients. 	Merging the hypothesis generating and the hypothesis-testing stages The descriptive stage generates information (1) through epidemiological information organized by (a) time, (b) place and (c) person and (2) through hypothesis generating interviews. Surveys conducted in the absence of a hypothesis clearly defined on the basis of this type of information blur the distinction between the two stages of the investigation and may seriously impair the capacity to formulate a conclusion.
6. Test hypotheses with an analytical study	<ul style="list-style-type: none"> - Write a mini-protocol to spell out the hypotheses to test and the study design to use. - Conduct an analytical study (case control or cohort). 	Believing that a questionnaire constitutes a study protocol The analytical step is a careful epidemiological study. It requires a design and an analytical plan before it is initiated. A case control study is not always the answer. Do not rush to the questionnaire but rather

		follow each of the 10 steps. If you do a study, write a one-page mini-protocol in bullet format.
7. Draw conclusions	<ul style="list-style-type: none"> - Analyze the analytical epidemiological study. - Formulate conclusions that explain the facts observed. 	<p>Having excessive confidence in the conclusions</p> <p>The final conclusions of an investigation are not reached as soon as a p value happens to be under 0.05. Formulating conclusions requires review of causality criteria, examination of the proportion of cases exposed to the suspected source, discussion of other possible explanations and a double check to see whether the source identified or the hypothesis considered explains all the descriptive findings.</p>
8. Compare hypothesis with established facts, additional studies	<ul style="list-style-type: none"> - Conduct an environmental assessment guided by the results of the analytical study. - Review literature. - Discuss conclusions with colleagues, peers and supervisors. 	<p>Rushing to conduct an environmental assessment</p> <p>In most cases, your environmental assessment will be guided and focused by the analytical epidemiology findings to further confirm a hypothesis. It is not a fishing expedition conducted at the early stages of the investigation where all kinds of samples are tested in the absence of any hypotheses to try to find an answer.</p>
9. Communicate findings	<ul style="list-style-type: none"> - Write a one-page draft summary report to leave in the field before departure. - Communicate findings with supervisors, the laboratory and local public health authorities. 	<p>Failing to communicate the results to decision-makers</p> <p>An investigation is not complete until the results have been communicated to those who need the information to act. A number of target audiences will need to receive the information in an adapted medium to engage in what they should do. Sending the report to a supervisor is not sufficient.</p>
10. Execute prevention measures	<ul style="list-style-type: none"> - Formulate clear, specific feasible recommendations on the basis of your findings (Who? What? When? How?). - Ensure implementation of the recommendations. - Evaluate the relevance and effectiveness 	<p>Formulating general recommendations that are not based upon findings</p> <p>Recommendations need to focus on those interventions that would have prevented the outbreak or that will control it. They should be guided by the results of the investigation, based upon evidence, focused and feasible. Do not re-formulate all the</p>

	of the recommendations.	recommendations of hygiene but focus on the specific ones that are the key issue in the outbreak.
--	-------------------------	---



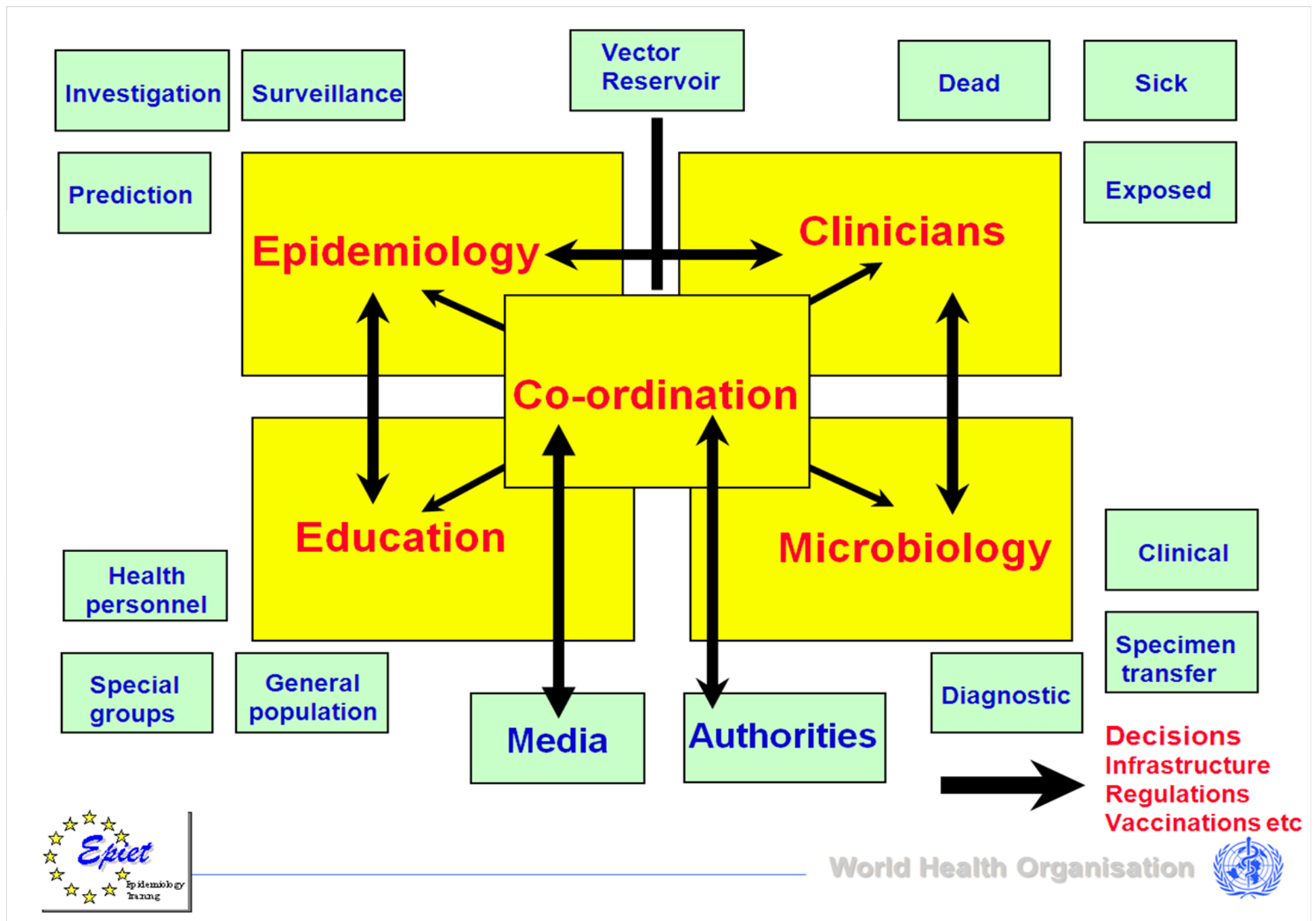
The outbreak management team

Last modified at 3/1/2015 8:17 PM by Arnold Bosman

Once the proper authorities have decided that a public health response is needed, then usually an outbreak team will be in charge of the response. The response is often focussed around the [10 steps of an outbreak investigation](#). For many of those steps, multidisciplinary expertise is required. For example steps 2 and 3 (confirm diagnosis and establish a case definition) often require close collaboration between field epidemiologist and public health microbiologist.

Steps 5 and 6 (determine population at risk and generate hypotheses) may need additional disciplines, depending on the context of the outbreak. For example with a vector borne disease, input from an entomologist may be required, in order to determine the spread of the vector and hence define the population at risk.

All of those different disciplines need to work in harmony, focussed on the same priorities to get through the 10 steps as quickly as possible. In large teams, coordinating the full outbreak response is the role of the outbreak manager. This person may have any of the disciplines as a professional background, with sufficient experience of having worked in many different outbreaks. Once you take the job of the outbreak coordinator, however, you need to be able to 'take some distance from your own discipline' and focus on the management tasks. So whether you are originally an epidemiologist, public health microbiologist, clinician, entomologist or public health decision taker, your job as coordinator is to focus on the process, and to align the work of all involved disciplines.



in smaller outbreaks and usually at the sub-national public health level it could happen that all of those responsibilities are carried by one person. In that case the challenge is to know when to 'switch your hat', for example in one moment you are designing questionnaires and data entry screens, and the other moment you are negotiating with the director of a nursing home to include the kitchen staff in the survey, and an hour later you may be at the ward, taking blood samples from cases, while at the end of the day you explain to the reporter of the local newspaper whether or not there are risks for the visitors of the nursing home.

It is just important to remember that in principle nothing changes in the steps that need to be taken in outbreak response. A main difference between small local outbreaks and large national ones, if the size of the outbreak team,

and the need to coordinate and communicate at every level and during every step of the way.

Multiple disciplines, one team, one objective

Whether the manager of the outbreak team has a background in epidemiology, microbiology, health science or generic public health, he/she shares one single objective with the whole team:

- **To maximise the scientific quality of the investigation in a complex environment**

The specific outbreak investigation objectives may vary according to the context of the outbreak and usually include "identify the source of the outbreak" or "identify the main risk factor for infection". The various disciplines contribute their expertise to achieve those objectives.

The field epidemiologist within the team is in charge of all epidemiological aspects of the investigation ([study design](#), [questionnaire](#), data handling, [descriptive analysis](#), [statistical inference](#)) and brings this expertise in the team discussion.

The public health microbiologist is in charge of all microbiological and diagnostic aspects of the investigation (human, veterinarian, food, water, environmental, specimen processing, selection and interpretation of diagnostic tests, biosafety aspects).

In addition, the manager needs to know just enough of each of the disciplines within the team, to allow setting priorities and to decide in situations where specialists in the team cannot reach agreement.

Operational challenges in outbreak investigations

The aim of coordinating an outbreak investigation is to maximize the scientific quality of the investigation in a complex environment. What could be the operational challenges? Well, here are a few:

- It is an unexpected event, so the public health system may not have been prepared to respond
- Need to investigate quickly, especially if the source may still expose other people
- Pressure for answers: family members, patients, journalists, politicians may want to have answers that you are yet unable to provide
- Multiple agencies may have responsibilities: who will coordinate and which responsibility has priority?
- Work will need to be carried out at many levels
- You are in the media spotlight
- Often, your investigation will have legal implications, as law enforcement agencies may want to use your evidence
- You may arrive late; all kind of actions may have been started and you may need to cancel some decisions.
- You may be working in a foreign country
 - You may be unaware of local sensitivities
- You will need to feedback to various people
- Stress, long working hours
 - finding time to rest and relax

You will need a structural approach to manage all challenges at one. Consider the following aspects:

Reasons for inviting you

- You may have specific expertise
- They may need more resources
- Perhaps they want to share responsibility
- There may be political or mass media pressure to invite you as an expert
- It may be mandatory or in guidelines
- They may need you to confirm local findings
- They may expect you to perform specialised investigations

Each of these reasons may require you to perform differently. It will be helpful to agree on some written 'Terms of Reference' (ToR). Such a written agreement should clarify:

- What are their expectations
 - expertise, tasks, time?
 - what local resources are available?
- What has already been done?
- What resources do you need to bring?
- What is your role ?
- Who is in charge?

Preparing to leave for the field

- Consult colleagues (microbiologist, vet, entomologist....)
- Review pertinent literature
- Decide who will lead the team
- Identify who provides support in head office
- One page report before leaving
 - objectives
- Arrange initial meeting for your arrival
- Discuss with your colleagues at the Institute to organise follow up of your ongoing projects !!!!!
- Organise your personal / family life

- Bring your 'Epi-Pack':
 - Computer, calculator, mobile phone
 - Software
 - USB, CDs
 - Notebook
 - File templates
 - Standard questionnaires
 - Handbooks, relevant articles
 - Camera
 - (Laboratory equipment)
 - Telephone, address list:
 - reference centers & persons
 - Maps, GPS
 - Others... (money, "health kit",)

When you arrive in the field

- Provide help - don't take charge
- Meet with key people
- Review and update status of problem
- Assess sensitivities
- Identify local resources and skills
- Discuss liaison
- Set up communications with base

Outbreak Control Team

- Membership
- Leadership
- Responsibilities
- Lines of communication
- Communication
- Decision making process

EPIET Lectures:

[Logistical Aspects of Outbreak Investigations](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Analysing a public health problem

Last modified at 12/8/2014 11:17 PM by Arnold Bosman

it is easier to talk about 'a public health problem', than to be able to define this in a [SMART](#) way. There are many approaches to problem analysis, and the choice of approach will depend on the professional culture. The common elements in problem analysis include: [\[1\]](#)

- specific question or assignment from policymakers
- Definition of objectives and specific study question
- Exploration of possible solutions (and the information & knowledge required for these solutions)
- Options for study designs to answer the questions

Assignments from policy makers

Policy makers aim to create conditions and practices that achieve certain strategic targets. In public health such a strategic target could be "protection from childhood diseases", and policies to achieve this can include childhood vaccination programmes. In real life, obstacles can occur that obstruct achievement of the targets, for example low vaccine coverage.

Policy makers may worry about determinants of low vaccine coverage, and may ask for "research to find answers that address this problem". Since this is not a very [SMART](#) request, we need to get a clear and concise idea of what exactly is the request and then to agree with the policy maker on a specific question that is researchable, and that can be answered by a study.

During your interview with the policymakers you may discover that their main concern is about loss of confidence in vaccines among parents, which leads them to reject vaccination for their children. In that case you can propose to answer the question "*what factors influence parents' decision to vaccinate their children?*". Do you see how much more specific (and researchable) this question is compared to "*find answers that address the problem of low vaccine coverage?*" The policymaker may tell you "yes, but that is what I meant, isn't it?". This may be true, but at least now you know specifically in which direction to aim your study. Then you may move to the next step, which is....

Definition of objectives and specific study question(s)

The question "*what factors influence parents' decision to vaccinate their children?*" may seem already [SMART](#) enough to start a study, yet there may be a number of pitfalls that we want to avoid.

What do we know of the vaccination system? How is vaccination offered, who does that, and what sources of information (regarding any aspect of vaccination) do parents have? Do schools require vaccination of their pupils? etc.

In other words, we need to understand the key elements of how vaccines are promoted, offered, administered, financed, monitored in our population, in order to enable us to decide if we can then study that factors as possible determinants for the decision of parents to vaccinate children. This will also allow you to

Explore possible solutions

Can each of these factors be changed by our policy makers or decision takers? If a factor cannot be changed, then it may not be of practical use to spend resources to study the influence of this factor on the outcome, *even though it may still be interesting to the researchers*. If our study fully depends on resources that we get from policy makers, we need to avoid that we study only what **we** think is interesting, but it mainly needs to enable them to do something with the results.

Never assume anything ! Always verify in a direct dialogue by asking questions.

Once we have restricted the aim of our study to specific objectives and questions we can address, then we move to the next step.....

Options for study designs to answer the questions

Choosing a study design depends a lot on the questions that you want to answer and on the budget available. Yet before looking even at study design, we may need to take a bigger step back and decide what type of research we may want to apply.

Empirical-Analytical research

This is what we mostly and traditionally do in [field epidemiology](#). We perform systematic observations of what happens in society (empirical) and analyse the observation results in an objective, controllable and reproducible way. [Cohort studies](#), [case control studies](#) and other [analytical study designs](#) belong to this class of research. Most of this research is quantitative or semi-quantitative: results are expressed in numbers.

The limitation of this type of research is that it usually requires strict definitions of hypotheses in advance. If we are unaware of the major determinant of the outcome of our interest, then this type of research will not allow us to find it. We only find what we define to look for.

Interpretative research

With interpretative research you explore the 'experiences of the study subjects'. What are their ideas, experiences,

hopes, fears? This type of research does not only look at 'the bare numbers', yet is usually qualitative in nature. It is often used in anthropology and sociology.

In our example of the interest of policy makers in determinants of parents' decision to vaccinate their children, we may decide to start with interpretative research, to get an overview of concerns and triggers for parents. Perhaps in a second stage, empirical-analytical studies can quantify the effect of those determinants (though this may not even be a necessary step, and it depends on the possible solutions that policy makers consider).

An example of methods used in this class of research is case study design [2]. Case studies are used in many sciences, including medicine. Though epidemiologists are often reluctant to extrapolate findings from limited number of cases to the whole population, single case studies can increase knowledge in complex areas enormously. For example the description of Mr Phineas Gage, a mid 19 century American railroad foreman who lived for 12 years with a hole in his head and a large portion of his brain's frontal lobes destroyed, gave essential neuro anatomical insights.

Sometimes we just need to recognize when it is relevant to publish observations on a single individual.

Critical-emancipatory research

This third class of research is not explicitly quantitative nor qualitative. Any type of data collection method may be used, with exception of 'experimentation'. Usually the researchers aim to contribute to processes in society that promote emancipation of groups. Their view is not only critical towards society, yet also towards the own study results.

References:

1. Verhoeven, Nel. [What is applied research? Methods and techniques for higher education] (Dutch language). Boom Lemma uitgevers, Den Haag, Netherlands. 2011
2. [The Qualitative Report Volume 14 Number 1 March 2009 42-60](http://www.nova.edu/ssss/QR/QR14-1/diaz-andrade.pdf) <http://www.nova.edu/ssss/QR/QR14-1/diaz-andrade.pdf>



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

SMART definitions

Last modified at 10/31/2011 9:34 PM by Arnold Bosman

SMART definitions (e.g. of objectives) are

- **S**pecific
- **M**easurable
- **A**ction oriented
- **R**ealistic
- **T**ime-related

Each time when you prepare a definition, ask yourself: *is it smart?*



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Cohort studies

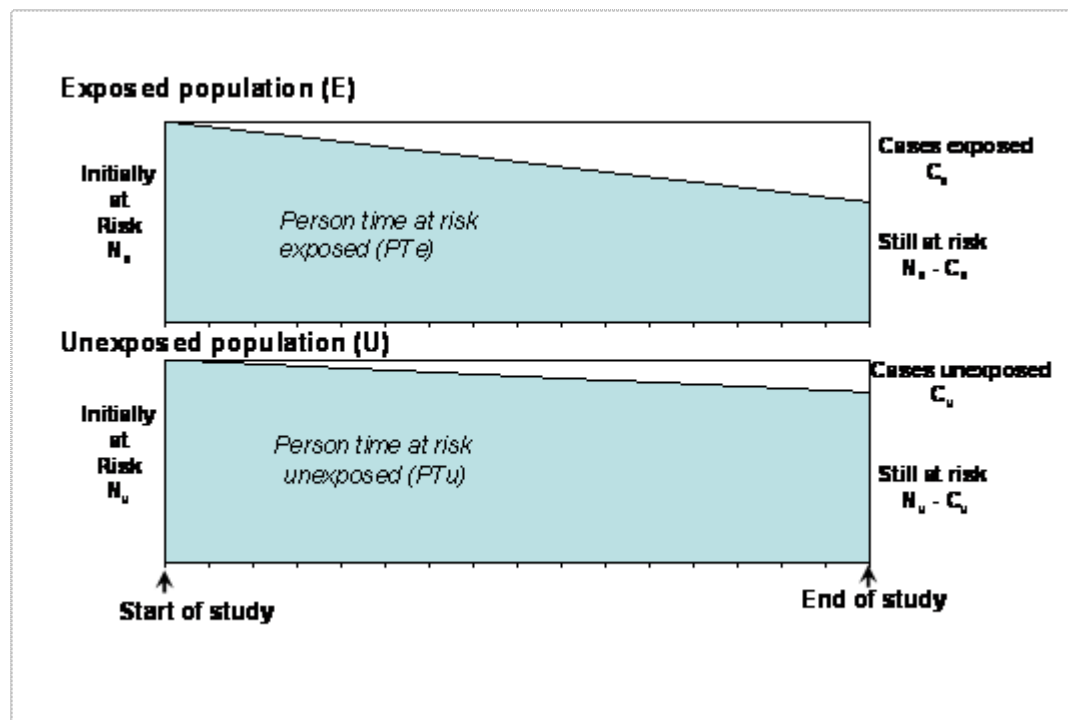
Last modified at 1/19/2015 10:10 PM by Arnold Bosman

Cohort Studies

In Romans' time, a cohort of legionnaires consisted of a group of soldiers sharing the same military events for a certain period of time. In epidemiology we consider that a *cohort* consists of people belonging to the same population and sharing similar experience for a defined period of time.

Cohort studies involve the comparison of [disease incidence](#) over time (risk or rate) between two subsets of a population (two [cohorts](#)). One of the 2 cohorts is exposed to a certain characteristic (exposure). The other is not. All other things being equal between the two cohorts but for their exposure. In both cohorts we measure [occurrence of disease](#) over the specific study period. However whenever the condition of "all other things being equal" is not met, the comparison might be wrong.

The following graph adapted from Rodrigues [\[1\]](#) illustrates occurrence of cases over time in the two cohorts. Initially Ne persons are exposed and Nu persons are unexposed. The number of persons who are disease free decreases over time (shaded area). The number of cases (non shaded area) increases over time but more in the exposed cohort. At the end of the study, respectively C_e and C_u cases have occurred in the exposed and unexposed cohorts. The shaded area represents the cumulative time during which persons were at risk of developing disease in each of the cohorts during the entire follow up period.



In this example, the Risk of disease in the exposed cohort (R_e) = C_e / N_e and the Risk of disease in the unexposed cohort (R_u) = C_u / N_u .

In a cohort study, we can compare those 2 risks, in order to see if exposure has an effect on the risk. One comparison is to look at the difference: **$R_e - R_u$** is also called the risk difference. The risk difference shows us what the absolute increase (or decrease) of the risk is when exposure occurs.

Another comparison is to see how the relative increase (or decrease) occurs after exposure: this is the Risk Ratio: **R_e / R_u** . This is also called 'Relative Risk'.

In addition to risks, we can also measure rates in cohort studies: in such a situation, the observation time in the cohort is taken into account in the denominator (for example: 51 cases per 1200 person-years). When we compare the rate of cases in the exposed cohort with the rate of cases in the unexposed cohort, then we consider that a Rate Ratio.

References

1. Rodrigues L. Kirkwood BR. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. Int J Epidemiol 1990 Mar;19(1):205-13



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Case-control studies

Last modified at 1/26/2016 4:45 PM by Vladimir Prikazsky

Case control studies

In cohort studies the denominator represents the exposure experience of the source population. If it is the exposure status as observed at the beginning of the cohort we will compute a risk. If we allow for exposure to vary overtime we will compute a rate which takes into account the time spent by each individual in the exposed and unexposed cohorts over time.

The main constraint in cohort studies is the necessity to collect information on exposure from large populations (to have denominators for the exposed and unexposed cohorts). We will see below that instead of collecting exposure information from the entire study population we can use a sample of it to calculate or estimate the risk ratio or the rate ratio. In other words, instead of using the entire cohort denominator we will use a sample of it. This sample is also frequently called a control group and it is used to represent the exposure experience of the source population.

The rationale behind using a sample of the denominator comes from the following formula for risk and rate ratio which can alternatively be expressed as follows:

For risk

$$RR = \frac{\frac{Ce}{Ne}}{\frac{Cu}{Nu}} = \frac{\frac{Ce}{Cu}}{\frac{Ne}{Nu}}$$

For rates

$$RR = \frac{\frac{Ce}{P_{Te}}}{\frac{Cu}{P_{Tu}}} = \frac{\frac{Ce}{Cu}}{\frac{P_{Te}}{P_{Tu}}}$$

From the above formula we already see that if we take an unbiased random sample of Ne and Nu the ratio of exposed to unexposed (Ne/Nu) will not be modified and therefore the risk ratio will remain the same (Ne/Nu or a sample of it gives the same risk ratio if sampling is unbiased). The same applies if we use person years at risk (PT).

We have generally speaking three major ways to select a sampled control group which reflect three ways to estimate exposure experience in the source population.

1. Controls are selected from people who are still free of the disease at the end of the study period (N_e - C_e and N_u - C_u). We will call the related study design a [traditional case control study](#) since it is the design most frequently used. The exposure measured reflects the exposure experience or status of people still free of disease at the end of the cohort.
2. Controls are randomly selected from the population present (at risk) at the beginning of the study (N_e and N_u in the above graphic). The related study design is called a [case cohort study](#). The exposure measured reflects exposure status at the beginning of the cohort.
3. Controls are selected proportionally to the person-time contributed by exposed and unexposed cohorts (PT_e and PT_u). The related study design is called a [density case control study](#). The exposure measured reflects the varying exposure of people at risk along the cohort.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Cross-sectional Studies

Last modified at 10/24/2011 7:43 PM by Arnold Bosman

This article is a result of new content structure of the FEM Wiki.

You are invited to contribute suitable content such as definitions, scope, examples and other related material. Feel free to link to other external resources.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

[Statistical Concepts](#)

Concepts in sampling

Last modified at 2/6/2012 2:52 PM by Arnold Bosman

In order to end up with a representative sample of sufficient size to perform the study, we need to define several elements:

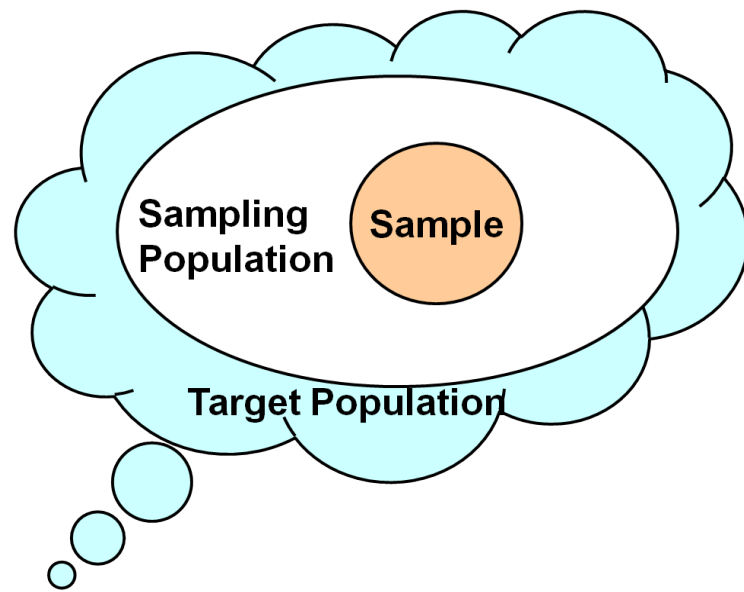
What is the sampling unit?

This is the smallest element of the sampling population in which we will make the observations of the study. Usually this is an individual person; the observations we make by questionnaires or interviews targeted at individual members of the study population. However it could be a household, a school, or even entire towns or cities.

For example if we want to study the relationship between religion and vaccine coverage, this would be done by studying "the population of towns in the country" and to measure within each town the distribution (proportions) of various religions and the vaccine coverage.

What is the sampling fraction?

This is the proportion of the target population that is included in the sample. For example, if the total size of the target population that we want to study is 100.000 people, and in the sample we include 500, then the sampling fraction is 0.5%



What is the sampling frame?

In the ideal situation, we have a comprehensive and exhaustive registry of all units of our population. Such complete population registers at national level make it possible to draw samples directly from this registry, according to our preferred scheme.

However, what do we do if such a system does not exist? What if we have no way of knowing each last individual in the population. Then we go looking for the best list of registered units that we believe is representative of the whole population. When we have such list of all the sampling units from which sample is drawn then such a list is the sampling frame.

For example we can have a list of all children < 5 years of age, or of all households in a province, or of all health care units in a district.

Which sampling scheme to use?

There are different methods (sampling schemes) of selecting sampling units from sampling frame, for example:

- Simple random sampling
- Systematic sampling
- Stratified sampling
- Probability proportional to size sampling
- Cluster sampling
- Quota sampling
- Convenience sampling or Accidental Sampling
- Line-intercept sampling
- Panel sampling

<<<THIS CHAPTER REQUIRES COMPLETION AND FINDING A DEDICATED EDITOR. INTERESTED? Write below !>>

Stratified Analysis

Last modified at 4/12/2016 7:39 AM by Vladimir Prikazsky

The existence of effect modifiers or confounding factors requires measuring an effect in subgroups (strata) of the study population. We perform a stratified analysis. The effect modifier or confounding factor can have two or several categories. Each of them is a stratum in the stratified analysis. We measure the effect between exposure and outcome in each of the various levels taken by the effect modifier or the confounding factor.

The relevant table looks as follows

	Cases	Total	Attack Rate	Risk Ratio
Stratum 1	a_1	Te_1	Re_1	RR_1
	c_1	Tu_1	Ru_1	
Stratum 2	a_2	Te_2	Re_2	RR_2
	c_2	Tu_2	Ru_2	
Stratum 3	a_3	Te_3	Re_3	RR_3
	c_3	Tu_3	Ru_3	
Stratum 4	a_4	Te_4	Re_4	RR_4
	c_4	Tu_4	Ru_4	
TOTAL	Sa	STe	Re	RR
	Sc	STu	Ru	

To conduct a stratified analysis we can identify six major steps which have a specific chronology:

1. Conduct a crude analysis

Measure the effect (RR or OR) of the exposure of interest on the outcome in our study. Compute the confidence limits of this effect.

2. Identify the potential [effect modifiers](#) or [confounding factors](#)

Those variables are identified from the crude analysis of the data or identified a priori from literature review. They include the other identified risk factors (variables which are associated with outcome) and variables which can be subdivided in several sub groups of public health interest (age, gender, etc.). When the effect modifier or confounding factor is not binary (Yes-No) we create as many strata as there are categories of exposure in that variable.

3. Measure the effect of exposure on outcome within each stratum

Measure the effect of the exposure on outcome within each of the strata (RR₂ to RR₄ above).

4. Look for [effect modification](#)

If the effect differs between strata, we then suggest that effect modification is present. This should be supported by a test for homogeneity between strata and a reflection on the biological plausibility of the varying effect among strata. Since effect varies among strata we need to present the results by stratum. An overall effect (crude effect) is less informative since not illustrating the information given by the effect measured in each stratum.

5. Look for [confounding](#)

Compare the crude measure of effect to a weighted measure (e.g. [Mantel-Haenszel](#)).

If the crude and weighted measures differ by more than 15-20%, the crude measure of effect may have been confounded. The weighted measure of effect is therefore more appropriate than the crude measure of effect. The crude measure of effect can be compared to the range of value taken by the stratum specific effects: if it lies outside the range of stratum-specific values, then confounding is likely.

6. Are effect modification and confounding present?

If both effect modification and confounding are present, the interpretation of a measured effect is complicated (a variable can be both a confounding factor and an effect modifier). In that event a [multivariable analysis](#) taking into account confounding and interaction is needed [1].

[References](#)

1. Hosmer DW, Lemeshow S. Model-Building Strategies and Methods for Logistic Regression. 2nd ed. New Jersey, USA: John Wiley & Sons Inc; 2000.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Questionnaire Design

Last modified at 10/30/2011 11:32 PM by Arnold Bosman

Summary

Questionnaires are one of the most important tools for field epidemiologists. An [advantage](#) of using a questionnaire is that it enables you to reach a large number of persons at once. On the other hand, using a questionnaire might be a [disadvantage](#) if you have little information on the subject and do not ask the right questions.

Questionnaires are used in outbreak investigations, and applied research as well as in evaluation of surveillance systems. The success of an investigation or a research project largely depends on a high response from the target population and the amount of valid information that was obtained.

There are different [types of questionnaires](#): Questionnaires can either be filled out by the respondent (self-administrated) or by an interviewer in a telephone or a face-to-face interview (interviewer-administrated). The suitable format depends on the study question, the target population and the available resources. Similarly, there are different [types of questions](#), open and closed. Answers to open questions will provide you with additional information and comments which you might have considered beforehand. [Closed questions](#) provide a limited choice of answers and are therefore easy to analyse.

A well designed questionnaire will provide appropriate data which allow answering your research question. It will minimise potential sources of [bias](#), thus increasing the validity of the questionnaire. A well designed questionnaire is much more likely to be completed. Therefore, creating a good questionnaire is crucial for the success of your project.

This chapter should help you to distinguish between the different types of questionnaires, questionnaire administration and question formats. You can follow the [10 steps](#) for designing a questionnaire or check the quality of your questions with the useful [hints](#) that are given. For example, the [seven golden rules](#) will guide you in perfecting your questionnaire. You can also find useful information on [piloting questionnaires](#) and [validated questionnaires](#).

See also the following EPIET Lectures:

[Questionnaire Design](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Data collection instrument

Last modified at 9/7/2014 9:35 PM by Vladimir Prikazsky

Writing a data collection instrument: A practical guide

Version 2 - 5 September 2014

A. Checklist for self or external assessment

Points	Tick	Item
1.		Items on the instrument match the objectives of the study and the analysis plan: All items are relevant and no item is missing.
2.		Items do not suggest any specific answer through including information or other elements that could influence respondents.
3.		The instrument does not collect identifiers or surrogate identifiers (e.g., dates of birth). These are collected separately.
4.		For questionnaires, items are worded as full questions that can be asked directly to participants.
5.		The instrument includes a skip pattern that anticipates the need to jump specific items according to some answer.
6.		When multiple answers are possible, the instrument specifies whether one or multiple answers are acceptable.
7.		Questionnaires are 'participant friendly' (introduction, thank you statement, logic flow, clear wording, respect and tact).
8.		The instrument specifies how to collect information (e.g., structured observation, interview, record

		review)
9.		Items are well formulated (Clear, avoid negatives, specific, focused. Aim at collecting information about one topic and does not confuse different issues that should be the focus of two different questions)
10		The instrument is field worker friendly (Instruction for administration, guidance, auto-coding and numbering).

B. 10 common errors seen in data collection

1. Data collection instrument not matching the objective of the study and / or the analysis plan	
<i>Description of the error</i>	<i>Rationale to change</i>
The data collection instrument is not developed on the basis of the objectives of the study. It does not follow the analysis plan. Some items may be missing and some may be unnecessary.	The data collection instrument is a logical deduction of the analysis plan, not the reverse. All the items on the instrument must be written in anticipation of the analysis.
<i>Examples of the error</i>	<i>Correction needed</i>
<ul style="list-style-type: none"> Missing items (for example: For a study on factors associated with receiving antenatal care among pregnant women, no items on the geographical and financial access to antenatal care) Presence of items that do not match the objectives of the study (for example: For a study on factors associated with receiving antenatal care among pregnant women, list of items about the kind of care that was given during antenatal care) Excessive length, unnecessary questions 	<ul style="list-style-type: none"> Analyze the problem, conduct a pilot qualitative study and make sure you make a good inventory of the items that need to be covered as per the study objectives. Stick to items that match the study objective (for example: Stick to items on the knowledge, attitude and practices that may increase or decrease the probability of receiving antenatal care) Shorten, focus on the objectives of the study

2. Items influencing the participants	
<i>Description of the error</i>	<i>Rationale to change</i>
The item is worded in a way that influences the participant to provide a specific answer.	For objective data collection, no items should suggest any answer.
<i>Examples of the error</i>	<i>Correction needed</i>

Do you know that the routine antenatal check up includes three visit?	Split in two questions: "What should a woman do for her health when she expect a baby? (Do not suggest answers but have an option ready for "antenatal visit" if mentioned)". If antenatal visits are mentioned, ask "How many antenatal visits should take place?"
---	---

3. Data collection instrument collecting identifiers or surrogate identifiers

<i>Description of the error</i>	<i>Rationale to change</i>
The data collection instrument contains information that can identify the study participant directly (Name, address, phone number) or indirectly (Date of birth).	All identifiers and surrogate identifiers must be eliminated from the instrument that may be accessible to many people: Data entry clerk, other staff.
<i>Examples of the error</i>	<i>Correction needed</i>
<ul style="list-style-type: none"> • Name and address on the questionnaire • Date of birth on the questionnaire 	<ul style="list-style-type: none"> • Replace by ID code and keep track of identifier in separate identifier sheet kept under lock and key by the primary investigator. • Stick to year of birth if possible.

4. Questionnaire items not fully worded as questions

<i>Description of the error</i>	<i>Rationale to change</i>
Some questions on a questionnaire are not fully worded as questions ready to be asked to study participants.	For quality assurance purposes, each participant needs to hear the question in the exact same way. Thus. The exact wording should be proposed by the primary investigator and not left for the field worker to decide.
<i>Examples of the error</i>	<i>Correction needed</i>
<ul style="list-style-type: none"> • "Education of woman" 	<ul style="list-style-type: none"> • "Did you attend school?" If yes: "What is the highest class that you attended?"

5. Absence or inappropriate skip pattern

<i>Description of the error</i>	<i>Rationale to change</i>
The items follow each other on the instrument without any anticipation that some items may not apply to some	"Skip patterns" plan the use or non-use of specific questions according to the answer given to previous

people given their response to a previous question.	questions. These allow smooth administration of the questionnaire and avoid asking questions that do not apply a particular person (which could generate confusing answers).
Examples of the error	Correction needed
<ul style="list-style-type: none"> “Have you heard about tetanus vaccination during pregnancy?”, then “Did you receive tetanus vaccination during your pregnancy?” and “How many doses did you receive?” 	<ul style="list-style-type: none"> Use: “What care should a woman receive during pregnancy?”, then skip other questions if tetanus is not mentioned. If mentioned, ask “Did you receive tetanus vaccination during your pregnancy” then skip other questions if answer is “No”. Then ask: “How many doses did you receive?” if answer was yes.

6. Unclear multiple answer options

Description of the error		Rationale to change	
Items have multiple answer options but do not specify whether one of more than one answer options is acceptable.		From an analysis plan point of view, the number of acceptable answer options changes the nature of the question. If more than one answer option is acceptable, logically, that is equivalent to as many items as there are answer options with a dichotomous yes / no answer. If only one answer option is acceptable, then it is one item with a categorical variable.	
Examples of the error		Correction needed	
What are the reasons why you did not register your pregnancy at the health centre?	a. Services were not available	What is the main reason why you did not register your pregnancy at the health centre? (Choose one)	a. Services were not available
			b. It was too far
	b. It was too far		c. I had no time
			d. I did not know it was needed
	c. I had no time	Among these factors, which ones contributed to the fact you did not register your pregnancy at the health	a. Services were not available Y/N
			b. It was too far Y/N

	d. I did not know it was needed	centre?	c. I had no time Y/N
			d. Didn't know it was needed Y/N

7. The questionnaire is not participant-friendly

<i>Description of the error</i>	<i>Rationale to change</i>
The questionnaire is thought as a good data collection tool but it has not been polished for use with study participants. Common errors include the absence of an introduction, the absence of thank you statement, the absence of logical sequence, jargon and inappropriate tone.	While the initial stage of the development of the questionnaire need to be centred around the analysis plan, the tool then need to be adapted for use with study participants.
<i>Examples of the error</i>	<i>Correction needed</i>
<ol style="list-style-type: none"> 1. Absence of introduction, transition sentences and thank you note. 2. Poor logical sequence 3. Have you had multiple sex partners in the last 12 months? 4. Did you receive iron and folic acid tablets? 	<ol style="list-style-type: none"> 1. Add introduction, transitions ("I will now ask you about your past pregnancies") and thank you statement. 2. Re-order sequentially, logically, from the general to the specific, from the public to the private. 3. How many sexual partners have you had in the last 12 months? 4. Did you receive the tablets against anemia?

8. Lack of clarity about the data collection procedure

<i>Description of the error</i>	<i>Rationale to change</i>
The instrument is unclear as to whether the information needs to be collected through interview, review of document or observations.	To standardize the data collection procedure, specify exactly how the information needs to be collected if it is not through interviews.
<i>Examples of the error</i>	<i>Correction needed</i>
<ul style="list-style-type: none"> • Type of house: a. Brick b. Mud 	<ul style="list-style-type: none"> • <i>Observe the house and write the type:</i> a. Brick b. Mud

<ul style="list-style-type: none"> How many doses of tetanus toxoid did you receive? 	<ol style="list-style-type: none"> Review the vaccination card. Number of tetanus toxoid doses received according to the card: ____ If no vaccination record, ask the participant: How many tetanus vaccine doses did you receive?
---	--

9. Poorly formulated questionnaire items

<i>Description of the error</i>	<i>Rationale to change</i>
Items are poorly formulated: They are unclear (e.g., excessive use of negative), unspecific or unfocused.	Poorly formulated questionnaire items will confuse the field workers and the study participants.
<i>Examples of the error</i>	<i>Correction needed</i>
<ul style="list-style-type: none"> Do you think it is important to attend antenatal clinic? 	<ul style="list-style-type: none"> What are the benefit that antenatal clinic provide?
<ul style="list-style-type: none"> Did you miss vaccination because you thought you did not need it? 	<ol style="list-style-type: none"> How many vaccination doses did you receive? Do you think the vaccination was needed in your case?
<ul style="list-style-type: none"> Did you miss any of the antenatal visits? 	<ul style="list-style-type: none"> How many antenatal visits did you attend?

10. The questionnaire is not friendly to the field workers

<i>Description of the error</i>	<i>Rationale to change</i>
The instrument is not field worker friendly: It contains no instruction for administration; no guidance, no auto-coding features and no items are numbered.	A questionnaire that is easy to use for field workers will be better filled.
<i>Examples of the error</i>	<i>Correction needed</i>
<ol style="list-style-type: none"> Absence of instruction for administration. No auto-coding No numbering of the questions 	<ol style="list-style-type: none"> Insert guidance for administration in italic or in a different font so that field workers know if the text is for their guidance or a question to read. Insert a column for auto-coding. Number each questionnaire item.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

[General Communication](#)

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Health literacy

Last modified at 4/17/2015 8:07 AM by Arnold Bosman

Health literacy can be defined as the capacity that an individual has to access and effectively use health-related information, in order to promote and maintain good health. While literacy can enable people to understand and communicate health information and concerns, when these are applied to a health context, it is called health literacy. A person can be literate and still have limited health literacy. In the report Healthy People 2010, the U.S. Department of Health and Human Services define it as "the degree to which individuals have the capacity to obtain, process and understand basic health information and services needed to make appropriate health decisions." Examples of actions that require health literacy skills include properly reading and adhering to a care or prevention program as well as being able to use the available healthcare services rationally and ponder individual behavioural change. Increasing health literacy rates is a means to empower patients and contribute to downgrade inequalities towards a healthier, safer, more demanding society.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

[General Communication](#)

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Health education

Last modified at 4/17/2015 8:08 AM by Arnold Bosman

One of the main tasks of health education is to inform about lifestyles and behaviours that prevent people from various diseases. In this sense, health education aims to influence a person's knowledge, attitudes and behaviours connected to health in a positive way. It is a process during which people learn how to take care about their own and other people's health. Initiatives can either focus on improving existing medical problems or preventative education (e.g. prevent people from acquiring various diseases or guide them on how to live with a disease), in any combination of planned learning activities. Health Education models of intervention have evolved in the past three decades. The field is quite diverse in Europe on what concerns approaches and levels of integration with public health programs.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Social marketing

Last modified at 4/17/2015 8:10 AM by Arnold Bosman

Even though the word “marketing” is often associated with advertising and promotion, the application of the methods to social non-for-profit causes and programmes has proven to be a helpful tool to enhance the effectiveness of efforts to protect and improve public health. Using social marketing tools to conduct public health improvement programs can help to clarify goals and improve success with limited public health resources. Health-related social marketing aims at improving people’s health and quality of their life in concrete social, political and economic environment. It requires familiarity with the audience to whom health messages are being addressed, knowledge about relevance and importance of the problem to the target groups, ability of the group to tackle the problem and potential to promote change of an existing situation in a concrete environment. Understanding social marketing principles and techniques is key to developing public health programs that can promote knowledge or positive behaviours as well as reduce risky ones.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Risk communication

Last modified at 12/7/2016 10:38 AM by Vladimir Prikazsky

Risk communication needs to be considered at all stages of risk management. It is a sustained communication process established with a diverse audience about the likely outcomes of health and behavioural attitudes. The main goal is to engage communities in discussions about environmental and health-related risks to create public understanding about their outcomes and approaches to deal with them. Risk communication can be about specific health-related choices, e.g. the perceived risks associated with getting immunised, or related to behaviours, as the risks associated with sexual behaviour. This approach requires a profound understanding of the distinction between the different dimensions and models of behavioural sciences.

Ten golden rules for risk communication

1. Never lie
2. Never say 'no comment'
3. There is never an 'off the record'
4. Be short, get to the point and always think of the audience
5. Stay calm and confident
6. Use simple language
7. Stay in control
8. It is OK to say "I don't know, but I'll find out"
9. Don't speculate
10. beware of reporters' tactics

What do people remember from an interview on TV:

- 55% Body language (professional, interested, calm, nervous, intimidated etc)
- 38% Tone of voice (concerned, calm, worried, relaxed, happy)
- 07% Words (content)

In radio interviews, usually only 9 seconds of the interview will be used within the story they want to present. Make sure such a "sound bite" complies to the following:

- No jargon
- Simple language
- Positive active verbs

Apply the "27/9/3 rule": Maximum 27 words, average 9 seconds, no more than 3 messages

The 3 C's of communication:

- Consistent content
 - Explicit information
 - Accuracy is key
- Clear message
 - Messages that leave no one guessing.
- Courtesy conveys respect

Another option of the 3 C's - clear, concise, consistent.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

[General Communication](#)

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Crisis communication

Last modified at 4/17/2015 8:11 AM by Arnold Bosman

While risk communication is ongoing, crisis communication is a reactive communication effort in the face of an unforeseen event. It is often unpredictable and unexpected, develops suddenly, takes uncontrolled course and evokes uncontrollable reactions. An open, honest, and ongoing interaction with the audience remains essential to successful communication during crisis. Focus is kept on the message, the time of its announcement and the media used, and some common advice to improve include: Do not allow a delayed reaction; Do not abandon pro-active action or allow for no action at all; Do not allow lack of preparedness to communicate in a crisis; Do not ignore the needs and expectations of the mass traditional media; Do not allow for lack of communication with external partners and stakeholders; Do not allow lack of internal communication in health systems and organisations; Do not allow information chaos; Do not play down the complexity of audience diversity; Promote careful elaboration and proactive planning of potential actions related to crisis communication as a crucial element in eliminating the unexpected characteristic of a crisis and probably prevent it or at least avoid its uncontrolled course;



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

[General Communication](#)

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Health advocacy

Last modified at 4/17/2015 8:13 AM by Arnold Bosman

Advocacy is one strategy to raise levels of familiarity with an issue and promote health and access to quality health care and public health services at the individual and community levels. When trying to gain political commitment, policy support, social acceptance and systems support for a particular public health goal or programme, a combination of individual and social actions may be used to try to affect change. This is one way of understanding Health Advocacy. The adoption of a health advocacy model can focus on an educational dimension when it identifies emerging public health issues that require action. It encompasses gathering information on existing practice related to public health, related legislation monitoring and providing feedback on how specific regulations impact local groups and communities. It may also help guiding health policy reforms. Often, health advocacy is carried out using mass and multi-media, direct political lobbying and community mobilization. It may materialize within an institution or through public health associations, patients' organisations, private sector and NGOs. All health professionals have a major responsibility to act as advocates for public health at all levels in society.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

[General Communication](#)

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Outbreak communication

Last modified at 4/17/2015 8:14 AM by Arnold Bosman

Acknowledging that “communication expertise has become as essential to outbreak control as epidemiological training and laboratory analysis”, in 2005 the World Health Organization created Communication Guidelines aiming at clarifying the specific communication challenges faced by public health officials as well as the best practices for communicating with the public during an outbreak of a communicable disease. An effective outbreak communication is one of the tools that can help achieve the public health goal of bringing an outbreak under control as quickly as possible, with as little social disruption as possible. The guidelines identify some fundamental aspects for outbreak communication practice: Trust; Announcing early; Transparency; Understanding the public;



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

[Infection control and hospital hygiene](#)

Introduction to Public Health and basic concepts

Statistical Concepts

Burden of HAIs

Last modified at 4/21/2016 7:24 AM by Vladimir Prikazsky

This article is a result of new content structure. No text available yet.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Types of HAIs

Last modified at 4/21/2016 7:26 AM by Vladimir Prikazsky

- Catheter-associated urinary tract infections (**CAUTI**)
- Central line-associated blood-stream infections (**CLABSI**)
- Surgical site infections (**SSI**)
- Ventilator-associated pneumonia (**VAP**)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

[Infection control and hospital hygiene](#)

Introduction to Public Health and basic concepts

Statistical Concepts

Main pathogens and resistance

Last modified at 4/20/2016 3:46 PM by Vladimir Prikazsky



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Risks and Rates

Last modified at 11/4/2010 10:21 PM by Aileen Kitching

Risks

Risks(synonyms: incidence proportion, attack rate) are expressed as a percentage.

In a population, the [risk](#) of a disease measures the proportion of people that develops this disease between two specified points in time (T_0 and T_1). It is calculated as the number of persons developing the disease during the observation period divided by the number of persons present at the beginning of the observation period. The denominator represents the population at risk of developing the disease at the beginning of the period of observation. The time spent by each individual in the observation period is not taken into account.

$$\text{Risk} = \frac{\text{Number of new cases occurring between } T_0 \text{ and } T_1}{\text{Number of persons present in the population at } T_0}$$

A risk is a probability. It will range from 0 to 1, or 0% to 100% if expressed as a percentage. The risk of disease is therefore also called incidence proportion.

In an example in which 50 out of 200 residents of a nursing home developed gastroenteritis between 12 and 20 May 2005, the risk of gastroenteritis is $50/200 = 0.25$ or 25 %.

In intervention epidemiology risk is used for short periods of follow up like outbreaks during which all individuals are assumed to be followed for the same period of time.

In outbreaks the term risk is frequently replaced by [attack rate](#), which is also expressed as a percentage. Even though widely accepted in the epidemiological community, the term "attack rate" is technically incorrect, as an attack rate is not a rate but a risk. In outbreaks, an food specific attack rate measures how many people who ate a certain type of food (or had an other exposure) became ill (else, "were attacked by that food").

The measurement of risk shows the probability of developing illness during a specific period of time. Without knowledge of the time period it is however impossible to interpret a risk [1]. A risk of death of 2% will imply different meanings if it expresses the risk of death during one month, one year or a 30 year follow up.

Rates

Rates are a measure of occurrence of a phenomenon [2]. A rate is calculated as the number of events that have occurred, divided by the total time experienced by the population under observation, usually expressed in person years. Rates are usually multiplied by a power of 10, to convert the rate into a decimal or whole number which is easy to interpret.

$$\text{Rate} = \frac{\text{Number of events occurring between } T_0 \text{ and } T_1}{\text{Total time experienced by population between } T_0 \text{ and } T_1}$$

In a study, it is possible that each person experiences the same event more than one time.

Incidence rates are a subset of rates, in which we are interested at quantifying *new* events ("new cases" of disease)

The term **rate** is sometimes used incorrectly in epidemiology. For example to replace the term risk (as in **attack rates** and **case fatality rates**).

References

1. Rothman KJ; Epidemiology: an introduction. Oxford University Press 2002, p.24-28.
2. Porta, M. A dictionary of Epidemiology, Fifth edition. Oxford University press, 2008.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Experimental Studies

Last modified at 9/14/2011 2:12 PM by CeRC

This article is a result of new content structure of the FEM Wiki.

You are invited to contribute suitable content such as definitions, scope, examples and other related material. Feel free to link to other external resources.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Descriptive Studies

Last modified at 3/1/2015 10:01 AM by Arnold Bosman

Descriptive epidemiology studies and summarizes patterns of disease or of disease determinants in terms of time, place and person. The results are used to understand a population's health status, generate hypotheses about the causes of diseases, and inform program planning and evaluation[1]. In other words, descriptive epidemiology describes the distribution of disease (recall that epidemiology is the study of the distribution and determinants of disease in populations)[2]. This is done by describing a health outcome by different characteristics of person (race, age, or sex, for example), place (geographic location), and time (a specific year or a span of time). For example, the case fatality of cholera in 1854 in London was 40% ([John Snow, cholera outbreak in London](#)).

In a certain sense, public health surveillance may be considered as an ongoing descriptive epidemiological study and you may find many examples of ways to describe disease and determinants in terms of time, place and person in the [articles on descriptive data analysis](#).

References:

1. Aschengrau, Ann, and George Seage. *Essentials of epidemiology in public health*. Jones & Bartlett Learning, 2008.
2. Public Health Social Network - what is descriptive epidemiology ([website accessed 1 March 2015](#))



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

[Introduction to Public Health and basic concepts](#)

Statistical Concepts

Transmission routes

Last modified at 11/1/2011 12:22 PM by Arnold Bosman

The pathway of causative agents from a [source](#) to infection of a susceptible host is called 'transmission route'. The characteristic of the transmission route depends mainly on the characteristics of the causative agent and those of the host. Some micro organisms are restricted to a limited number of transition routes, whereas others can follow many different pathways to infect their hosts. It is useful to have detailed knowledge about the specific transmission routes of pathogens, since this gives practical information of effective [control measures](#) by interrupting the spread of the infection within the population.

Direct transmission

This means direct and immediate transfer of infectious agents to a susceptible host. This may be through direct contact such as touching, biting, kissing or sexual intercourse, or by the direct projection of droplet (droplet spread) spraying onto eyes, nose or mouth of other people during sneezing, coughing, spitting, singing or talking. Droplet spread is usually limited to short distances, such as 1 meter or less).

Direct transmission routes are linked to behavior, and most interventions that target this particular transmission usually aim to educate people to reduce risk behavior (e.g. condom use, using facial masks while contacting patients, sneeze in handkerchiefs or sleeves, etc)

Vertical transmission

A specific form of direct transmission is that between mother and child during pregnancy or childbirth.

Indirect transmission

When transmission of infectious organisms occurs from a source through **objects (vehicles)** or **insects (vectors)** we call this indirect transmission. Transmission through **vehicles** is usually linked to **processes**, such as food production, food handling, cleaning procedures in day care centers, hygiene procedures in medical facilities etc.

Vehicle-borne

Infectious agents can reach susceptible hosts through transport on inanimate objects (fomites) such as toys, handkerchiefs, soiled clothes, bedding, medical instruments, food, water, blood products or any other substance that can be contaminated. Some vehicles allow multiplication of the infectious agent (e.g. salmonella in food), though this is not always the case. [Intervention measures](#) to reduce the risk of vehicle-borne transmission aim to enhance the safety of procedures (safe food production, education of food handlers, hand washing protocols in health care, etc).

Vector-borne

When insects transfer infectious agents to susceptible hosts, they act as 'vectors' of the infection. [Intervention measures](#) to reduce the risk of vector-borne transmission aim to control the size of the vector population. This may include spraying of insecticides and reducing the breeding spaces for insects.

Mechanical:

This includes simple mechanical carriage by crawling or flying insects, and does not require multiplication of the micro organisms.

Biological:

When the micro organisms multiply within the vector and / or undergo developmental cycles within the insect, then this is part of biological vector-borne transmission of infectious diseases. In such cases, an incubation period is required (starting from the moment of introduction of the infectious agent into the vector) before the vector itself becomes infective. In such situations, infected insects may transmit the pathogen (vertically) to the next generations of offspring.

Airborne transmission

Microbial aerosols are suspensions of particles (fluid or solid) in the air consisting partially or wholly of microorganisms. They may remain suspended in the air for prolonged periods of time (as opposite to droplets, that are too large in diameter and fall to the ground relatively fast). This transmission route works particularly efficient for viruses such as the measles virus (a coughing patient may produce an infectious aerosol in a corridor that can remain suspended to infect others passing by several minutes or longer after the patient has left).

References:

1. David L. Heymann (ed.). Control of Communicable Diseases Manual. 19th Edition.
2. Patrick L. Remington, MD; William N. Hall, MD, MPH; Irving H. Davis, PE, CIH; Anita Herald, MD; Robert A. Gunn, MD, MPH. Airborne Transmission of Measles in a Physician's Office. JAMA. 1985;253(11):1574-1577. doi: 10.1001/jama.1985.03350350068022



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Probability

Last modified at 2/2/2012 6:38 AM by Arnold Bosman

When we ask ourselves the question "how certain am I that a specific event will occur?", then we are interested in **probability**. This is very much the same as asking the question 'what is the **risk** that this will happen?'.

Probability can be expressed as a percentage, permillage or in other words: a [proportion](#).

A related, yet slightly different concept is [odds](#).



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Measures of Disease Impact

Last modified at 10/13/2015 9:01 AM by Vladimir Prikazsky

Epidemiology is not just about identifying risk factors for disease but also about evaluating control measures or [public health interventions](#) to reduce or eliminate the effect of these risk factors. It is therefore important to be able to predict the impact of removing a particular exposure (or risk factor) on the incidence of disease in the population. This information can help policy makers decide on how best to allocate resources to ensure the most beneficial impact on public health.

Many diseases are caused by more than one exposure. For example, primary hepatic cancer may be caused by exposure to excess alcohol consumption, hepatitis B infection or hepatitis C infection. In order to assess the potential public health impact of a hepatitis B vaccination strategy on the incidence of primary hepatic cancer, we need a way of [quantifying disease burden](#) associated specifically with hepatitis B infection.

In order to do this, we need a way of measuring the proportion of the disease that can be attributed to the exposure. The relative risk (or risk ratio) is used as a measure of the effect of an exposure on an individual's risk of disease. However, to assess the impact more generally we also need to know the number of individuals that are exposed (the prevalence of exposure). This chapter therefore begins by exploring the concepts of [relative risk versus attributable risk](#).

Impact among the exposed and in the population

Measures of impact should help to answer questions like these:

- How much of the disease can be attributed to a particular exposure?
- How much of the disease can be prevented by eliminating a particular exposure?

For the public health policy maker it is helpful to answer these questions from two perspectives:

- What is the impact on people who are exposed to the risk factor?

What is the impact on the population as a whole?

This chapter explains how impact may be measured in both the [exposed](#) group and in the entire [population](#). It gives examples of how these measures are calculated and explains what they mean.

Details are given of how to calculate each of the following measures:

- [Attributable risk among the exposed](#)
- [Attributable fraction among the exposed](#)
- [Attributable fraction in cohort studies](#)
- [Preventable fraction in cohort studies](#)
- [Attributable fraction in case-control studies](#)
- [Attributable risk in the population](#)
- [Attributable fraction in the population](#)

Impact numbers

In clinical medicine, the number needed to treat (NNT) is used as a measure of treatment effect. It is the number of persons that need to be treated to achieve one beneficial outcome (e.g. cure) or to prevent one adverse outcome (e.g. relapse).

However, this measure has limited usefulness in a public health context when the impact of an exposure on the risk of disease is being assessed. In this situation we are more interested in calculating, for example, the number of people in a population among whom one case may be attributed to the exposure. This chapter therefore concludes with a brief discussion of [impact numbers](#). These are a range of measures that have been developed to express these public health concepts.

Further reading

Suggestions are given for further reading about the [general principles](#) of measuring impact, and some [examples](#) of the use of measures of impact in field epidemiology.

Learning Objectives

After reading this chapter, you will be better able to:

- understand the difference between relative and attributable risk
- understand the difference between attributable risk among the exposed and attributable risk within the population
- calculate and interpret attributable risks
- understand the concept of impact numbers.

EPIET Lectures:

Measures of Impact



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Bias

Last modified at 3/21/2015 12:51 PM by Theodore Lytras

What is bias?

If an epidemiological study is viewed as a measurement exercise [1][2], then we need to consider how much we can trust the measurement (*risk*, *rate*, effect) obtained from that study. Can we use it to safely describe (accurately estimate) the *association* between an exposure (potential causal characteristic [2]) and a disease/outcome, or to conclude that a risk factor really does cause the disease/outcome in the population in which the study was done?

Bailey et al [3] define an *association* as a 'statistical dependence between two or more events, characteristics, or other variables'. According to Rothman [2], a measure of association compares what happens in two distinct populations (or sub-populations), although these two populations may correspond to one population in different time periods (e.g. before and after an event). Relative *measures of association* (e.g. relative risk/ risk ratio, rate ratio, odds ratio) estimate the size of an association between exposure and disease/outcome (*strength of association*), and indicate how much more likely people in the exposed group are to develop the disease/outcome than those in the unexposed group [3]. The presence of an association does not necessarily imply a *causal* relationship.

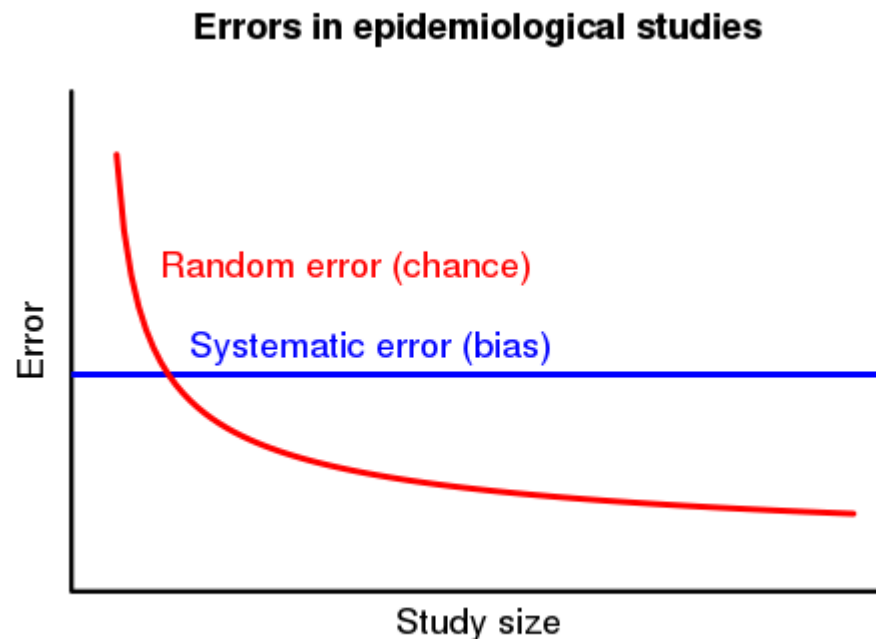
Before we can conclude that an observed *association* between an exposure (risk factor) and outcome (e.g. disease) as measured in our study is *causal*, and may reflect the true situation in the population, we first need to exclude other possible reasons why we might have obtained that result and be sure that the measurement/ result has been estimated with little error. We need to consider whether the result could be due to *systematic error* (*bias* or *confounding*) or to *random error* (due to chance). If we consider that the results reflect the true situation in the population, they then need to be interpreted according to *causality criteria*.

Random error reflects the amount of variability in the data [1]. Assessment of *random error* aims to distinguish findings (variations of observed values from the true population values) due to chance alone (findings that we cannot readily explain) from findings we could replicate if we repeat the study many times. *Precision* is the opposite of *random error*, and an estimate with little *random error* can be described as precise [2].

In epidemiological studies, *biases* are *systematic errors* that result in incorrect estimates when measuring the effect of exposure on risk of disease/outcome. Any error that results in a systematic deviation from the true [association](#) between exposure and outcome is a bias [\[3\]](#). *Validity* is the opposite of bias, and an estimate that has little systematic error can be described as valid [\[2\]](#). Biases may distort the design, execution, analysis and interpretation of studies [\[4\]](#). Some authors define bias more broadly. Daly's definition - defining bias as any factor or process that tends to produce results or conclusions that differ systematically from the truth - thus includes errors in analytical epidemiology and errors of interpretation [\[5\]](#).

Distinguishing random errors from systematic errors

As described in the following graphic adapted from Rothman [\[1\]](#), there is a way to distinguish [random errors](#) from [systematic errors](#). If we increase the size of a study until it is infinitely large (increase our sample size), [random errors](#) (due to chance) can be reduced to zero and corrected for. However, [systematic errors \(biases\)](#) are not affected by increasing the size of the study and will remain.



Modified from Rothman. 2002

In this chapter, we will focus on [systematic errors \(bias\)](#). Epidemiologists frequently classify [bias](#) into three broad categories: [selection bias](#) (bias in the way the study subjects are selected), [information bias](#) (bias in the way the study variables are measured), and [confounding](#) (described in a specific chapter).

[Selection biases in case-control studies](#) include among others: [case ascertainment \(surveillance\) bias](#), [referral bias](#), [diagnostic bias](#), [non-response bias](#), and [survival bias](#). [Selection biases in cohort studies](#) include: [healthy worker effect](#),

diagnostic bias, non-response bias, and loss to follow-up.

The term "misclassification" is frequently used to describe [information bias](#), the mechanism of which can be [differential](#) or [non-differential \(random\)](#). [Misclassifications](#) might be introduced by the [observer](#) ([interviewer bias](#), [biased follow-up](#)), by the study participants ([recall bias](#), [prevarication](#)), or by measurement tools (e.g. questionnaires).

References..

1. Rothman KJ. Epidemiology - An Introduction. New York: Oxford University Press; 2002.
2. Rothman KJ, Greenland S, Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
3. Bailey L, Vardulaki K, Langham J, Chandramohan D. Introduction to Epidemiology. Black N, Raine R, editors. London: Open University Press in collaboration with LSHTM; 2006.
4. Sackett DL. Bias in analytic research. J Chronic Dis. 1979;32(1-2):51-63.
5. Daly LE, Bourke GJ. Interpretation and Uses of Medical Statistics. 5th ed. Blackwell Science.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

[Introduction to Public Health and basic concepts](#)

Statistical Concepts

Applied Immunology

Last modified at 11/30/2013 9:28 AM by Arnold Bosman

Objectives of this chapter:

- Understand the difference between innate and addaptive immune system
- Define basic components of adaptive immune system
- Define important terms in immunology
- Explain major applications of immunology

Immunology is a broad branch of biomedical science that covers the study of all aspects of the **immune systems** in all living organisms. It deals with the physiological functioning of the immune system in states of both health and disease.

This chapter focusses on how the immune system deals with infectious diseases.

The immune system is the defense against all the pathogens that we may encounter during our life and it consists of 2 main systems: the **innate immune system** and the **adaptive immune system**.

The innate immune system

This is a general level of defense and is called 'innate' because we are born with it and it is genetically encoded; these traits that protect us from pathogens can also be passed to our offspring. The best general defense barrier we have is our skin; it is the physical barrier that protects our vital organs and is able to capture and kill organisms that try to penetrate it.

The best [routes](#) that pathogenic organisms can take to successfully penetrate our body is through our airways, via the food and drinks that we consume or by being injected through the skin directly into our blood (e.g. through an insect bite). However, also these 'ports of entry' have well developed innate defenses. Pathogens that enter our nose and throat are assaulted by chemicals and physical weapons. This includes the mucus (slime) that the epithelial cells in our airways produce, the tiny hairs (cilia) that use wiping movements to move the mucus (containing the captured

pathogens) out of our body and last but not least pathogen-eating cells (phagocytes).

These physical defenses above aim to prevent that pathogenic organisms enter our bloodstream or penetrate our cells. However, even when that fails, the innate defenses still have some options left:

For example when a virus enters a cell, it is recognized by the cell as an invader within minutes and the cell initiates a cascade of chemical signals to mobilize its defenses. Some molecules that the cell produces turn the intensity of the defense up or down and some molecules give signals to neighboring cells to start similar defense responses. Again other molecules damage the virus itself. If that is not enough to kill the virus, then sometimes the cell self-destructs in order to prevent the virus from further spreading.

All these defenses are examples of innate defenses, which aim to contain any invading pathogen. These responses also trigger a more specific defense that are specifically adapted to attack the pathogenic organism: the adaptive immune system. Virtually every successful vertebrate virus has developed mechanisms to circumvent the deadly traps that our innate immune system has put in place in our defense. This is why the adaptive immune system is so important.

The adaptive immune system

This system develops as we are exposed to pathogens throughout our lives: it is an acquired immune system. This system has a characteristic that the innate immune system lacks: immunological memory. Cells from the adaptive immune system are able to remember which pathogen they encountered, and store that memory for future rapid defense responses throughout our lives. The innate and the adaptive immune systems work closely together, for example some of the innate system's cells are so called **antigen presenting cells**, have the task to capture antigens from pathogens that have intruded in the human body and to present those **antigens** to the adaptive system, in order to start producing specific antibodies against the pathogen.

The key players of the adaptive system are T-cells (named so because they originate in the Thymus) and B-cells (originating from the bone marrow). Each of these cells have proteins in their outer coat that can bind with foreign antigens. There is almost an infinite range of different foreign antigens that the proteins on our B and T cells can recognize. This is similar to finding the key that fits a specific lock. The Antigen Presenting Cells of the innate immune system present foreign antigens that they have detected to our B and T cells in order to stimulate them for a response. This process usually takes place in the lymph nodes in our body.

B-cells have as one of their tasks to produce **antibodies**, that can then circulate in our blood stream to neutralize circulating pathogens while they are still outside of our cells. T-cells can directly kill an infected cell in our body, which will then stop the further replication and spread of the infecting pathogen. How do T-cells know that one of our cells is infected? Remember we said above that an infected cell can send chemical signals to mobilize our defenses? Well, some of these chemical signals attract T-cells. The next step is that the T-cells recognize some foreign antigens (from the invading pathogen) on the surface of our infected cells. That is usually the trigger for the T-cell to start killing the infected cell. This is a very effective and efficient way of dealing with pathogenic organisms, because only infected cells and only the pathogens are attacked, and our healthy cells are left alone.

There are disorders of our immune system that leads to attacks of our healthy cells: these so called 'auto-immune disorders' will not be discussed here.

Once T-cells and B-cells have been triggered for the very first time to respond to a specific infecting pathogen (thanks to the complex interaction with the innate system), they will remember this throughout our lives. This immunological memory will allow the T-cells and B-cells to respond immediately each time when this same pathogen tries to infect us again. This is the key concept in [\[\[wiki:/fem/w/wiki/vaccination.aspx|vaccination\]\]](https://www.wikipedia.org/wiki/Vaccination), where we artificially introduce harmless foreign antigens from dangerous pathogens, in order to trigger a lifelong immunological memory, without actually having to go through the dangerous primary natural infection.

Summary of our defenses

So in short, we have several lines of defenses:

- physical barriers such as our skin, the mucus and the sweeping cilia (hair cells) in our airways, the acid in our stomach
- mechanism to detect invasion by foreign organisms
- control and command centres (such as our lymph nodes) where targeted responses are coordinated, once detected
- our actual defense weapons: B- and T-cells and the antibodies, all grown to neutralize and kill invaders

Pretty sophisticated eh? With such specialized defenses, we are bound to be safe!

Unfortunately, the pathogenic organisms that surround us day by day have developed various smart ways of evasive action, to circumvent our defenses. Curious how? Read further [here](#).

references:

- Introductory Course EPIET/EUPHEM - Immunology lecture.
- Epidemics - the Dynamics of Infectious Diseases by Dr. Marcel Salathé, Dr. Ottar N. Bjornstad, Dr. Rachel A. Smith, Dr. Mary L. Poss, Dr. David P. Hughes, Dr. Peter Hudson, Dr. Matthew Ferrari, Dr. Andrew Read.[Coursera course by PennState university](#).



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Case Definitions

Last modified at 1/6/2012 1:52 PM by Arnold Bosman

Summary

One of the core tasks that epidemiologists do is to measure disease occurrence in a population, for example within a [surveillance system](#). The second major task is to compare those occurrences between sub groups of the population, analyse and interpret those differences (e.g. in [outbreak investigations](#) or other field investigations).

To measure disease occurrence we need to first count patients with a specific disease. To do so epidemiologists first define the disease. This is the case definition. It is a set of standard criteria used to decide whether a person can be counted as having a particular disease or not. By using a standard case definition we make sure that all counted cases of the same disease have been identified the same way regardless of whom has identified the case. We should emphasize here that the epidemiological meaning of a case definition may differ from the clinical and biological meaning. The case definition is a tool to count cases. It is not a tool to make a diagnosis and treat a patient. A standard case definition implies that some people with the disease will not comply with the case definition criteria and that some without the disease may be counted as cases.

Learning Objectives

After reading this chapter, you will be better able to:

- understand how to [define a case](#) in the context of public health investigations
- understand the difference between a case definition in public health and a clinical diagnosis in the health care setting
- apply [various levels of case definitions](#) (e.g. possible, probable, confirmed)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Impact numbers

Last modified at 6/12/2016 8:15 PM by Christian_Friedrich

Impact numbers are a relatively new concept. They are measures derived from case-control and cohort studies that are intended to be simple to understand and to help compare the population impact of different interventions [\[1\]](#) [\[2\]](#).

They are analogous to the concept of the number needed to treat (NNT) used in clinical medicine. This is the number of persons with a disease that, on average, must be treated in order to achieve one beneficial outcome (e.g. cure) or to prevent one adverse outcome (e.g. relapse).

The impact number reflects the number of people in each population (the whole population, the cases, all those exposed, and the exposed cases) among whom one case is attributable to the particular exposure or risk factor.

The following types of impact number are described below:

- [population impact number \(PIN\)](#)
- [case impact number \(CIN\)](#)
- [exposure impact number \(EIN\)](#)
- [exposed cases impact number \(ECIN\)](#)
- [population impact number of eliminating a risk factor \(PIN-ER-t\)](#).

The table below shows these impact numbers calculated from [the example](#) of the study of drunk driving and automobile related deaths.

Table. Impact numbers estimated from a cohort study (n=10,000) of drunk driving and automobile related deaths, Anystate, 2010

Impact measures	Abbrev.	Rate or number

Death rate in drunk drivers	le	150/1,000
Death rate in non-drunk drivers	lu	14/1,000
Death rate in all drivers	lpop	18/1,000
Attributable risk among exposed	ARe	136/1,000
Attributable fraction among exposed	AFe	0.91
Population attributable risk	ARpop	4/1,000
Population attributable fraction	AFpop	0.22
Population impact number	PIN	250
Case impact number	CIN	4.5
Exposure impact number	EIN	7.4
Exposed case impact number	ECIN	1.1
Population impact number of eliminating a risk factor	PIN-ER- <i>t</i>	40

Population impact number (PIN)

This is the number in the whole population among whom one case is attributable to the exposure or risk factor. It can also signify, for a protective factor, the number in the whole population among whom one case will be prevented by the exposure or intervention.

It is equivalent to the reciprocal of the population attributable risk (ARpop).

$$PIN = 1 / AR_{pop}$$

In the example of drunk driving and driving related deaths ([table - death from drunk driving](#)) there were 4 deaths per 1,000 drivers in one year attributable to drunk driving. We thus have:

$$PIN = 1 / (4 / 1,000) = 250$$

This means that, for every 250 people in Anystate, there is one driving related death attributable to drunk driving on average per year.

Case impact number (CIN)

This is the number of people with the disease or outcome among whom one case is attributable to the exposure or risk factor. It can also signify, for a protective factor, the number of people with the disease among whom one case will be prevented by the exposure or intervention.

It is equivalent to the reciprocal of the population attributable fraction (AF_{pop}).

In the example of drunk driving and driving related deaths ([table - death from drunk driving](#)), 22% of driving related deaths in the population could be attributed to drunk driving. We thus have:

$$CIN = 1 / 0.22 = 4.5$$

This means that, for every 4.5 driving related deaths, one is attributable to drunk driving on average.

Exposure impact number (EIN)

This is the number of people with the exposure among whom one excess case is attributable to the exposure.

It is equivalent to the reciprocal of the attributable risk in the exposed (AR_e).

In the example of drunk driving and driving related deaths ([table - death from speeding or drunk driving](#)), there were 136 deaths per 1,000 drunk drivers in one year attributable to drunk driving. We thus have:

$$EIN = 1 / (136 / 1,000) = 7.4$$

This means that, for every 7.4 drunk drivers, there is one driving related death attributable to drunk driving on average per year.

Exposed cases impact number (ECIN)

This is the number of exposed cases among whom one case is attributable to the exposure.

It is equivalent to the reciprocal of the attributable fraction in the exposed (AF_e).

In the example of drunk driving and driving related deaths ([table - death from speeding or drunk driving](#)), 91% of

driving related deaths among drunk drivers could be attributed to drunk driving. We thus have:

$$ECIN = 1/0.91 = 1.1$$

This means that, for every 1.1 drunk drivers with a driving related death, one driving related death is attributable to drunk driving on average per year.

Population impact number of eliminating a risk factor (PIN-ER-t)

This is derived from the population attributable fraction. It is calculated by multiplying the population size (n) by the risk of an event in the next t years (I_{pop}) and by the population attributable fraction (AF_{pop}) [3].

$$PIN\ ER\ t = n \times I_{pop} \times AF_{pop}$$

n = population size

I_{pop} = incidence in population (over t years)

AF_{pop} = attributable fraction in population

In the example of drunk driving and driving related deaths ([table - death from drunk driving](#)), there were 10,000 drivers in the study (n), 18 deaths per 1,000 drivers in one year (I_{pop}), and 22% of driving related deaths in the population could be attributed to drunk driving (AF_{pop}). We thus have:

$$PIN\ ER\ t = 10,000 \times 0.018 \times 0.022 = 40$$

This means that up to 40 (of 180) driving related deaths per year in Anystate could potentially be prevented by eliminating drunk driving.

References

1. Heller RF, Dobson AJ. Disease impact number and population impact number: population perspectives to measures of risk and benefit. *Br Med J* 2000;**321**:905-2.
2. Heller RF, Dobson AJ, Attia J, Page J. Impact numbers: measures of risk factor impact on the whole population from case-control and cohort studies. *J Epidemiol Community Health* 2002;**56**:606-10.
3. Heller RF, Buchan I, Edwards R, Lyratzopoulos G, McElduff P, St Leger S. Communicating risks at the population level: application of population impact numbers. *Br Med J* 2003;**327**:1162-5.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Various Levels of Case Definition

Last modified at 9/27/2011 9:49 PM by Arnold Bosman

When facing a new disease or when at the beginning of an investigation it is usual to design different level of case definition from very sensitive (suspect case) to very specific (laboratory confirmed case). Usually cases are temporarily classified as suspect until laboratory results are known.

During an outbreak of botulism in Italy in 2004 two levels of case definition were developed as follows:

A 'probable botulism case' was defined as:

- a person who had dined at restaurant A on February 22 or 24, 2004

and

- had experienced diplopia, **or** blurred vision, **or** fixed/dilated pupils

and

- at least one of the following symptoms: dysphagia, dry mouth, dysarthria, upper/lower extremity weakness, dyspnoea, or severe constipation.

Those who met the **probable case** definition **and** had **laboratory-confirmed** botulism were considered **definite cases**.

With the sophistication of laboratory methods and availability of many diagnostic tests for a single disease some case definition may become lengthy and resemble more a complex decision tree as for example the WHO SARS case definition. The following is one of the WHO regularly updated case definition during the SARS epidemic in the west pacific WHO region in 2003.

Suspect SARS case: definition for public health surveillance

1. A person presenting after 1 February 2003 with a history of:

- high fever (>38° C);

AND

- cough, or breathing difficulty
- AND one or more of the following exposures during the 10 days prior to onset of symptoms:

AND one or more of the following:

1. close contact* with a person who is a suspect or probable case of SARS
2. history of travel, to an area with recent local transmission of SARS **
3. residing in an area with recent local transmission of SARS

2. A person with an unexplained acute respiratory illness resulting in death after 1 November 2002, but on whom no autopsy has been performed AND one or more of the following exposures during to 10 days prior to onset of symptoms:

1. close contact, * with a person who is a suspect or probable case of SARS;
2. history of travel to an area with recent local transmission of SARS
3. residing in an area with recent local transmission of SARS

** Close contact means having cared for, having lived with, or had direct contact with respiratory secretions or body fluids of a suspect or probable case of SARS.*

*** Areas in which there are reported foci of transmission of SARS are updated on the WHO website (<http://www.who.int/csr/sarsareas/en/>). Current areas (1 April 2003) with reported foci of transmission are:*

Probable SARS case: definition for public health surveillance

A suspect case with radiographic evidence of infiltrates consistent with pneumonia or respiratory distress syndrome (RDS) on chest X-ray (CXR)

OR

A suspect case of SARS that is positive for SARS coronavirus by one or more assays.

OR

A suspect case with autopsy findings consistent with the pathology of RDS without an identifiable cause.

Exclusion criteria: A case should be excluded if an alternative diagnosis can fully explain their illness.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Case to case study design

Last modified at 9/14/2011 2:17 PM by CeRC

Case to case studies are types of case control studies used when the disease of interest can be sub-classified in two or several groups that have specific risk factors. In a case to case study cases with a particular sub-type of a disease are compared to cases with another subtype. For example during a listeriosis outbreak, cases with the outbreak sub-type would be compared to sporadic cases (the controls).

Some assumptions are made. Non outbreak cases (the controls), if infected with the outbreak subtype would have been classified as cases. They come from the same population which gave rise to outbreak cases. They represent exposure (e.g. food consumption) in the source population for outbreak cases. This is probably the major issue. Are sporadic cases of listeriosis representing food consumption in the general population? This may not always be true. Non epidemic cases may be more likely to be exposed than the overall source population. We may therefore underestimate the odds ratio.

Some advantages lie with case to case studies. Cases are readily available. Since all subjects in the study are sick there also may be less differential recall between cases and controls.

Case to case studies may be a convenient design when information is available for the sub class of cases used as controls. However, as in any case control study, investigators need to be very cautious and verify that exposure in the control group reflects accurately exposure in the source population for cases.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Case cross over studies

Last modified at 10/25/2011 6:40 PM by Arnold Bosman

Among [cohort designs](#), cross over studies are intervention studies in which the same group of people is exposed to two different interventions in two separate periods of time. This requires that the effect of the intervention is short enough not to impact on the effect of the second intervention and that a time gap between the two interventions is respected. [\[1\]](#)

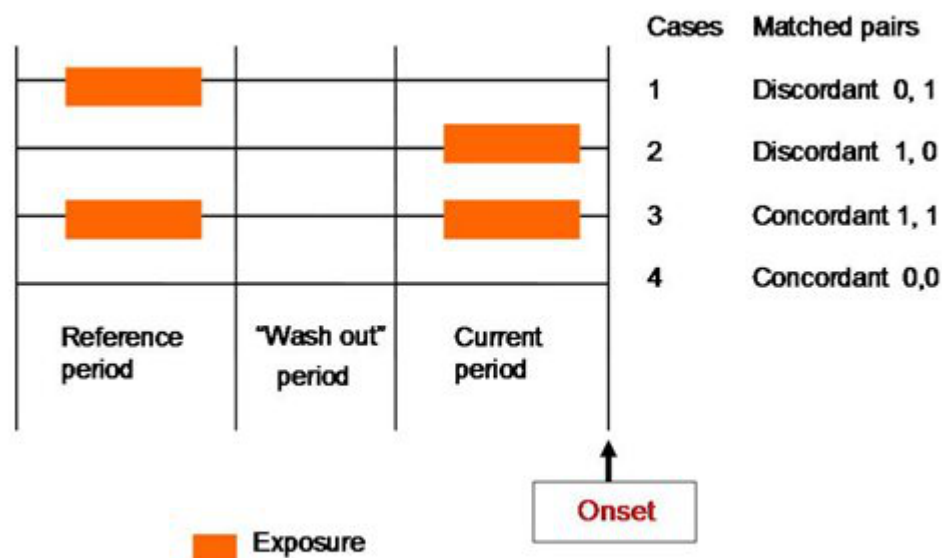
Case cross over studies are the case control version of crossover studies. This concept was introduced by Maclure et al. [\[2\]](#) [\[3\]](#) In a case cross over design all subjects are cases and exposure is measured in two different periods of time. The general principle is to find an answer to the question: "Was the case-patient doing anything peculiar and unusual just before disease onset?" or "Did the patient do anything unusual compared to his routine?". The assumption is that if there are triggering events, these events should occur more frequently immediately prior to disease onset than at any similar period distant from disease onset.

In case cross over studies, instead of obtaining information from two groups (cases and controls), the exposure information is obtained from the same case group but during two different periods of time. In the first period exposure is measured immediately before disease onset. In the second period exposure is measured at an earlier time (supposed to represent background exposure in the same person). Exposure among cases just prior disease onset is then compared to exposure among the same cases at an earlier time. Each case and its matched control (himself) are therefore automatically matched on many characteristics (age, sex, socio economic status, etc.)

To illustrate that point Maclure used the following example. Let suppose we study the role of heavy physical activity in the occurrence of myocardial infraction (MI). Using a case cross over design we could document exposure to heavy physical activity among cases in the hour immediately preceding MI. We would then document exposure to heavy physical activity among those same cases at another earlier time.

The following figure illustrates periods of exposures taken into account in a case cross over study.

Source: Adapted from Jean Claude Desenclos, InVS, France



In the above figure the period immediately before onset is called the « current » period and the other period "the reference period". The two periods are separated by a "wash out period" in order to avoid that exposure in the reference period is mixed with exposure in the current period. The reference period of exposure is used to reflect average exposure experience among cases. Case 1 was unexposed in current period (just prior to onset) and exposed in the reference period. Case 2 was exposed just prior onset and unexposed in the reference period. Case 3 was exposed in both periods and case 4 in none.

From the above we should consider that the same case and its 2 periods of exposure constitute a matched pair. Cases 1 and 2 are discordant pairs and cases 3 and 4 concordant. This is why with a case cross over design a matched pair analysis is required. Only discordant matched pairs will be used in the analysis (see chapter on [matching](#) for rational).

In addition some characteristics of exposure and outcome are noteworthy.

Exposure should change over time in the same person and over short period of time.

Exposure should not be changing in a systematic way over time. In the example of physical activity let's suppose we have documented exposure in the hour immediately before onset and that we have documented reference exposure two days before at the same time. This would not be appropriate if physical activity occurs in a systematic timing (every second day at the same time).

Exposure should have a short term effect. Duration of exposure effect should be shorter than average time between two routine exposures in the same individual. The effect of a first exposure should have stopped before the next exposure.

Induction time between exposure and outcome should be short.

Disease must have an abrupt onset. Case cross over are not appropriate if the exact date/time of onset is not available or if abrupt onset does not exist (some chronic disease).

Several reference time periods can be used to document average exposure among cases. In that instance, an average of time being exposed is computed and compared to exposure just prior disease onset. The efficiency of the case cross over method increases with the number of reference periods included.

As in any [case control study](#) the capacity to properly document exposure should be identical in the two periods of time. In case cross over designs information biases are a sensitive issue.

Even if [confounding](#) is controlled since a case is its own control, within-person confounding can occur. In the example of heavy physical activity and MI, another factor (anger) may be linked both to exposure (heavy physical activity) and outcome (MI).

Case cross over and food borne outbreaks.

Case cross over design was sometime used by epidemiologists to try to identify a food item as the vehicle for a food borne disease outbreak. Several of the above listed points merit to be challenged. A recall (exposure) period of around three days may be too large to use this design. In addition food habits (average exposure) do not happen randomly in an individual. Finally, comparing consumption of a potentially infected food item in the "current" period to average consumption of a similar un-infected food item in the reference period does not relate to the same exposure.

Consumption of a food item could be identical in the current and reference time periods and still only the food item in the current period was contaminated.

References

1. Rothman KJ. Epidemiology. An Introduction. New York, Oxford University Press 2002.
2. Maclure MA, Maclure M, Robins MJ. The case-crossover design: a method for studying transient effects on the risk of acute events. American Journal of Epidemiology 1991; 133(2):144-153
3. Maclure, M, Mittleman MA. Should we use a case-crossover design? Annu. Rev. Public Health. 2000; 21:193-221



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Choosing a Reference Group

Last modified at 9/14/2011 3:54 PM by CeRC

Making comparisons is fundamental to epidemiological investigations and studies. We need to compare [risk or rates of illness in exposed and unexposed group](#), or [odds of exposure in cases and controls](#). Without making comparisons with a reference group, we cannot say from data analysis that an [association](#) with a given outcome is anything other than spurious. Such a reference group is designated as the control group in case control studies and the unexposed group in cohort studies. For the field epidemiologist, difficulties more often arise in choosing controls for case control studies than in choosing an [unexposed group in cohort studies](#). This section will focus mainly on the former.

In order to [define a control group](#), it is helpful to be clear about who the cases are, in other words, to start with a case definition. The case definition then helps to define the population from which the cases arise, the source population. This population is also the population from which controls should be drawn. The most important principle to follow is that controls should be representative of the source population. Cases can be defined in any way that the investigator decides, but this definition is key to determining the source population of cases, and hence the source population of controls.

There are many ways of [choosing controls](#). This section reviews some of the more common types of controls, their advantages and disadvantages.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Descriptive data analysis

Last modified at 5/21/2013 4:35 PM by testuser1

This part of the FEMWIKI deals with basic data analysis to describe population data (e.g. surveillance data) in [time](#), [place](#) and [person](#).

This chapter in this section is incomplete, and more extensive content is welcome.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

The idea of Statistical Inference

Last modified at 9/15/2011 10:14 PM by Arnold Bosman

The general idea of statistical inference, is to find out a certain "truth" about a population, by investigating a **sample**, rather than the entire population. The investigation can be **descriptive** (for example to find out the *true occurrence* of a disease) or **analytical** (for example to *test the hypothesis* that people who have eaten home preserved green olives are more at risk of developing botulism than those who did not eat those olives).

Statistical Inference is the process of drawing conclusions about the entire population, based on the investigation of a sample. So it is a form of *generalisation*.

This process differs from [causal inference](#), which is explained elsewhere.

Significance tests

In order to make the conclusions *objective*, statistical tests are usually applied, with the aim to reach a **decision** ('yes' or 'no') on a difference (or 'effect'), on a probabilistic basis, on observed data. Such statistical tests are also called significance tests, which all have in common that they require a **Null Hypothesis** (H0): "There is no difference (no effect) between the groups that we compare".

A Null Hypothesis (H0) will always have a complementary **Alternative Hypothesis** (H1): "There is a difference between the groups that we compare" (in other words: **the Null Hypothesis is not true**).

The aim of a significance test is to help us **decide** to reject the Null Hypothesis or not.

In our example, we could write the Null Hypothesis like this:

"There is no difference in occurrence of botulism in the population between the people that have eaten home preserved green olives (=exposed) and those that did not (=unexposed)".

Such hypothesis makes it easier to design a study to test this: we need to take a representative sample of the people

that were exposed and a representative sample of those who were unexposed. In both samples we measure the occurrence of botulism, and we compare the results.

The next challenge is: how different do the results need to be to make us decide to reject the H_0 ?

This is the point where the p-value will help our decision. This value will tell us what is the probability (p) to find the difference that we have observed (between our samples) if the Null Hypothesis H_0 is true. The lower this p-value, the lower the probability that chance alone can explain the difference between the results in our samples when there really is no difference in the total population.

This requires that we investigate and quantify the [probability to be different from the expected](#).

Making a decision on H_0 .

If we have convinced ourselves that the occurrence of botulism is significantly different between the exposed (who ate olives) and the non-exposed, then we can decide to reject the Null Hypothesis.

Now in taking a decision on H_0 , we can make two possible errors:

- The null hypothesis is true but rejected: Type I error (α -error)
- The alternative hypothesis is true, but the null hypothesis is not rejected: Type II error (β -error)

Please note that statistical tests only allow us to decide to reject H_0 or not to reject. This is different from deciding to accept H_0 , or accept H_1 .

Problems in applying significance tests in observational studies

In these examples we have applied significance tests to an observational study: an outbreak has occurred within a population at risk (guests in a restaurant) and retrospectively we test hypotheses on data observed from events that took place before we formulated the hypotheses.

One of the criticisms often given regarding the interpretation of such epidemiological studies is that no random assignment of subjects to groups (exposed, non-exposed) took place. The aim of randomisation is to get an equal distribution of other risk factors which have not been measured (or even discovered). The gold standard for such studies is the randomised controlled trial, preferably where the investigators and subjects are blinded to the assignment in exposed and unexposed.

In such designs where everything except the exposure of interest is randomised, the significance tests produces a p-value that truly reflects the probability that chance produced the differences in results between study groups.

In observational studies, we have to be aware that we observe 'experiments of nature' (such as outbreaks) where the assignment of people to exposed and non-exposed is rarely a fully random process. For this reason, many critics say that the p-value in such circumstances should be considered to have a descriptive nature and caution should be exercised in case of statistical inference.

Part of this problem is related to concepts of [bias](#) and [confounding](#).



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Measuring incidence rates

Last modified at 10/28/2010 8:40 PM by Lisa Lazareck

Cohort studies measuring incidence rates

The computation of effects with incidence rates is similar to calculation of effects from incidence proportions (risk).

The incidence rate of disease in exposed (IRE) and unexposed (IRu) can be computed as follows:

$$IRE = \frac{\text{number of cases exposed}}{\text{sum of person-time at risk among exposed population}}$$

$$IRu = \frac{\text{number of cases unexposed}}{\text{sum of person-time at risk among unexposed population}}$$

A rate difference can be computed: **Rate difference** = $IRE - IRu$

The relative effect of the exposure on disease occurrence can be measured by computing the rate ratio minus 1.

$$\text{Relative effect} = \frac{IRE}{IRu} - 1$$

The rate ratio is:

$$\text{Rate ratio} = \frac{IRE}{IRu}$$

Example

Breast cancer cases and person-years of observation for women with tuberculosis repeatedly exposed to multiple x-ray fluoroscopies and unexposed women with tuberculosis

Radiation exposure	Person-years	Breast cancer	Rate/10000 p-y	Rate ratio	Rate difference	Relative effect
Yes	28010	14	14.6	1.86	6.7	0.86
NO	19017	15	7.9			

Source: Boice & Monson [\[1\]](#)

One can express the result by saying that the relative effect is 0.86 which would suggest an 86 % increased rate of breast cancer among exposed. One can also express the results by saying that the rate of breast cancer is 1.86 times higher in the exposed cohort than in the unexposed cohort.

References

1. Boice, J. D., and R. R. Monson. Breast cancer in women after repeated fluoroscopic examinations of the chest. *Journal of the National Cancer Institute* 1977 59: 823–832.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Measures of Disease Occurrence

Last modified at 10/8/2017 10:04 PM by Arnold Bosman

Learning Objectives:

After reading this chapter, you will be better able to:

- appreciate the definition of risk, rate, prevalence and odds;
- appreciate the difference between measures of disease occurrence and measures of effect;
- familiarize with the different terms and synonyms which are used to describe risk, rates, prevalence;
- understand that the epidemiological jargon is not always correct (an attack rate is actually a risk);
- identify what is the best measure to calculate in different study designs.

Disease occurrence

The measures to be used depend on the study design, but also on what we want to measure. Measures of disease occurrence are used when we are interested only in quantifying an event (the outcome), and our analysis does not extend further to take into account exposures. When we want to relate the effect of a certain exposure to an outcome, we will then need to use what we call [measures of effects](#).

A [risk](#) represents a proportion of the number of people developing the disease divided by the number of people in the population. It can be presented as a proportion (ranging from 0 to 1) or as a percentage (ranging from 0% to 100%), and expresses the probability of an outcome (health event, disease etc) in a certain group. Although time units are not expressed, the concept of risk implies that we are observing a population for a specific time period. Risks may also be expressed per 10,000 or per 100,000 population; this is sometimes called '**cumulative incidence**', or simply '**incidence**'.

[Rates](#) are used when the occurrence of an event relates to units of time (for example the number of deaths per 100 persons-years). When measuring the occurrence of a new event in relation to units of time, we refer to [incidence rates](#).

In cohort studies performed during an outbreak investigation, [attack rates](#) are often calculated to have a measure of the proportion of people who experience the outcome of the study. Indeed, attack rates (as well as [case fatality rates or ratio](#)) are risks and are an example on how epidemiological jargon might be misleading. Though they are called rates, they are proportions because they do not relate to units of time.

Sometimes it might be convenient to approximate [rates into risks](#), as the interpretation of risks is easier.

Prevalence is a proportion of how many events (for example, people with disease) are present at a specific point in time in a population. It is expressed as a percentage.

The [odds](#) of an event ("odds", always plural) are the probability that this event will occur divided by the probability that the event will not occur. Therefore, a value of the odds of the event occurring can range from 0 to infinity. Odds are a measure rarely used, though the ratio of two odds ([odds ratio](#)) are one of the most commonly used [measures of effect](#) in epidemiology.

Whenever using any of these measures it is important to consider the context in which they are used and whether or not they properly express what we want to measure.

EPIET Lectures:

[Measures of Disease Occurrence](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Relation between risk and incidence rate

Last modified at 9/14/2011 2:06 PM by CeRC

It is easier to interpret [risks](#) than [rates](#). Therefore, it might sometimes be convenient to convert an incidence rate into a risk using the following formula [\[1\]](#):

Risk = Rate x Time

Suppose that we have a population of 1000 persons in which the incidence rate of cancer X is 6 cases per 1000 person-years ($6 / 1000 \text{ yr}^{-1}$). If we follow this population for 30 years the risk of cancer X in the population over that 30 years is: $6 / 1000 \text{ yr}^{-1} \times 30 \text{ years} = 0.18$ or 18%. Among the 1000 persons present at the start of the follow-up, 180 cases of cancer X will occur. If the follow up was 15 years the related risk would be 9 %. The above formula does not take into account the decrease of the population at risk over time and cannot be used when risk is large. It also assumes that rate remains constant over time.

References

-
1. Rothman KJ; Epidemiology: an introduction. Oxford University Press 2002, p.33-38.

Traditional case-control studies

Last modified at 3/1/2015 9:21 AM by Arnold Bosman

Traditional case-control studies

Let suppose that we are at the end of the follow up period and have respectively C_e and C_u cases, and $Ne - C_e$ and $Nu - C_u$ persons still free of disease (non cases), in the two cohorts. If the disease is rare it is obvious that persons free of disease at the end of the study period reflect the exposure experience of the source population. If the disease is frequent, exposure among persons free of disease at the end of the study may be lower than in the source population (since exposure increases the risk of disease).

If the disease is rare we can use a sample of non cases at the end of the study period to estimate the risk ratio. Using non cases to estimate the source population exposure experience is the principle of traditional case control studies.

Let's call "c" and "d" respectively the number of exposed and unexposed in the sample. If sampling is done independently from the exposure status we would expect **if the disease is rare**

$$\frac{c}{d} = \frac{Ne - C_e}{Nu - C_u} = \frac{Ne}{Nu}$$

or equivalently

$$\frac{c}{Ne - C_e} = \frac{d}{Nu - C_u}$$

If the above is true the risk ratio estimated from a traditional case control study can be represented as:

$$\frac{I_{Pe}}{I_{Pu}} = \frac{\frac{a}{Ne - C_e}}{\frac{b}{Nu - C_u}} = \frac{a}{b} \times \frac{Nu - C_u}{Ne - C_e} = \frac{a}{b} \times \frac{d}{c}$$

The quantity ad/bc is the odds ratio. It represents the ratio of the odds of disease among exposed divided by the

odds of disease among unexposed.

However if the disease is not rare a large part of Ne/Nu is represented by future cases who are more likely to be exposed than non cases. Consequently, the odds ratio may dramatically overestimate the risk ratio.

$$\frac{Ne - Ce}{Nu - Cu} \text{ may not be equal to } \frac{Ne}{Nu}$$

To illustrate this point let's now move to the example of a food borne outbreak in a nursing home with 200 residents and 74 cases of gastroenteritis. The epidemic curve is consistent with a point common source of infection and example 4 shows the results of a retrospective cohort study. It suggests that the risk of gastroenteritis is 3.4 times higher among residents who consumed a specific food item compared to those who did not

Example 1: Occurrence of gastroenteritis among residents of nursing home A according to consumption of a specific food item.

Specific food item	Total	Cases	Risk	Risk ratio
Yes	60	44	73.3%	3.4
No	140	30	21.4%	Reference
Total	200	74	37.5%	

Let's suppose investigators would have preferred to conduct a traditional case control study (case – non cases study) rather than a retrospective cohort. In a traditional case control study controls are selected from people who are free of the disease at the end of the study period. The OR is a good estimate of the risk ratio if the disease is rare.

Example 2: Consumption of a specific food item among cases and various samples of residents of a nursing home

Consumption	Cases	50% sample of non cases	OR	50% sample of source population	RR
Yes	44	8	10.1	30	3.4
NO	30	55		70	

Using as controls a 50% sample of the non cases the odds ratio equals 10.1, overestimating the risk ratio by a factor of

three. This should not come as a surprise, though. When selecting controls from non-cases, and since the disease is frequent (the overall risk of gastroenteritis is 37.5%), the control group is no longer representing the distribution of exposure in the source population. The frequency of exposure in the control group selected from non cases is 7.3% and was 30% in the source population.

If instead we had done a [case cohort study](#) and chosen a 50% random sample of the source population, the sample (if unbiased and ignoring random variation) would be likely to provide the same proportion of exposed (30%) than in the source population. The risk ratio obtained (3.4) would again be similar to the risk ratio observed in the [cohort study](#).

When to use a traditional case control study?

Traditional case control studies are an easy and very convenient way to conduct epidemiological studies when the disease is rare. Because of its simplicity it is the most popular method. It has been extremely useful to epidemiologists in the past 50 years. Provided that the disease is rare the odds ratio provides a good estimate of the risk ratio. However, it should not be used when disease incidence is high. This particularly applies to investigations of food borne outbreaks with very high incidence.

NB. "Traditional case-control studies" are a type of case-control studies, where controls are simply non-cases. For this reason, you may find them quoted as "case-non-case studies" in literature.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Case-cohort study

Last modified at 10/17/2014 9:54 AM by Ioannis Karagiannis

Case cohort studies

In case-cohort studies, we aim to achieve the same goal as in cohort studies, but more efficiently, using a sample of the denominators of the exposed and unexposed cohorts [1]. Properly conducted case-cohort studies provide information that should mirror what could have been learned from a cohort study.

We will call "source population" the population which gives rise to cases. The source population includes exposed and unexposed cohorts and in that source population we could have conducted a cohort study comparing risk or rates of disease between exposed and unexposed cohorts.

If, instead, we decide to do a case-cohort study, we will include the same cases and classify them as exposed or unexposed. In other words, we start by choosing the cases, which is a case-control study characteristic. Instead of getting exposure information from all individuals constituting the denominators of exposed and unexposed cohorts, which would have been a cohort study characteristic, we only use a sample of them. The purpose of this sample is to estimate the relative size of exposed and unexposed components of the source population (the proportion of exposed in the source population at the beginning of the cohort).

To do so, we select a random sample from the entire source population. If that sample is unbiased (sampling done independently from exposure status) we expect (disregarding sampling variation) the distribution of exposed and unexposed persons in the sample to reflect the exposure distribution in the source population at the beginning of the cohort. This is an important aspect of case-cohort studies. The sample should be representative of the population giving rise to cases (the source population) regarding exposure.

One way to imagine case-cohort studies is therefore to think of them as nested within cohorts of exposed and unexposed people. Any case cohort study could be thought off as nested from the source population. The sample group (control group) is a sample of the denominator present at the beginning of the cohort.

From a cohort study measuring risk of disease in exposed and unexposed cohorts we can draw the following results table:

Table 1

Exposure	cases	Population at risk	IP	Risk ratio
Yes	a	Ne	a/Ne	a/Ne / c/Nu
No	b	Nu	c/Nu	

If, instead of studying the entire denominators of exposed and unexposed, we were sampling them (let's say 10%) we would have the following table:

Table 2

Exposure	Cases	Sample from source population
Yes	a	Ne/10
No	b	Nu/10

Obviously, the risk of disease cannot be computed from the above table, since denominators sampled from exposed and unexposed cohorts are only a sampling fraction of these two populations. However, if risk can no longer be computed for exposed and unexposed, the risk ratio remains the same. If in the risk ratio calculation we replace the denominators by the 10% samples representing them, we obtain the same value for the risk ratio.

$$\text{Risk Ratio} = \frac{\frac{a}{\frac{Ne}{10}}}{\frac{b}{\frac{Nu}{10}}} = \frac{\frac{a}{Ne}}{\frac{b}{Nu}}$$

When the sample is randomly selected from the source population the risk ratio computed using the sample equals the risk ratio computed within the entire cohorts.

Since we are randomly selecting controls from the source population as it was at the beginning of the study (before disease occurrence), it may happen that persons who will later become a case will be selected as controls. Therefore some persons may appear both in the case and control groups. This should not come as a surprise. In a cohort study cases are counted in the numerator and denominators of exposed and unexposed. The same applies to case cohort studies since we use a sample of exposed and unexposed people of the source population. We are not concerned by the disease status of the control group but by its exposure status. The aim of the control group is to properly reflect the exposure in the source population and this source population originally includes people who will later become cases. Excluding future cases would lead to overestimating the risk ratio, this particularly when disease occurrence is high.

When to conduct a case cohort study?

Case-cohort studies are not very popular. Their concept is not well understood to the point that some journals would reject a case cohort study on the reason that the control group includes cases. Case cohort studies are a very suitable design when disease incidence is high. They provide a direct estimate of the risk ratio. They are not suited when exposure changes over time (if exposure is measured at the beginning of a follow up period and differs from the overall exposure experience during the entire study period).

NB. Case-cohort studies are a type of case-control studies, where controls are simply representative of the source population in terms of exposure (as controls should always be). In literature, you may find "case-cohort studies" quoted as "case-control studies" and "traditional case-control studies" quoted as "case-non-case studies", since, in the latter, controls are actually non cases.

References

1. Rothman KJ Epidemiology. An introduction. Oxford University Press, New York, 2002.
2. Le Polain de Waroux O, Maguire H, Moren A. The case-cohort design in outbreak investigations. Euro Surveill. 2012;17(25):pii=20202. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20202>

Density case control studies

Last modified at 10/28/2010 8:36 PM by Lisa Lazareck

Density case control studies

In cohort studies, incidence rates are sometimes called incidence density rates. By similarity we will call "density based sampling" a sampling method in which the sample used as controls will represent the person time experience of exposed (P_{Te}) and unexposed (P_{Tu}) cohorts in the source population. [1] Thus the probability of any person from the source population to be selected in the sample is proportional to his/her person-time contribution to the denominators of the incidence rates in exposed and unexposed cohorts.

Incidence (density) rates in exposed and unexposed cohorts of the source population can be expressed as follows:

$$IR_e = \frac{a}{PT_e}$$

$$IR_u = \frac{a}{PT_u}$$

Where "a" is the number of cases exposed, "b" the number of cases unexposed, "P_{Te}" the total person-time accumulated by exposed persons and "P_{Tu}" the total person-time accumulated by the unexposed group.

If instead of studying the entire denominators of person time being exposed and unexposed we were sampling them (let's say 10%) we would have the following table:

Table 1

Exposure	Cases	Sample from source population
Yes	a	P _{Te} /10
No	b	P _{Tu} /10

Obviously from the above table the incidence rate of disease cannot be computed since person time denominators sampled from exposed and unexposed are only a sampling fraction of these two populations. However, if incidence rate can no longer be computed for exposed and unexposed, the rate ratio remains the same. If in the rate ratio calculation we replace the person time denominators by the 10% samples representing them, we obtain the same value for the rate ratio.

$$\text{Rate Ratio} = \frac{\frac{a}{\frac{P_{Te}}{10}}}{\frac{b}{\frac{P_{Tu}}{10}}} = \frac{\frac{a}{P_{Te}}}{\frac{b}{P_{Tu}}}$$

In this study design the sample (control group) is randomly selected from the person time experience of the source population. As a consequence the rate ratio computed using this sample is equal to the rate ratio computed within the cohort study done with the entire person time denominators of the source population.

The next issue is obviously about how to select a sample and make sure it represents the person-time experience of the exposed and unexposed cohorts in the source population. It is in fact quite simple. Each time a case occurs, an individual (or several) is randomly selected from the source population which is still free of the disease at the time of the case onset. This is sometimes called prospective case control study. A mathematical explanation of this rational can be found in Rodrigues *et al.* [2]

For each person contributing time in the source population experience, the time that this person is eligible to be selected in the sample is the same time during which she is also eligible to become a case if the disease should occur. Selecting an individual at the time of disease onset in a case leads us to select a sample among people still at risk and therefore proportionally to their time participation so far in the study. People who have left, are dead or who are already cases cannot be selected from that time on. This is also meaning that a selected individual who is still at risk of disease can later become a case in the study.

Let us suppose that Boise and Monson [3] had decided to do a density case control study instead of a person-time cohort study. They would have identified the 56 cases that occurred in the two cohorts and selected a sample series of 470 women. The sample series group should be sampled so that the person time distribution of the sample mirrors the person time distribution of the source population. If randomly selected and unbiased, this would give us 280 exposed and 190 unexposed in the sample (59.6 % of the sample is exposed which is equal to the proportion of exposed person time in the source population, 28010 / 47027).

Example

Cases and sample selected from breast cancer cases and person-years of observation for women with tuberculosis repeatedly exposed to multiple x-ray fluoroscopies and unexposed women with tuberculosis

Radiation exposure	Person-years	Breast cancer	Rate/10000 p-y	Rate ratio source population	Sample	Rate ratio sample
Yes	28010	41	14.6	1.86	280	1.86

No	19017	15	7.9		190	
Total	47027	56	11.9			

Source: Boice & Monson [3]

The controls represent person years at risk experience among exposed and unexposed. Controls are selected concurrently from those still at risk when a case occur. A person selected as a control can later become a case, the opposite is not possible since a case is no longer at risk of developing disease (with a non recurrent disease). A control which later becomes a case is kept in both groups.

Using density sampling allows us to compute a rate ratio which is equal to the rate ratio we would have obtained if a cohort study had been conducted to compare rates between exposed and unexposed cohorts in the source population. Density case control studies require an analysis matched on time of disease onset and control selection.

When to conduct a density case control study?

Density case control studies are suited for estimation of rate ratios (incidence density rate ratios). They are simple to conduct. In fact, to select a control among persons still free of disease, at the time a case occurs, is common practice frequently called prospective case control study. It provides a good estimate of the rate ratio. Density case control studies are suited when unequal length of follow up occurs for study members.

References

1. Rothman KJ. Epidemiology. An Introduction. New York. Oxford University Press 2002
2. Rodrigues L. Kirkwood BR. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. Int J Epidemiol 1990 Mar;19(1):205-13
3. Boice, J. D., and R. R. Monson. *** cancer in women after repeated fluoroscopic examinations of the chest. *Journal of the National Cancer Institute* 1977 59: 823–832.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Effect Modification

Last modified at 9/15/2011 10:22 PM by Arnold Bosman

If the magnitude of the [risk ratio](#), [rate ratio](#), [odds ratio](#) or [risk difference](#) varies in different sub groups (strata) of the study population, there is effect modification. This differs from confounding, where we generally believe that the measure of effect (i.e. the RR, OR etc.) will be the same in each of the strata defined by levels of the confounding variable. Where the measure of effect does differ by the effect modifying variable, it is unreasonable to combine the results from the different strata (as is the case using the Mantel-Haenzel methods for a confounding variable).

Cohort study (hypothetical)

In the hypothetical cohort study carried out to measure the effectiveness of a vaccine on preventing occurrence of disease X; vaccine effectiveness (VE) can be derived from the risk ratio (RR) using the formula: $VE = (1 - RR) * 100$ (to express VE as a percentage).

Vaccination	Denominator	Cases	Risk/1000	Risk Ratio	VE (%)
Yes	301545	150	0.50	0.27	71
No	298655	515	1.72		

The risk ratio for the entire cohort of 0.29 implies a VE of 71%. But we also looked at VE across different age groups of the population. Within each age group, a risk ratio comparing the risk of disease X between the vaccinated and unvaccinated was computed.

Age Group	Vaccination	Denominator	Cases	Risk/1000	Risk Ratio	VE (%)
< 1 year	Yes	35625	38	1.07	0.87	13

	No	24375	30	1.23		
1 - 4 years	Yes	44220	34	0.77	0.42	58
	No	46780	86	1.84		
5 - 9 years	Yes	78200	50	0.64	0.19	81
	No	75000	250	3.33		
10 - 24 years	Yes	83400	18	0.22	0.15	85
	No	82600	120	1.45		
> 24 years	Yes	60100	10	0.17	0.40	60
	No	69900	29	0.41		

We observe that the risk ratio ranges from 0.15 to 0.87 according to various age groups, consequently, neither is VE equal for the various age groups. This suggests that age is modifying the protective effect of the vaccine. Age is called an effect modifier.

Since the data suggest different vaccine effectiveness by age group it would not be logical to summarise the table and give only an overall vaccine effectiveness (e.g. 71%). It is important to describe the VE by age groups. When effect modification is suggested by the data, it is important to present stratum specific results that provide more information than an overall effect.

Case control study

The same reasoning can be applied to a case control study. A case control study conducted in France in 1995 suggest that storing eggs for longer than 2 weeks in the home increases the risk of gastroenteritis (OR = 3.8) in children [1]. However if the analysis is stratified in two seasons, summer and others, the odds ratio is higher in summer (OR = 6) than in other seasons (OR = 2.3), suggesting that the increased risk of gastroenteritis with duration of home eggs storage expresses itself differently according to the season. Here, season is an effect modifier of the association between duration of storage of eggs in the home and the occurrence of *Salmonella enteritidis* gastroenteritis.

	Duration of storage	Cases	Controls	OR	95% CI

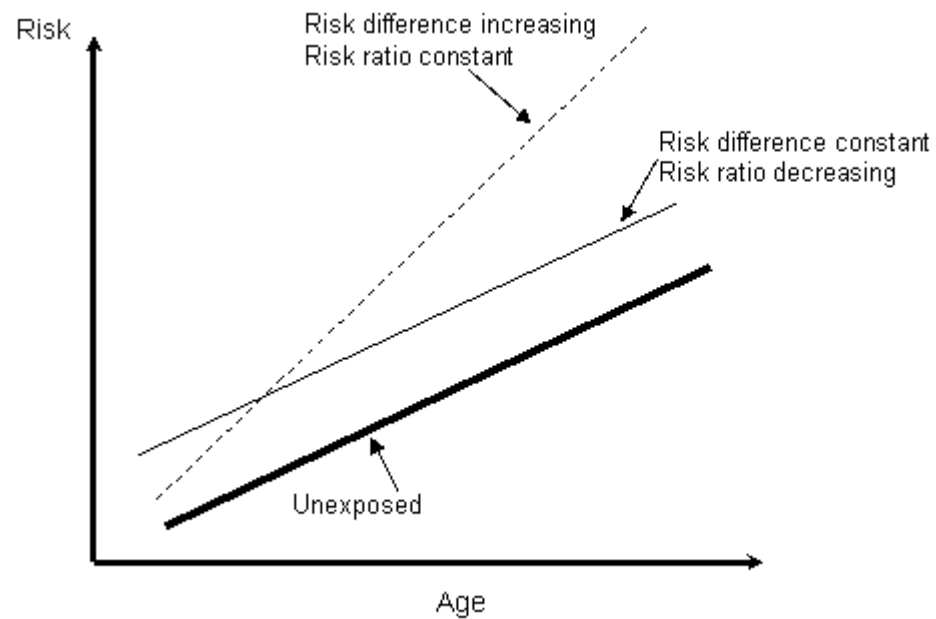
Overall	≥ 2 weeks	12	2	6	1.3 - 26.8
	< 2 weeks	52	64		
Summer	≥ 2 weeks	19	5	3.8	1.4 - 10.2
	< 2 weeks	84	100		
Seasons other than summer	≥ 2 weeks	7	3	2.3	0.6 - 9.0
	< 2 weeks	32	36		

Specific statistical methods are used to look for effect modification and test the homogeneity of stratum specific risk ratios or odds ratios. The most popular tests include the Woolf test, Breslow-Day, X^2 for trends, etc. Details of the various methods can be found in referenced books and articles [2].

Assessing risk differences between exposed and unexposed cohorts

In the two above examples, effect modification was assessed by comparing the risk ratios or odds ratios between different sub-groups (strata) of a population. However we sometimes use risk difference to identify how risk varies between exposed and unexposed cohorts.

The following example is a classic illustration of the difficulty to conclude on the presence or not of effect modification according to the type of effect measure we use (risk ratio or risk difference). In the figure, the risk of hypothetical disease X is compared between exposed and unexposed according to age. The risk increases with age linearly among unexposed (bold line). For the exposed groups two alternatives are presented. First the line representing the increase of risk with age among exposed (plain line) is parallel to that of unexposed. The risk difference (RD) is constant and the RR decreases with age. Alternatively (dotted line) if risk increases with a bigger slope among exposed, RD increases with age and RR is constant. This is why some authors would use the term effect-measure modification rather than effect modification to make sure that the type of effect measure (RR or RD) is specified [3;4]. Some also refer to "an effect modifier of the risk difference" or alternatively of the risk ratio.



Example (hypothetical)

A cohort study collects information on drinking, exposure to ceramic dust and subsequent liver cancer. The table shows the risk (over 1 year, per 100,000 persons) derived from the study.

	No ceramic dust	Ceramic dust
Drinker	10	50
Non-drinker	1	5

Among those exposed to ceramic dust, the relative risk of liver cancer between drinkers and non-drinkers is 10 (50/5). Among the unexposed, the relative risk between drinkers and non-drinkers is 10 (10/1).

The risk difference between the drinkers and non-drinkers who are not exposed to ceramic dust is $10 - 1 = 9/100,000$ persons. The risk difference between drinkers and non-drinkers who are exposed to ceramic dust is $50 - 5 = 45/100,000$ persons.

The difference in effect modification between these scales reflects statistical interaction - which refers to the deviation from the underlying model. This is different from [biological interaction](#).

References

1. Delarocque-Astagneau, E., Desenclos JC, Bouvet P, Grimont PA. Risk factors for the occurrence of sporadic *Salmonella enterica* serotype enteritidis infections in children in France: a national case-control study. *Epidemiol Infect* 1998 Dec;121(3):561-7.
2. B.R.Kirkwood, J.A.C.Sterne. *Medical Statistics*. Second ed. Massachusetts: Blackwell Science Ltd; 2003.
3. *A Dictionary of Epidemiology*. Fifth ed. New York: Oxford University Press; 2008.
4. K.J.Rothman, S.Greenland, T.L.Lash. *Modern Epidemiology*. Third ed. Philadelphia, USA: Lipincott Williams and Wilkins; 2008.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Confounding in studies

Last modified at 10/30/2011 11:30 PM by Arnold Bosman

A confounding variable is one which is associated with both the exposure and the disease. It confounds the [measured association](#) (RR or OR). A variable is NOT a confounder if it lies on the [causal pathway](#) between the exposure and disease.

Confounding should always be addressed in studies investigating [causality](#). Because the confounding variable is not evenly distributed between exposed and unexposed (in a [cohort study](#)) or between cases and controls ([case-control study](#)), and it is also a risk factor for the disease, the measured association is distorted or biased. The bias could be negative (an underestimate) or positive (an overestimate) or could even reverse the apparent direction of effect.

Cohort Study (hypothetical)

A cohort study was conducted during the investigation of an outbreak of a vaccine-preventable disease among 2000 children. Teachers had noticed that boys were more likely to be ill than girls.

One of the study objectives was to compare the risk of disease between boys and girls.

The following table illustrates the crude results. The risk of illness is 82% among boys compared to 18% among girls (risk ratio = 4.52). We would then suggest that risk of illness among boys was 4.5 times higher than among girls.

Gender	Cases	Total	Attack Rate	RR
Boys	819	1000	82%	4.52
Girls	181	1000	18%	ref

Given vaccination is likely to be a confounding variable affecting whether a person becomes ill with a vaccine preventable disease or not; the study population was divided in two strata, the vaccinated, and unvaccinated.

Stratified	Gender	Cases	Total	Attack Rate	RR
Unvaccinated	Boys	814	950	86%	1.00
	Girls	86	100	86%	ref
Vaccinated	Boys	5	50	10%	0.95
	Girls	95	900	11%	ref

In the above example the distribution of vaccinations differs between boys and girls, therefore vaccination is associated with gender (5% of boys are vaccinated compared to 90% of the girls). In addition vaccination is associated to (a protective factor for) occurrence of disease. The risk of illness is 86% among unvaccinated and 10.5% among vaccinated. However, if we only consider the unvaccinated group, there is no longer a difference in occurrence of disease between boys and girls. Likewise, among the vaccinated group, the difference in disease occurrence between the genders is negligible;

The apparent association between gender and disease was confounded by vaccination status. Stratification according to the confounding variable showed that the association between gender and disease was absent.

Case-control study

A large case-control study was conducted in Sweden, to determine whether occupational magnetic fields were associated with female breast cancer [1]. 20,400 cases of breast cancer were identified from the cancer registry, and 116,227 controls were selected randomly from the population register of all women between 1976 and 199 gainfully employed in Stockholm or Gotland County in Sweden. Exposure assessment was based on information about occupation obtained from the Swedish census, by linking a new job-exposure matrix to the occupation type.

Amount of exposure	Cases	Controls	Odds Ratio
<0.1	2939	16835	ref
0.10-0.19	11369	60859	0.93

This suggests that there is no association between levels of magnetic field exposure and development of cancer. But if a person is exposed to a greater amount of magnetic field over a longer time, then surely that would have an effect?

Age at diagnosis	Amount of Exposure	Cases	Controls	OR	95%CI	
<50yrs	<0.1	840	9222	ref		
	0.10-0.19	2833	29282	0.94	0.87	1.02
≥50yrs	<0.1	2083	7613	ref		
	0.10-0.19	8536	31577	1.01	0.96	1.07

Showing that the same proportion of cases developed breast cancer regardless of exposure levels and age at diagnosis - i.e. when controlled for age, being exposed to a higher level of magnetic field has no effect on developing breast cancer. Therefore, in this example, age is not a confounder of the relationship between magnetic field exposure and breast cancer.

Simpsons Paradox

Simpsons Paradox refers to the reversal of the direction of an association when data from several groups are combined to form a single group.

Here is the crude result from the comparison of two treatment types (A and B) on kidney stones [\[2;3\]](#).

	Cases Cured	Total cases	Percentage cured	RR
Treatment A	273	350	78%	ref
Treatment B	289	350	83%	1.06

This analysis shows that Treatment B ought to be the preferred option.

If the size of the kidney stones is a confounding variable for the effect of the treatment; the analysis must be stratified.

		Cases Cured	Total cases	Percentage cured	RR
Small stones	Treatment A	81	87	93%	1.07
	Treatment B	234	270	87%	ref
Large stones	Treatment A	192	263	73%	1.06
	Treatment B	55	80	69%	ref

--	--	--	--	--	--

For both smaller and larger kidney stones; Treatment A resulted in the higher proportion of patients cured.

Two facts are evident:

- 1. Those with small stones tend to be given Treatment B preferentially, while those with large stones are provided with Treatment A. So there are dominating proportions - the patients are not evenly distributed between treatment groups regardless of the confounding variable (size of stones). There is an association between the size of the stone and the treatment option offered.
- 2. The confounding variable has a large effect on the outcome: those with large stones, even if given the better treatment (A), will see less success than those with small stones. There is an association between the size of stone and the proportion of success.

The apparent association between treatment type and outcome is confounded by the uneven distribution treatment between the two groups and by the fact that the percentage cured differs between the two groups.

The existence of two kidney stone sizes with unequal proportion of treatments and of success confounds the measured effect.

References

1. Forssen UM, Rutqvist LE, Ahlbom A, Feychting M. Occupational Magnetic Fields and Female Breast Cancer: A Case-Control Study using Swedish Population Registers and New Exposure Data. *Am J Epidemiol* 2005 Feb 1;161(3):250-9.
2. S.A.Julious, M.A.Mullee. Confounding and Simpson's paradox. *BMJ* 309[6967], 1480. 3-12-1994.
3. C.R.Charig, D.R.Webb, S.R.Payne, J.E.A.Wickham. Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy. *BMJ* 292, 879-881. 29-3-1986.

See also the following EPIET Lectures:

[The Third Factor](#)

The Mantel Haenszel Method

Last modified at 4/3/2013 8:08 PM by Munir

The most popular method used to compute a weighted risk ratio or odds ratio is the Mantel Haenszel method, which can be used for risk ratios or rate ratios.

From the following table

		Cases	Total
Stratum 1	Exposed	a_1	Te_1
	Unexposed	c_1	Tu_1
	Total		T_1
Stratum 2	Exposed	a_2	Te_2
	Unexposed	c_2	Tu_2
	Total		T_2

The Mantel Haenszel risk ratio (RR_{MH}) can be computed as follows:

$$RR(MH) = \frac{\sum_1^i (a_i T u_i) / T_i}{\sum_1^i (c_i T e_i) / T_i}$$

In which:

a and c are the number of cases exposed and unexposed in a stratum

Te and Tu are the total number exposed and unexposed in a stratum

T is the total of a stratum

The sums \sum are calculated for the i strata.

Returning to the example of the cohort study with vaccinated girls and boys:

Crude RR

Gender	Cases	Total	Attack Rate	RR
Boys	819	1000	82%	4.52
Girls	181	1000	18%	ref

Stratified RRs

	Gender	Cases	Total	Attack Rate	RR
Unvaccinated	Boys	814	950	86%	1.00
	Girls	86	100	86%	ref
			1050		
Vaccinated	Boys	5	50	10%	0.95
	Girls	95	900	11%	ref
			950		

In our example the crude measure of effect (the risk ratio) was 4.5. The weighted measure of effect calculated with the Mantel Haenszel method is close to 1. It is obtained as follows:

$$RR_{MH} = \frac{\sum (a_i T_{ui} / T_i)}{\sum (c_i T_{ei} / T_i)} = \frac{[(814 \cdot 100) / 1050] + [(5 \cdot 900) / 950]}{[(86 \cdot 950) / 1050] + [(95 \cdot 50) / 950]} = \frac{82.2}{82.8} = 0.99$$

The relative difference between the weighted and the crude measures of effect is more than 15% ($4.5/0.99 \cdot 100 = 450\%$) therefore suggesting that, in our hypothetical study, vaccination is confounding (is a confounding factor for) the association between gender and disease. Had a stratified analysis been omitted, the data may lead to the conclusion that being a boy was a risk factor for the disease.

The adjusted RR 0.99 is presented, which concludes that this is the measure of association between gender and disease. This is different from [effect modification](#), where two RRs would be presented.

Mathematically, the adjusted estimate is a weighted average of the stratum specific measures of the risk ratio. It will therefore always lie within the range of the stratum specific measures of the effect. (i.e. in the example above; 0.99 is between the range 0.95 and 1.00 - the stratum specific RRs).

For a case control study the Mantel Haenszel odds ratio (OR_{MH}) can be computed as follows:

Stratified	Risk Factor	Cases	Controls	Totals
Stratum 1	Exposed	a_1	b_1	
	Unexposed	c_1	d_1	
				T_1
Stratum 2	Exposed	a_2	b_2	
	Unexposed	c_2	d_2	
				T_2

$$OR(MH) = \frac{\sum_1^i (a_i d_i) / T_i}{\sum_1^i (c_i b_i) / T_i}$$

In which:

a and c are the number of cases exposed and unexposed in a stratum,

b and d are the number of controls exposed and unexposed in a stratum.

T is the total for a stratum

The sums \sum are calculated for the i strata.

It can become customary to 'eyeball' the data: comparing the crude measure to the range of the stratum-specific measures. If the crude measure is not included in the range between stratum-specific measures, confounding may exist.

A watertight method for identifying confounding variables exists. It requires the construction of a causal diagram summarizing the knowledge and assumptions between all exposures, confounders and disease outcome; which is then analysed using graphical algorithms [1].

References

1. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999 Jan;10(1):37-48.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Model building strategies

Last modified at 10/8/2010 2:01 PM by Fernando Simon

The aim of model building is to select the variables which will result in the best model to explain the observed data. Model building will be based both on methods, experience and common sense. The epidemiologist, not the software package, is responsible for the analysis and model building process.

The most frequent approach to model building is to achieve the smallest model (number of variables) that still explain the data. The smallest is chosen because it is also the more stable. Another objective is also to provide the best possible control of confounding within the data set.

The selection of variables should start with a careful univariate analysis of each variable. This involves defining if the variable is best described as a dichotomous, polytomous or continuous and verifying linearity assumptions. This also involves, prior to the logistic regression analysis, doing a careful stratified analysis by the means of 2xn contingency tables. This provides a unique way to look at the data (what is in each cell of 2x2 tables, including zeros).

Once the univariate analysis is completed we will select all variables with a statistical test leading to a p-value below a predefined cut-off level. A cut-off level of $p\text{-value} < 0,25$ is often used. We should also include all variables we believe have a biological or public health importance. According to literature the use of more conservative or traditional level ($p\text{-value} < 0,05$) does not always allow for identifying all variables known to be important. One should also keep in mind that a group of variables which are not individually important in the model may play a collective role (confounding).

Several methods can be used to assess the fit of a best model. They include:

- forward or backward step by step approach monitored by the analyst,
- stepwise forward or backward (the software uses a precise algorithm to add or drop variables),
- the best subset method.

Following the achievement of the best model fit, the importance of each variable should then be verified by

comparing the crude association and the results of the model including comparison of confidence intervals and its statistical significance. The process of adding, fitting, dropping refitting continues until all variables in the model are judged either statistically or biologically important.

Once we have a model with all relevant variables we then should consider if interaction terms should be added. This implies that categories or linearity assumptions have been verified for polytomous and continuous variables.

[<<Back to Logistic regression](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Advantages and Disadvantages of Questionnaires

Last modified at 5/26/2011 5:31 PM by CeRC

Advantages

The main advantage of using questionnaires is that a large number of people can be reached relatively easily and economically. A standard questionnaire provides quantifiable answers for a research topic. These answers are relatively easy to analyse.

Disadvantages

Questionnaires are not always the best way to gather information. For example, if there is little previous information on a problem, a questionnaire may only provide limited additional insight. On one hand, the investigators may not have asked the right questions which allow new insight in the research topic. On the other hand, questions often only allow a limited choice of responses. If the right response is not among the choice of answers, the investigators will obtain little or no valid information.

Another setback of questionnaires is the varying responses to questions. Respondents sometimes misunderstand or misinterpret questions. If this is the case, it will be very hard to correct these mistakes and collect missing data in a second round.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Types of Questionnaires

Last modified at 5/26/2011 5:31 PM by CeRC

Self-administrated questionnaires

Questionnaires can be self-administrated, i.e. administrated to the potential responders by mail, E-mail or Internet. They fill in the questionnaire without the help of a member of the investigation team.

[Self-administrated questionnaires](#) have the advantages to be cheap and easy to administer. They preserve confidentiality and can be completed at the respondent's convenience without the influence of the interviewer. However, self-administrated questionnaires can result in low response rates, because people feel less motivated to respond. Questions can be misunderstood easily without the help of an interviewer. In addition, there can be a considerable time lag between the first sending of the questionnaire and the collection of all questionnaires, particularly if questionnaires are sent out by mail.

In recent years, web-based questionnaires have become increasingly popular among epidemiologists. Advantages include the increased timeliness and potentially higher response rate compared to mailed questionnaires. A web-based questionnaire is easily set up and the corresponding link can be sent to a large number of persons at low or no expense. Also, investigators using web-based applications can collect data without having to do the time-consuming data entering themselves. Web-based questionnaires are particularly useful in populations with high internet literacy such as young people or company employees. It is less useful for surveys in the general population as persons with little computer literacy are unlikely to respond. Some web-based applications are free while others need to be purchased [1, 2, 3, 4].

Interviewer-administrated questionnaires

[Interviewer-administrated questionnaires](#) can be used in face to face or telephone interviews. They can be used easily to interview less literate or illiterate people. Interviewers can help to clarify ambiguous questions and the answers are available more quickly than in a mailed questionnaire. The most important disadvantage is the bias which can be

introduced by different interviewers' perceptions and interpretations of the answers (interviewer bias). Also, in large surveys, more than one interviewer is needed to carry out all interviews, thus resulting in an increase in needed resources. Questionnaires need to be short (up to a maximum of 10 min), especially for telephone interviews. Also, telephone interviews are not the optimal setting to ask about sensitive issues such as sexual behaviour.

References

1. Survey Monkey <http://www.surveymonkey.com/>
2. LimeSurvey <http://www.limesurvey.org/>
3. QuestionPro <http://www.questionpro.com/web-based-survey-software.html>
4. QuestBack <http://www.questback.com/>



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Format of closed questions

Last modified at 5/26/2011 5:32 PM by CeRC

Closed questions can have different formats. The following formats are most widely used.

Single choice

The most frequently used format is a question which can be answered by "yes/no" or a number (age in years, number of sexual partners etc).

Did you travel to a foreign country in the last month? Please tick the right box.	
Yes	<input type="checkbox"/>
No	<input type="checkbox"/>
Don't know	<input type="checkbox"/>

Checklist

Alternatively, a checklist can be provided:

Which of the following outdoor activities did you do last week? Please tick the appropriate activities.	
Running	<input type="checkbox"/>
Walking	<input type="checkbox"/>
Hiking	<input type="checkbox"/>
Cycling	<input type="checkbox"/>
Swimming	<input type="checkbox"/>

None of the above	
-------------------	--

Rating Scale

The same information can be asked in a more detailed way in a rating scale:

Did you do use sunscreen during the following outdoor activities during the past six months? Please indicate how often you used sunscreen.				
	Always	Sometimes	Seldom	Never
Running				
Walking				
Cycling				

The rating scale could be numerical:

Do you think that information on the risk of sunburns would be useful for you? Please circle								
Not at all Useful	1	2	3	4	5	6	7	Very Useful

Or analogue:

How severe is your pain in this moment on a scale from 0-10? Please put the tick on the line	
0 (No pain) ----- -----	10 (Very severe pain)

Lickert Scale

Another option is the use of the Lickert scale to measure attitudes.

Sunburns cannot be avoided during outdoor activities.	
No, I strongly disagree	
No, I disagree quite a lot	
No, I disagree just a little	
I'm not sure about this	
Yes, I agree just a little	
Yes, I agree quite a lot	
Yes, I strongly agree	



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Bias in Questionnaires

Last modified at 5/26/2011 5:33 PM by CeRC

Definition of bias

Bias can be defined as a systematic error in an epidemiological study. Please see chapter 2 on "Biases in Epidemiological Studies" for more detailed information.

Information bias

The first main bias results from the type of information obtained (information bias). A recall bias is present when cases are more likely to remember their exposure than controls. In addition, the interviewer or observer bias can occur when different interviewers have different interpretations of the similar questions or their responses.

A recall bias can be reduced by increasing the timeliness of the survey, i.e. keeping the interval between the event of interest and the survey as short as possible. The interviewer bias can be reduced by developing a questionnaire with clear instructions. Also, interviewers should be trained thoroughly and perform test interviews prior to the start of the survey.

Selection bias

Selection bias results from a systematic error in the selection of the study population. For example, if those who respond to the questionnaire differ from those who do not respond, a [nonresponder bias](#) is present (1). For example people who only own mobile phones will be excluded if the telephone interviews are restricted to landlines. Since mobile phone exclusive users tend to be younger and wealthier than the general population, a bias is introduced (2).

It is therefore important to ensure an overall high response rate, for example by offering incentives to participate or send reminders to non-responders (3). Interview partners should be chosen randomly if telephone or household interviews are performed. For example, this can be done by asking to interview the person in a household whose birthday was last. If possible, information on demographic characteristics of the non-responders should be obtained. This could be achieved by a simple non-responder survey or by collecting demographic characteristics in the

population registration office. However, obtaining information on on-responder is time-consuming and not always successful. A non-responder bias can also be corrected during the analysis , by standardising the results by age, sex etc..

References

1. Hennekens CH and Buring JE. Epidemiology in Medicine. Philadelphia 1987. PP 171-2
2. Kempf AM and Remington PL. New Challenges for Telephone Survey Research in the Twenty-First Century. Annu. Rev. Public Health 2007. 28:113-26
3. Silman AJ and Macfarlane GJ. Epidemiological studies. A practical guide. Cambridge 2002. PP 132-7



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Ten Steps to Design a Questionnaire

Last modified at 5/26/2011 5:31 PM by CeRC

Designing a questionnaire involves 10 main steps:

1. Write a study protocol

This involves getting acquainted with the subject, making a literature review, decide on objectives, formulate a hypothesis, and define the main information needed to test the hypothesis.

2. Draw a plan of analysis

This steps determines how the information defined in step 1 should be analysed. The plan of analysis should contain the measures of association and the statistical tests that you intend to use. In addition, you should draw dummy tables with the information of interest. The plan of analysis will help you to determine which type of results you want to obtain. An example of a dummy table is shown below.

Exposure	nr Cases (%)	Total	Attack Rate	RR (CI95%)
Tomato salad				
Chicken breast				

3. Draw a list of the information needed

From the plan of analysis you can draw a list of the information you need to collect from participants. In this step you should determine the type and format of variables needed.

4. Design different parts of the questionnaire

You can start now designing different parts of the questionnaire using this list of needed information.

5. Write the questions

Knowing the education and occupation level of the study population, ethnic or migration background, language knowledge and special sensitivities at this step is crucial at this stage. Please keep in mind that the questionnaire needs to be adapted to your study population. Please see "Format of Questions" section for more details.

6. Decide on the order of the questions asked

You should start from easy, general and factual to difficult, particular or abstract questions. Please consider carefully where to place the most sensitive questions. They should be rather placed in the middle or towards the end of the questionnaire. Make sure, however, not to put the most important item last, since some people might not complete the interview.

7. Complete the questionnaire

Add instructions for the interviewers and definitions of key words for participants. Insure a smooth flow from one topic to the next one (ex. "and now I will ask you some questions about your own health..."). Insert jumps between questions if some questions are only targeted at a subgroup of the respondents.

8. Verify the content and style of the questions

Verify that each question answers to one of the objectives and all your objectives are covered by the questions asked. Delete questions that are not directly related to your objectives. Make sure that each question is clear, unambiguous, simple and short. Check the logical order and flow of the questions. Make sure the questionnaire is easy to read and has an clear layout. Please see the [Hints to Design a good Questionnaire](#) section for more details.

9. Conduct a pilot study

You should always conduct a pilot study among the intended population before starting the study. Please see the [Piloting Questionnaires](#) section for more details.

10. Refine your questionnaire

Depending on the results of the pilot study, you will need to amend the questionnaire before the main survey starts.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Hints to Design a Good Questionnaire

Last modified at 5/26/2011 5:32 PM by CeRC

What is a well designed questionnaire?

A well designed questionnaire has a good appearance, is short and simple and covers topics relevant to the study question. It has a logical structure and a nice [layout](#). Well designed questionnaires can attain a high response rate and allow for an easy data summarisation and analysis. The [Seven Golden Rules](#) are helpful to design appropriate questions.

Introducing the survey

Questionnaire should always be accompanied by a cover letter (if administrated by mail) or an introduction by the interviewer. The introduction should include information on:

- who you are and who you work for;
- why you are investigating;
- where you obtained the respondent's name from;
- how and where you can be contacted.

Confidentiality should be guaranteed. The time requested to fill in the questionnaire or the length of interview should be indicated correctly. Most importantly, the introduction should clarify the usefulness of the study to the potential respondents and convince them to participate.

The first page of a questionnaire should include the return address and the study title in bold. All pages should bear an identifying mark or a unique identifier, page numbers and directions for interviewers or interviewees. The items should be numbered. If you are choosing to send a questionnaire by mail, the sending should always include a self-addressed and prepaid envelope to facilitate a response.

Order of the questions

Be aware that the order of the questions asked might influence the answers. It is recommended to group the

questions by topic. The starting questions should be simple, relevant to main subject and non-threatening in order to put the participants at ease and catch their interest. Although frequently done, neither demographic nor personal questions are a good start for the interview or a written questionnaire. The first questions should serve to get the participants "in the mood" for the topic of interest.

The biggest challenge in designing a questionnaire lies in keeping the focus on the research question. Sidetracking should be avoided at all costs. Avoid collecting unnecessary information that does not help answering your research question. However, demographic information should be collected.

Conclude the survey

Conclude the questionnaire by thanking the respondents for their participation and ensuring them that their participation was really helpful. At this stage participants should have the opportunity to ask questions on the study or the subject they were interviewed on. If needed, ask for permission to make further enquiries and record the telephone numbers of those who consented to do so.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Piloting Questionnaires

Last modified at 5/26/2011 5:33 PM by CeRC

As mentioned before, it is very important to pretest ("pilot") the questionnaire. The questionnaire should be piloted with a similar group of people to your intended subjects. The aim of a pilot survey is to obtain estimates about the expected response rates, data quality, the validity and comprehensibility of the questionnaire (1).

It allows highlighting problems such as inappropriate questions or ambiguity before starting the real survey. The effects of an alternative wording should be tested as well. A pilot questionnaire should also include questions about the overall impression on respondents and interviewers in order to include their comments. It might be useful as well to ask non-respondents for the reasons for not participating.

Often the pilot study leads to several amendments before the survey starts. Therefore, there should be sufficient time allowed for this phase in the original time schedule of the study.

References

1. Silman AJ and Macfarlane GJ: Epidemiological studies. 2nd edition 2001, Cambridge. Pp 138-141



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Validated questionnaires

Last modified at 5/26/2011 5:33 PM by CeRC

The information obtained in a study should be valid, i.e. reflecting the true situation (1).

A questionnaire is considered validated if the questionnaire has been shown to have a high reliability and internal consistency. A questionnaire with a high reliability would elicitate the same answers if applied to the same population again. Internal consistency is measured by comparing the answers to questions measuring the same concepts.

Validated questionnaires exist for a large number of study questions, for example [quality of life](#), pain and chronic diseases. Validated questionnaires are widely used in social sciences.

The ideal would be to use validated questionnaires in all investigations. In intervention epidemiology, however, validated questionnaires are very uncommon. In field epidemiology, epidemiologists rely frequently on already used standard questionnaires, for example for food-borne outbreaks. However, the situation differs for each outbreak, each study and each country. Re-using standard questionnaires will not necessarily point towards the exposure of interest. Therefore you need to generate a hypothesis first using a trawling questionnaire. This will help you to design an appropriate questionnaire for your study.

References

-
1. Silman AJ, MacFarlane GJ :Epidemiological Studies. A practical guide. 2nd edition 2002, Cambridge. Pp 111-122



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

CAUTI

Last modified at 5/4/2016 10:05 AM by Diamantis Plachouras

1. Definition.

The definition of catheter-associated urinary tract infections (CAUTI) according to the CDC is a UTI where an indwelling urinary catheter was in place for more than two calendar days on the date of event (day 1 being the day of device placement).

2. Burden of disease.

Catheter-associated urinary tract infections (CAUTI) are the most prevalent health-care associated infections (HAIs) accounting for one third (27%) of HAIs in developed countries according to the ECDCs annual report from 2008. The attributable mortality of CAUTI is low but the high frequency of catheter use in health care settings resulting in CAUTI, means that the burden of CAUTI is substantial with regard to prolonged hospital stays and increased antibiotic use. According to European studies, 15-25% of hospitalised patients and 5% of patients in elderly homes have a urinary catheter. More studies estimate that 41-58% of catheters in place are probably unnecessary. The risk for CAUTI increases by 5% for each day with a catheter. The annual costs for CAUTI accounts for £ 99 million every year (£ 1968 per episode) in the United Kingdom. A strong leadership and a systematic approach engaging all healthcare staff are crucial in order to achieve maximum effect.

3. How to prevent-specific requirements.

In addition to standard measures for the prevention of health-care associated infections (HAIs) described elsewhere up to 70% of CAUTI can be prevented by following evidence-based guidelines focusing on the catheter use [1]. A bundling strategy using selected evidence based activities undertaken simultaneously, can reduce significantly the incidence of CAUTI as well as the use of urinary tract catheter-days. Such bundled actions should focus on five clearly defined activities [2-4]:

- Avoid unnecessary urinary tract catheterisation by providing access to an evidence-based list of indications;
- Selection of catheter-material and size;
- Aseptic insertion techniques;
- Aseptic maintenance routines;
- Assess the need for maintaining the urinary catheter on a daily basis during rounds and promptly remove unnecessary urinary catheters.

3.1 Avoid unnecessary urinary catheterisation.

Minimize urinary catheter use and consider alternatives, for example suprapubic catheters, intermittent catheterisation, external condom catheters for males and diapers. Avoid bladder distension. Use portable bedside ultrasound device to assess urine volume. Make the indication for the urinary catheters clear among doctors and nurses. Urinary catheterisation is indicated in the following cases:

1. Acute urinary retention with or without obstruction;
2. Critically ill patients in need of precise measurement of urinary output;
3. During urological surgery;
4. Preoperatively for example in long operations, urologic surgery, large infusions during surgery;
5. Prolonged immobilization for example in cases with pelvic fractures and unstable thorax;
6. Improve comfort at end of life care;
7. Painful pressure ulcers and other wounds in genital area.

Clearly document the clinical indication for the urinary catheter, the time and date of insertion, the expected duration, the type of catheter and drainage system. Also note the planned date of removal of the catheter and the reasons for keeping the catheter.

3.2 Selection of catheter-material and size.

Use as small a catheter as possible ensuring proper drainage, to minimize trauma and risk of infection.

3.3 Aseptic insertion techniques.

- Insert catheters following aseptic techniques using sterile equipment;
- Perform hand hygiene before and after insertion or any manipulation of the catheter device or site;
- Gloves, drape and sponges should be sterile;
- Use a single-use packet of sterile lubricant jelly for insertion to minimise urethral trauma and discomfort;
- Urethral meatus cleaned with soap and water or sterile saline (0,9% NaCl).

3.4 Aseptic maintenance routines.

- Perform hand hygiene before and after any manipulation of the catheter device or the collecting bag;
- Maintain a sterile, continuously closed drainage system with a sampling port;
- Keep catheter properly secured to prevent movement and urethral traction;
- Keep collection bag below the level of the bladder at all times to prevent reflux;
- Maintain unobstructed urine flow and keep the catheter and collecting tube free from kinking;
- Empty collection bag when filled $\frac{3}{4}$ to prevent reflux, using a clean collecting container for each patient, using

appropriated hand hygiene, non-sterile gloves and gown. Avoid splashing and use goggles and a protective mask if necessary;

- Avoid contact of the draining spigot with the collecting container;
- Routine daily personal hygiene is all that is required for meatal and perineal cleansing;
- Collect urinary samples for cultures by aspirating urine from the needle-less sample port with a sterile syringe after disinfecting the port. Obtain larger volumes for special analyses aseptically from the tap of the drainage bag;
- Only change bag and catheter based on clinical indication such as infection, leakage, obstruction and when collecting system is damaged;
- If bladder irrigation is necessary due to anticipated obstruction, use closed continuous irrigation systems.

3.5 Daily review of urinary-catheter.

- Revise the need of catheter at daily rounds;
- Apply effective reminder systems to remove catheter, for example alerts and reminders for doctors or automatic stop orders 48-72 hours after insertion;
- If the catheter is judged to stay, clearly document why it is still in place and when it should be removed.

References.

1. Umscheid CA, Mitchell MD, Doshi JA, Agarwal R, Williams K, Brennan PJ. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. *Infection control and hospital epidemiology*. 2011;32(2):101-14.
2. Institute for healthcare improvement IHI. How to guide: Prevent catheter-associated urinary tract infections. Available from: www.ihi.org. 2011 [accessed 5th April 2016]
3. Loveday HP, Wilson JA, Pratt RJ, Golsorkhi M, Tingle A, Bak A, et al. epic3: national evidence-based guidelines for preventing healthcare-associated infections in NHS hospitals in England. *The Journal of hospital infection*. 2014;86 Suppl 1:S1-70.
4. Lo E, Nicolle LE, Coffin SE, Gould C, Maragakis LL, Meddings J, et al. Strategies to prevent catheter-associated urinary tract infections in acute care hospitals: 2014 update. *Infection control and hospital epidemiology*. 2014;35(5):464-79.

Originally contributed by Birgitta Lytsy (*Department of clinical microbiology and infection control, Uppsala University Hospital, Sweden*)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

CLABSI

Last modified at 5/4/2016 10:17 AM by Diamantis Plachouras

1. Definition.

The definition of central line-associated blood-stream infections (CLABSI) are systemic infections with the central venous catheter (CVC) as its source and when other sites of infection have been excluded (i.e. the isolation of the same microorganism from blood cultures and the CVC in significant numbers) [1]. CVCs are colonized by microorganisms on either the endoluminal or the external catheter surface beneath the skin and originate from microorganisms colonizing the patient's skin at the insertion site or the hands of the staff during insertion or contaminating the hub during care interventions. Coagulase-negative staphylococci, particularly *Staphylococcus epidermidis*, are the microorganisms most frequently implicated in CR-BSI. Other microorganisms commonly involved include *Staphylococcus aureus*, *Candida* spp and enterococci.

2. Burden of disease.

CLABSI represents 10% of all healthcare associated infections (HAIs) and are the fourth most common HAIs in acute care hospitals in Europe according to the ECDC's annual report from 2008. The prevalence of CLABSI in intensive care is higher and represents 30% of HAIs, being the second most common type of infection after respiratory infections in intensive care settings. The incidence of CLABSI is estimated to 2.7/1000 catheter days and the literature suggests that up to 70% of CLABSI could be prevented if adequate measures are undertaken [2]. In the most recent national prevalence survey in the UK, the Health Protection Agency reported that the prevalence of CLABSI was 0.5%, accounting for 7.3% of the HAIs detected. Mortality and morbidity from CLABSI are substantial and is costly for the health care system since CLABSI increase antibiotic use and length of stay in intensive care and the hospital. It is estimated that each year in the United States, central venous catheters may cause 80,000 catheter-related bloodstream infections and, as a result, up to 28,000 deaths among patients in intensive care units (ICUs) [3]. In a study from Spain the attributable mortality from CLABSI in intensive care was 10% and the median length of stay was 13 days longer for patients with CLABSI compared to controls [4].

3. How to prevent - specific requirements.

General conditions must be met to prevent HALs described elsewhere regarding education of staff, surveillance and infrastructure. The following recommendations are based on three recent guidelines [5-7].

A. Before insertion:

- indication: avoid unnecessary CVC insertion by providing access to an evidence-based list of indications;
- selection of site: select the most appropriate site for every patient. Choose subclavian veins before jugular and femoral veins when the catheter is placed under planned and controlled conditions. Use ultrasound guidance for jugular vein. Avoid femoral veins;
- type of catheter material: polytetrafluoroethylene (Teflon) and polyurethane catheters have been associated with fewer infections than catheters made of polyvinyl chloride or polyethylene. The choice of multi-lumen or single lumen catheter is still debatable. Use all-inclusive kits;
- skin asepsis before insertion: use alcoholic solution (70% isopropanol) containing chlorhexidine (0.5%) for skin preparation before insertion.

B. Principles for insertion:

- ensure maximum aseptic technique during insertion: the person inserting the catheter should wear a head cap, face mask, sterile body gown and sterile gloves, and use a full-size sterile drape. Use a checklist to ensure that aseptic technique is maintained. Another healthcare personnel person than the inserter should observe and document the insertion procedure. These healthcare personnel should be empowered to stop the procedure if breaches in aseptic technique are observed;
- choose the right dressing for the insertion site.

c. Principles for maintenance:

- catheter care and catheter site care: daily inspection of CVC site;
- hand hygiene and aseptic technique during care and maintenance and accessing the system;
- disinfect catheter hubs, needleless connectors, and injection ports before accessing the catheter;
- remove nonessential catheters: assess the need for continued intravascular access on a daily basis during multidisciplinary rounds;
- replacement strategies: routinely change of intravascular devices does not prevent CLABSI.

References.

1. O'Grady NP, Alexander M, Dellinger EP, Gerberding JL, Heard SO, Maki DG, et al. Guidelines for the prevention of intravascular catheter-related infections. Centers for Disease Control and Prevention. MMWR Recommendations and reports: Morbidity and mortality weekly report Recommendations and reports/Centers for Disease Control. 2002;51(Rr-10):1-29.
2. Umscheid CA, Mitchell MD, Doshi JA, Agarwal R, Williams K, Brennan PJ. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. Infection control and hospital epidemiology. 2011;32(2):101-14.
3. Pronovost P, Needham D, Berenholtz S, Sinopoli D, Chu H, Cosgrove S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. The New England journal of medicine. 2006;355(26):2725-32.
4. Olachea PM, Palomar M, Alvarez-Lerma F, Otal JJ, Insausti J, Lopez-Pueyo MJ. Morbidity and mortality

associated with primary and catheter-related bloodstream infections in critically ill patients. *Revista espanola de quimioterapia* : publicacion oficial de la Sociedad Espanola de Quimioterapia. 2013;26(1):21-9.

5. O'Grady NP, Alexander M, Burns LA, Dellinger EP, Garland J, Heard SO, et al. Summary of recommendations: Guidelines for the Prevention of Intravascular Catheter-related Infections. *Clinical infectious diseases* : an official publication of the Infectious Diseases Society of America. 2011;52(9):1087-99.
6. Elliott T, Timsit JF. epic3: guidelines for preventing infections associated with the use of intravascular access devices. *The Journal of hospital infection*. 2014;87(3):182.
7. Marschall J, Mermel LA, Fakhri M, Hadaway L, Kallen A, O'Grady NP, et al. Strategies to prevent central line-associated bloodstream infections in acute care hospitals: 2014 update. *Infection control and hospital epidemiology*. 2014;35(7):753-71.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

SSI

Last modified at 5/4/2016 11:19 AM by Diamantis Plachouras

1. Definition

A Surgical Site Infection (SSI) is an infection that occurs after surgery in the part of the body where the surgery took place. Surgical site infections are superficial when the skin only is involved. Deep SSI are more serious since they involve tissues under the skin, organs, or implanted material.

2. Burden of disease.

SSIs represent 17% of all healthcare associated infections (HAIs) and are the third most common HAIs in acute care hospitals in Europe according to the ECDCs annual report from 2008. The prevalence varies according to procedure and body site and prevalence figures should be stratified accordingly. Surgical procedures can be divided into three groups in terms of the risk of developing SSI; clean surgery (i.e. cardiac, orthopaedic), clean-contaminated surgery (i.e. urogenital, cholecystectomy) and contaminated surgery (i.e. colon surgery). The estimated overall incidence of SSI is 2%-5% and figures are higher for contaminated procedures.

SSIs are associated with longer post-operative hospital stays up to 7-10 days, additional surgical procedures, prolonged antibiotic treatment and increased readmission rates. SSIs may require intensive care and often result in higher mortality. The reported crude mortality rate after major surgery is 0.5-5%.

The attributable costs of SSI vary depending on the type of operative procedure and the type of infecting pathogen. Costs of SSI are believed to account for \$3.5 billion to \$10 billion annually in healthcare expenditure in the US [1].

3. How to prevent - specific requirements.

General conditions must be met to prevent HAIs described elsewhere regarding education, surveillance and infrastructure. It has been shown that 55% of SSIs are preventable when evidence-based guidelines are followed [2].

Deep SSIs are caused by bacteria inoculated into the wound during surgery and originate either from the patient

(endogenously) or from the operating team (exogenously). The most common pathogens in deep SSIs are skin flora (Coagulase-negative staphylococci, *Staphylococcus aureus*) and gut flora (*Enterobacteriaceae* and *Enterococcus* spp). The risk of SSI increases with the number of bacteria and the virulence of the bacteria inoculated in the wound and depends on the immune status of the patient. Antibiotic prophylaxis administered in the right dose and the right timing has been standard practice for any surgical procedures because evidence shows that it prevents SSI and mortality. The efficacy of prophylaxis is affected by antibiotic resistance and recent studies show that 39% to 51% of bacteria that cause infections after surgery are already resistant to standard antibiotics in the US [3]. In a world without efficient antibiotics for prophylaxis and for treating SSI the main goal for surgeons must be to focus on optimizing the patient before surgery and infection control interventions to prevent SSI. There is a substantial evidence base for preventing SSI compiled by the WHO [4], and SHEA-IDSA in the United states [1]. It is estimated that SSI are preventable by 40%–60% if adequate measures are undertaken systematically since SSI are multifactorial and several actions must be undertaken simultaneously. The outcome of SSI is dependent on surgical techniques and standardized operating procedures meaning that surgeons and intraoperative staff must be well educated, take a leading role and be the experts in the hard work of preventing SSI.

Interventions to prevent SSI can be divided into preoperative, intraoperative and postoperative. The following are the most important:

- preoperative measures:
 - consider medical factors influencing the immune system of the patient and eliminate known risk factors for SSI: i.e. stop smoking, optimize nutritional status, blood glucose if diabetes mellitus, optimize medication for COPD and heart failure, review medication list and exclude immunosuppressing drugs if possible, ensure a BMI < 30 if possible before surgery;
 - treat remote infections before surgery;
 - administer antibiotic prophylaxis with optimal timing and adequate dosing according to evidence-based standards and guidelines;
 - treat patient skin with soap or chlorhexidine combined with alcohol repeatedly;
 - remove hair when necessary using clippers;
 - wash and disinfect hands of operating team optimally;
 - use the WHO checklist before start of operation [5].
- Intraoperative measures:
 - the operating team should wear clean air suits, gloves, masks, caps and waterproof gowns according to EN-standards;
 - the surgeon should use double gloves and all persons present in the operating room should wear helmets, masks

and clean air suits if clean surgery;

- ensure the discipline in the operating room by minimizing movements, door openings and persons present;
 - optimal ventilation adapted for the procedure and risk of SSI. High flow ventilation and low air counts for clean surgery. Do not block the ventilation;
 - ensure that all instruments and equipment are adequately cleaned, disinfected and sterilized;
 - ensure the patient is normothermic with proper tissue levels of oxygen and blood-glucose throughout the operation.
- Postoperative measures:
- ensure optimal closure of wounds;
 - use drainage with closed system and removed it <24 hours;
 - leave wound dressing unchanged on as long as possible;
 - adequate cleaning of the operating theatre between patients.

References.

1. Anderson DJ, Podgorny K, Berrios-Torres SI, Bratzler DW, Dellinger EP, Greene L, et al. Strategies to prevent surgical site infections in acute care hospitals: 2014 update. *Infection control and hospital epidemiology*. 2014;35 Suppl 2:S66-88.
2. Umscheid CA, Mitchell MD, Doshi JA, Agarwal R, Williams K, Brennan PJ. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. *Infection control and hospital epidemiology*. 2011;32(2):101-14.
3. Teillant A, Gandra S, Barter D, Morgan DJ, Laxminarayan R. Potential burden of antibiotic resistance on surgery and cancer chemotherapy antibiotic prophylaxis in the USA: a literature review and modelling study. *The Lancet Infectious diseases*. 2015;15(12):1429-37.
4. WHO. WHO guideline for safe surgery. Safe surgery saves lives. 2009. Available from: http://apps.who.int/iris/bitstream/10665/44185/1/9789241598552_eng.pdf [accessed 5th April 2016].
5. WHO. Surgical safety checklist. Available from: http://www.who.int/patientsafety/safesurgery/tools_resources/SSSL_Checklist_finalJun08.pdf?ua=1wwwwho.int. 2009 [accessed 5th April 2016].



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Cluster Investigations

Last modified at 9/17/2011 6:45 PM by Arnold Bosman

This article is a result of new content structure of the FEM Wiki.

You are invited to contribute suitable content such as definitions, scope, examples and other related material. Feel free to link to other external resources.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Odds

Last modified at 10/11/2012 2:56 PM by Vladimir Prikazsky

Odds (no synonyms), are expressed as an absolute number.

The [odds](#) of an event ("odds", always plural) occurring is the probability (e.g. [risk](#)) that this event will occur divided by the probability that the event will not occur. It can also be expressed as the [probability](#) that an event will occur divided by "1 minus the probability that the event will occur". [\[1\]](#)

$$\text{Odds of event} = \frac{P}{1 - P}$$

This probability measure is popular in the world of gambling. If we compute the number of people putting money on one horse winning and the number of people putting money on the horse not winning (i.e. putting money on other horses) we can compute the odds of winning. For example among 3100 persons gambling on horses, 100 persons put money on horse "A" to win and 3000 do not (they bet on other horses). The odds of winning are then 1/30 (100/3100 divided by 3000/3100 which can be simplified as 100/3000 or 1 / 30). In fact in gambling the odds of not winning are preferred and expressed as a ratio X/1. In our example, 30/1, or in words "thirty to one". This means that for every Euro that you bet, you will receive 30 if you win.

Since in epidemiology we illustrate the population under investigation with a two-by-two table, we will use a [table](#) to describe how to calculate odds. In the two-by-two table the concept of exposure is also included. However, to calculate the odds of disease, it not needed to take into account that in our population some might have been exposed to a particular exposure and some not.

Example 1

	Developing the disease	Not developing the disease	Total
--	------------------------	----------------------------	-------

Exposed	a	b	a+b
Not exposed	c	d	c+d
Total	30	70	100

The [table](#) yields the following calculations:

$$\text{Risk of disease} = p = \frac{a}{a+b} = \frac{30}{100} = 0.3$$

$$\text{Odds of disease} = \frac{p}{1-p} = \frac{\frac{a}{a+b}}{\frac{(a+b)-a}{a+b}} = \frac{\frac{30}{100}}{1-\frac{30}{100}} = \frac{\frac{30}{100}}{\frac{100-30}{100}} = \frac{30}{70} = 0.43$$

Therefore to calculate the odds: divide the risk of getting the disease by the risk of not getting the disease. It is equal to the ratio of the number of people with the disease to the number of people without it in a particular population.

The odds is a measure rarely used in epidemiology. Most often the odds are used to express the [odds ratio](#). A disease-odds ratio is the ratio of the odds of having the disease among the exposed and the odds of having the disease among the unexposed [\[1\]](#). In other words, the odds ratio is the ratio of the odds of disease observed in 2 subsets of a population.

In you take again the [table](#) as an example, the disease-odds ratio will be equal to:

Odds of developing the disease among the exposed: a / b

Odds of developing the disease among the unexposed: c / d

Disease-odds ratio:

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a * d}{b * c}$$

As you see by comparing [example one](#), [two](#) and [three](#), the risk and the odds approximate each other when the event is rare. When the event occurs frequently the odds overestimate the risk of disease.

For this reason, in many situations (when the disease is rare) the [odds ratio](#) can estimate the risk ratio.

Example 2

	Developing the disease	Not developing the disease	Total
Exposed	a	b	a+b

Not exposed	c	d	c+d
Total	50	99 950	100 000

Risk of disease = $50 / 100000 = 0.00050000$

Odds of disease $(50 / 100000) / 1 - (50/100000) = 0.00050025$

When getting the disease is a rare event, the risk of disease approximates the odds of disease.

Example 3

	Developing the disease	Not developing the disease	Total
Exposed	a	b	a+b
Not exposed	c	d	c+d
Total	59	950	1000

Risk of disease = $50 / 1000 = 0.05000$

Odds of disease $(50 / 1000) / 1 - (50/1000) = 0.05263$

References

-
1. Porta, M. A dictionary of Epidemiology, Fifth edition. Oxford University press, 2008.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment
Introduction to Public Health and basic concepts

General Communication
Statistical Concepts

Infection control and hospital hygiene

Quantifying disease burden

Last modified at 9/20/2011 4:53 AM by Arnold Bosman

Introduction

The effect of a health hazard in a population depends on two things:

- the size of the risk for exposed individuals (relative risk)
- the number of individuals exposed (prevalence of exposure).

The product of these two factors is known as the *population attributable risk*.

If an exposure is uncommon, then the disease burden within the population will be small, no matter how great the risk to the individual, e.g. bat bites and rabies.

If an exposure is common, then the disease burden within the population may be large, even if the risk to the individual is small, e.g. egg consumption and infection with Salmonella Enteritidis [1].

The population attributable risk is an important public health concept since it provides a measure of the impact of the exposure on the population as a whole. It also indicates the reduction in disease burden that could be achieved if the risk factor was controlled or eliminated by effective preventive action.

Relative risk versus attributable risk

Risk can be expressed in relative terms (relative risk, standardised mortality ratio) or in absolute terms (absolute risk, number needed to treat) [2].

The *relative risk* (risk ratio) is the most commonly used measure of the effect of an exposure on an individual's risk of disease. It is the ratio of the occurrence (incidence) of disease in exposed people to the occurrence of disease in unexposed people.

The *attributable risk* is used to measure the public health impact of the exposure on the population as a whole, once a link between the exposure and the occurrence of disease has been established. It also indicates the potential impact of control measures.

There are two ways of measuring attributable risk:

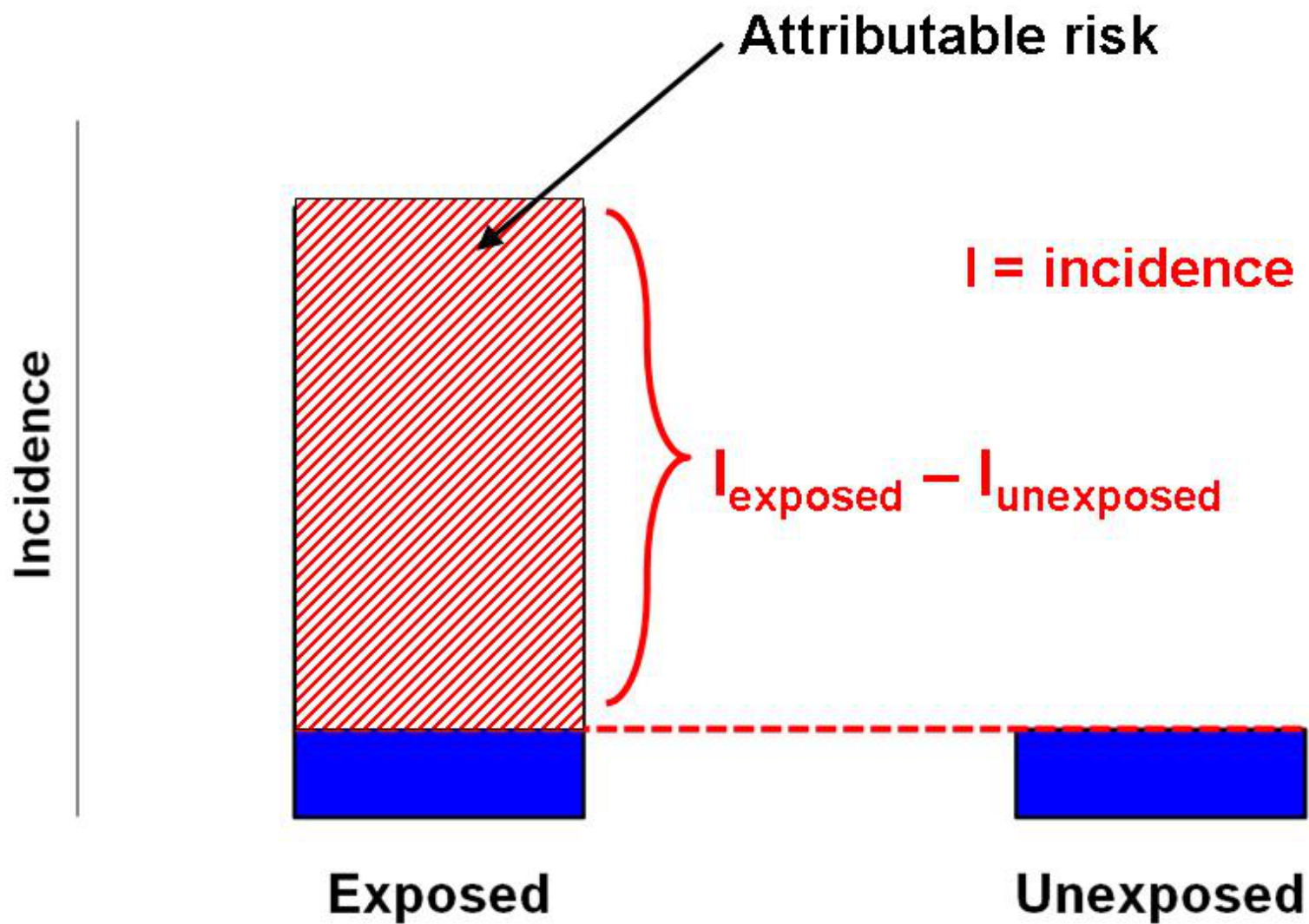
a) the attributable risk among the exposed (AR_e)

b) the attributable risk in the population (AR_{pop})

The attributable risk among the exposed (AR_e)

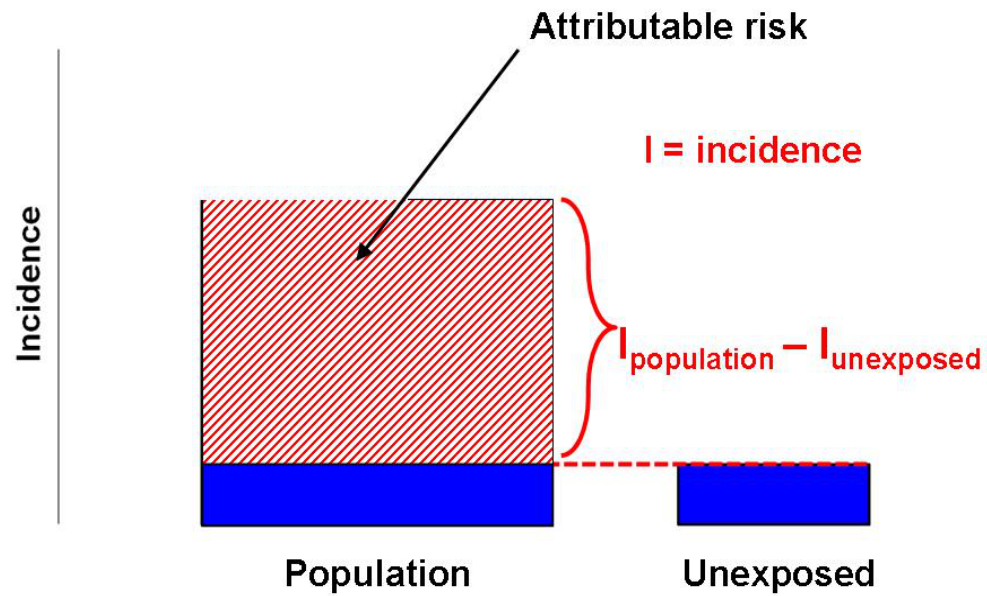
This is the proportion of cases among exposed individuals that can be attributed to the exposure.

Figure 1. Schematic illustration of attributable risk among the exposed



This is the proportion of cases in the general population that can be attributed to the exposure, and therefore prevented if an intervention exists.

Figure 2. Schematic illustration of population attributable risk



References

1. Northridge ME. Public health methods - attributable risk as a link between causality and public health action. *Am J Public Health* 1995;**85**:1202-3.
2. Irvine EJ. Measurement and expression of risk: optimizing decision strategies. *Am J Med* 2004;**117**:2S-7S.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Measures of impact among the exposed

Last modified at 5/26/2011 5:26 PM by CeRC

Attributable risk among the exposed (AR_e)

The attributable risk among the exposed (risk difference or excess risk) is the proportion of cases among exposed individuals that can be attributed to the exposure. It provides information about the absolute effects of the exposure [1].

EXAMPLE: Reducing automobile related deaths

Let us suppose that we are in charge of a prevention programme and that our goal is to reduce automobile-related deaths. However, we have a limited budget and we want to have the maximum impact on reducing deaths.

We decide to conduct a cohort study of 10,000 drivers to examine risk factors for automobile-related deaths. We are particularly interested in factors like drunk driving and speeding since we believe interventions are feasible.

We would like to quantify the disease burden (deaths) due to the exposure in each of the two groups (drunk drivers and speeding drivers). This means that in each exposed group we are aiming to measure how many of the deaths that occur are due to drunk driving and to speeding respectively.

First, we calculate the risk difference between the exposed and unexposed. This is known as the attributable risk among the exposed (AR_e):

$$AR_e = I_e - I_u$$

The study gives the following results:

Table. Risk of death from speeding or drunk driving, Anystate, 2010

Speeding	Total	No. of deaths	Risk of death	Attributable risk (exposed)
----------	-------	---------------	---------------	-----------------------------

	drivers			
Yes	2,000	100	50	50 - 10 = 40/1,000
No	8,000	80	10	
Drunk driving	Total drivers	No. of deaths	Risk of death per 1,000	Attributable risk (exposed)
Yes	300	45	150	150 - 14 = 136/1,000
No	9,700	135	14	

We can also express attributable risk as the percentage of all deaths among the exposed that can be attributed to the exposure. This is known as the attributable fraction among the exposed (AF_e):

$$AF_e = \frac{I_e - I_u}{I_e} \times 100$$

I_e = incidence among exposed

I_u = incidence among unexposed

If the risk factor is causal, then the attributable fraction among the exposed corresponds to the proportion of disease among the exposed that can:

- be attributed to the exposure
- be avoided by eliminating the exposure.

Attributable fraction in cohort studies

In a cohort study, the attributable fraction among the exposed (AF_e) is:

$$AF_e = \frac{I_e - I_u}{I_e} \times 100 = \frac{RR - 1}{RR} \times 100$$

I_e = incidence among exposed

I_u = incidence among unexposed

RR = risk ratio

In the example of speeding and drunk driving we therefore have:

Speeding

$$AF_e = \frac{I_e - I_u}{I_e} \times 100 = \frac{50 - 10}{50} \times 100 = 80\%$$

This means that (if speeding causes driving related deaths) 80% of driving related deaths among speeding drivers can be attributed to speeding. They could be avoided if speeding did not occur.

Drunk driving

$$AF_e = \frac{I_e - I_u}{I_e} \times 100 = \frac{150 - 14}{150} \times 100 = 91\%$$

This means that (if drunk driving causes driving related deaths) 91% of driving related deaths among drunk drivers can be attributed to drunk driving. They could be avoided if drunk driving did not occur.

These examples illustrate what happens if exposure increases risk of disease. If exposure prevents disease (e.g. vaccination), the attributable risk is often called the preventable fraction among the exposed (PFe).

We would then have the following:

$$PFe = \frac{I_e - I_u}{I_e} \times 100 = (1 - RR) \times 100$$

I_e = incidence among exposed

I_u = incidence among unexposed

RR = risk ratio

Table. Vaccine effectiveness in the population of Anystate, 2010

	Population	No. of cases	Cases per 1,000	Risk ratio (RR)
Vaccinated	306,045	150	0.49	0.28
Unvaccinated	298,655	515	1.72	Reference

Total	604,700	665	1.10	

To calculate the preventable fraction:

$$PFe = \frac{I_e - I_u}{I_e} \times 100 = (1 - RR) \times 100$$

$$PFe = \frac{I_e - I_u}{I_e} \times 100 = \frac{1.72 - 0.49}{1.72} \times 100 = (1 - 0.28) \times 100 = 72\%$$

The expected number of cases among the vaccinated population, if they were unvaccinated, is:

$$306,045 \times (1.72/1,000) = 526 \text{ cases}$$

We have calculated that the vaccine was able to prevent 72% of these cases (the preventable fraction).

The estimated number of cases that were prevented by the vaccination programme is therefore:

$$526 \times 0.72 = 379 \text{ cases}$$

Attributable fraction in case-control studies

$$AF_e = \frac{OR - 1}{OR} \times 100$$

$$PFe = (1 - OR) \times 100$$

AF_e = attributable fraction among the exposed

PFe = preventable fraction among the exposed

OR = odds ratio

Two assumptions are made in substituting OR for RR:

- that controls are representative of the general population
- that the prevalence of exposure is low [2].

Methods are also available for calculating attributable fractions for matched case-control studies [3].

Synopsis

Attributable risk among the exposed (AR_e)

- The *number* of cases (amount of disease) *among the exposed* that can be attributed to the exposure
- What is the risk among the exposed that is due to the exposure?
- This is calculated as the absolute difference between risk in the exposed and risk in the unexposed
- It assumes that the causal effect is entirely due to the risk factor

Synonyms:

- Attributable risk (exposed)
- Attributable benefit (exposed)
- Risk difference / Excess risk
- Rate difference / Excess rate
- Absolute risk reduction

Attributable fraction among the exposed (AF_e)

- The *proportion* of cases (percentage of disease) *among the exposed* that can be attributed to the exposure
- Attributable risk expressed as a proportion of the risk in the exposed
- What is the proportion of disease among the exposed that:
 - can be attributed to the exposure?
 - can be prevented if the exposure is eliminated?

Synonyms:

- Attributable fraction (exposed)
 - Attributable proportion / Attributable risk percent (exposed)
 - Aetiological fraction / Preventable fraction (exposed)
 - Relative risk reduction
-

References

1. Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol* 1988;**128**:1185-97.
2. Cole P, MacMahon B. Attributable risk percent in case-control studies. *Br J Prev Soc Med* 1971;**25**:242-4.
3. Kuritz SJ, Landis JR. Attributable risk ratio estimation from matched-pairs case-control data. *Am J Epidemiol* 1987;**125**:324-8.

Measures of impact in the population

Last modified at 5/26/2011 5:26 PM by CeRC

Attributable risk in the population (ARpop)

The second type of question we may ask relates to the excess risk of disease in the total population that is attributable to exposure. This is the attributable risk in the population (ARpop) or the population attributable risk. It is the proportion of cases in the general population that can be attributed to the exposure.

$$AR_{pop} = I_{pop} - I_u$$

I_{pop} = incidence in population

I_u = incidence among unexposed

It represents the reduction in risk we would achieve if the entire population was not exposed. It helps to identify which exposures are most relevant in the community and will yield most benefit from public health interventions [\[1\]](#) [\[2\]](#).

Attributable fraction in the population (AFpop)

The population attributable risk can also be expressed as a percentage of the total risk in the population.

This is known as the attributable fraction in the population (AFpop):

$$AF_{pop} = \frac{I_{pop} - I_u}{I_{pop}} \times 100 = \frac{RR - 1}{RR} \times 100$$

I_{pop} = incidence in the population

I_u = incidence among the exposed

Table. Risk of death from speeding, Anystate, 2010

Speeding	Total drivers	No. of deaths	Risk of death per 1,000	Attributable risk (population)
Yes	2,000	100	50	
No	8,000	80	10	
Total	10,000	180	18	$18 - 10 = 8/1,000$

Speeding

$$AF_{pop} = \frac{I_{pop} - I_u}{I_{pop}} \times 100 = \frac{18 - 10}{18} \times 100 = 44\%$$

This means that (if speeding causes driving related deaths) 44% of driving related deaths in the population can be attributed to speeding.

Table. Risk of death from drunk driving, Anystate, 2010

Drunk driving	Total drivers	No. of deaths	Risk of death per 1,000	Attributable risk (population)
Yes	300	45	150	
No	9,700	135	14	
Total	10,000	180	18	$18 - 14 = 4/1,000$

Drunk driving

$$AF_{pop} = \frac{I_{pop} - I_u}{I_{pop}} \times 100 = \frac{18 - 14}{18} \times 100 = 22\%$$

This means that (if drunk driving causes driving related deaths) 22% of driving related deaths in the population can be attributed to drunk driving.

AFpop can also be expressed as:

$$AF_{pop} = \frac{Pe \times (RR - 1)}{Pe \times (RR - 1) + 1} \times 100$$

The above formula is not valid if the RR is adjusted for confounders, as is often the case. In this situation one of the following alternatives is preferable:

$$AF_{pop} = \frac{PCe \times (Ie - Iu)}{Ie} \times 100 = \frac{PCe \times (RR - 1)}{RR} \times 100 = PCe \times ARe \times 100$$

Pe = proportion of the population exposed

PCe = proportion of cases exposed

Ie = incidence in exposed

Iu = incidence in unexposed

RR = risk ratio

ARe = attributable risk among exposed

If the risk factor is causal, then the population attributable risk depends on:

- the strength of the association (RR)
- the frequency of the exposure (Pe)

To have a large impact on the population, the exposure must be common.

Methods are also available for dealing with multiple exposure categories for a single risk factor [3], and for diseases caused by multiple risk factors [2] [4].

Sometimes, diseases are the result of complex interactions between risk factors. Methods to conceptualise and clarify these interactions have been developed. These include sequential attributable fractions [5] [6], and causal pies [7] [8].

Synopsis

Attributable risk in the population (ARpop)

- The *number* of cases (amount of disease) *within the population* that can be attributed to the exposure

- What is the risk within the population that is due to the exposure?
- Helps in determining the public health relevance of specific exposures within the whole community
- Assumes that the causal effect is entirely due to the exposure

Synonyms:

- Attributable risk (population)

Attributable fraction in the population (AFpop)

- The *proportion* of cases (percentage of disease) *within the population* that can be attributed to the exposure
- What is the proportion of disease within the population that:
 - can be attributed to the exposure?
 - could be prevented if the risk factor was eliminated?
 - could be prevented if everyone was exposed to the protective factor?

Synonyms:

- Attributable fraction (population)
- Population attributable fraction
- Attributable proportion (population)
- Aetiological fraction / Preventable fraction (population)
- Population attributable risk percent

References

1. Rockhill B, Newman B, Weiberg C. Use and misuse of population attributable fractions. *Am J Public Health* 1998;**88**:15-19.
2. Coughlin SS, Benichou J, Weed DL. Attributable risk estimation in case-control studies. *Epidemiol Rev* 1994;**16**:51-64.
3. Hanley JA. A heuristic approach to the formulas for population attributable fraction. *J Epidemiol Community Health* 2001;**55**:508-14.
4. Bruzzi P, Green SB, Dyer DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol* 1985;**122**:904-14.
5. Eide GE, Gefeller O. Sequential and average attributable fractions as aids in the selection of prevention strategies. *J Clin Epidemiol* 1995;**48**:645-55.
6. Rowe A, Powell KE, Flanders WD. Why population attributable fractions can sum to more than one. *Am J Prev Med* 2004;**26**:243-9.
7. Hoffmann K, Flanders WD. Estimating the proportion of disease due to classes of sufficient causes. *Am J*

Epidemiol 2006;**164**:1253-5.

8. Liao S-F, Lee W-C. Weighing the causal pies in case-control studies. *Ann Epidemiol* 2010;**20**:568-73.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Measures of disease impact - Further reading

Last modified at 6/3/2014 10:32 PM by Vladimir Prikazsky

General principles

Browner WS, Newman TB. Sample size and power based on the population attributable fraction. *Am J Public Health* 1979;**79**:1289-94.

Bruzzi P, Green SB, Dyer DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol* 1985;**122**:904-14.

Coughlin SS, Bencihou J, Weed DL. Attributable risk estimation in case-control studies. *Epidemiol Rev* 1994;**16**:51-64.

Fleiss JL. Inference about population attributable risk from cross-sectional studies. *Am J Epidemiol* 1979;**110**:103-104.

Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol* 1988;**128**:1185-97.

Hanley JA. A heuristic approach to the formulas for population attributable fraction. *J Epidemiol Community Health* 2001;**55**:508-14.

Heller RF, Dobson AJ, Attia J, Page J. Impact numbers: measures of risk factor impact on the whole population from case-control and cohort studies. *J Epidemiol Community Health* 2002;**56**:606-10.

Heller RF, Buchan I, Edwards R, Lyratzopoulos G, McElduff P, St Leger S. Communicating risks at the population level: application of population impact numbers. *Br Med J* 2003;**327**:1162-5.

Northridge ME. Public health methods – attributable risk as a link between causality and public health action.

Am J Public Health 1995;**85**:1202-3.

Rockhill B, Newman B, Weiberg C. Use and misuse of population attributable fractions. *Am J Public Health*

1998;**88**:15-19.

Rowe A, Powell KE, Flanders WD. Why population attributable fractions can sum to more than one. *Am J Prev*

Med 2004;**26**:243-9.

Taylor JW. Simple estimation of population attributable risk from case-control studies. *Am J Epidemiol*

1977;**106**:260.

Walter DS. Calculation of attributable risks from epidemiological data. *Int J Epidemiol* 1978;**7**:175-82.

Walter DS. Effects of interaction, confounding and observational error on attributable risk estimation. *Am J*

Epidemiol 1983;**117**:598-604.

Examples from field epidemiology

Bhunja R. Hutin Y. Ramakrishnan R. Pal N. Sen T. Murhekar M. A typhoid fever outbreak in a slum of South Dumdum

municipality, West Bengal, India, 2007: evidence for foodborne and waterborne transmission. *BMC Public Health*

2009;**9**:115.

Birmingham ME. Lee LA. Ntakibirora M. Bizimana F. Deming MS. A household survey of dysentery in Burundi:

implications for the current pandemic in sub-Saharan Africa. *Bull World Health Org* 1997;**75**:45-53.

Caceres VM. Ball RT. Somerfeldt SA. Mackey RL. Nichols SE. MacKenzie WR. Herwaldt BL. A foodborne outbreak of

cyclosporiasis caused by imported raspberries. *J Family Pract* 1998;**47**:231-4.

Crocetti E. Arniani S. Bordoni F. Maciocco G. Zappa M. Buiatti E. Effectiveness of influenza vaccination in the elderly in

a community in Italy. *Eur J Epidemiol* 2001;**17**:163-8.

Hennessy TW. Cheng LH. Kassenborg H. Ahuja SD. Mohle-Boetani J. Marcus R. Shiferaw B. Angulo FJ. Emerging

Infections Program FoodNet Working Group. Egg consumption is the principal risk factor for sporadic Salmonella

serotype Heidelberg infections: a case-control study in FoodNet sites. *Clin Infect Dis* 2004;**38 Suppl 3**:S237-43.

Pepin J. Saheb N. Coulombe MA. Alary ME. Corriveau MP. Authier S. Leblanc M. Rivard G. Bettez M. Primeau V.

Nguyen M. Jacob CE. Lanthier L.

Emergence of fluoroquinolones as the predominant risk factor for Clostridium difficile-associated diarrhea: a cohort

study during an epidemic in Quebec.

Clin Infect Dis 2005;**41**:1254-60.

Richardson G. Thomas DR. Smith RM. Nehaul L. Ribeiro CD. Brown AG. Salmon RL. A community outbreak of *Campylobacter jejuni* infection from a chlorinated public water supply. *Epidemiol Infect* 2007;**135**:1151-8.

Walter ND. Taylor TH. Shay DK. Thompson WW. Brammer L. Dowell SF. Moore MR. Active Bacterial Core Surveillance Team. Influenza circulation and the burden of invasive pneumococcal pneumonia during a non-pandemic period in the United States. *Clin Infect Dis* 2010;**50**:175-83.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Viewpoints for Causality (Bradford Hill)

Last modified at 9/15/2015 4:48 PM by James Stuart

Nine "viewpoints" for causality were set out by Bradford Hill [\[1\]](#) .

Strength of Association

Strong associations are more likely to be causal than weak ones. In common source outbreaks, we look for one (or two) vehicles of infection, so we would expect the risk ratio for the true source to be very high. They often are. In outbreaks of food borne disease, association between illness and exposure to a common source may lead to odds ratios exceeding 50 [\[2\]](#) [\[3\]](#) . On the other hand, not all strong associations are causal and may be due to [confounding](#). This can be illustrated by the association observed between multiple births and Down syndrome, an association that is actually confounded by age of the mother [\[4\]](#). The use of the [strength of the association](#) as a causal criterion can be misleading where confounding factors are not known.

Weak associations do not rule out causality and may still have public health importance. This can apply when an exposure is common in a population. e.g. passive smoking and lung cancer (Risk Ratio = 1.3) [\[5\]](#).

Consistency

Repeated observations of an association in different populations under different circumstances showing the same or similar results suggest that the results of a single study are not due to chance. Over 100 studies into the association between smoking and lung cancer demonstrate increased risk of the disease. Note that the weighted results of similar studies, if statistically homogeneous, can be combined together in a meta- analysis.

However, consistency can also be misleading. Five studies compared the risk of dying from meningococcal disease after administration of oral antibiotics before admission to hospital versus no administration. All 5 studies showed lower mortality in those given antibiotics [\[6\]](#). The results were consistent and statistically homogeneous, thus in favour of a protective role for oral antibiotics before hospitalisation. However, doctors are more likely to prescribe oral antibiotics if they do not suspect meningococcal disease i.e. in patients with milder illness. We can infer that the

association observed is confounded by disease severity. Consistency of results in observational studies may simply be due to the presence of the same confounding factors.

Epidemiologists should be cautious not to mistake statistical significance for consistency. Different studies may show similar effect measures but with different levels of significance (including significant and not significant) and still be consistent.

Specificity

The criterion of specificity requires that a cause leads to a single effect, not multiple effects.

It is not very helpful in establishing causality. The fact that one agent contributes to multiple diseases should not be evidence against its role in any one disease. Smoking, for example, can lead to many ill effects in the smoker. The specificity criterion has repeatedly been used by those against 'smoking as a cause of lung cancer' as their main argument.

Temporality

Exposure must precede the disease.

This is the only criterion that is fundamental to postulating a cause and effect relationship and fits our intuitive understanding of causality.

However the example of Mumps Measles Rubella (MMR) vaccine and autism illustrates that our self-taught causal inference can be erroneous. Some parents observe the beginning of autism shortly after an injection of MMR. Quite naturally, they attribute this illness to the vaccine. But autism often begins at the age when MMR is given. Studies have shown that autism is just as likely to occur before as after an MMR injection, and that children who have not been given MMR are just as likely to get autism as those who have [7].

It is sometimes difficult to document sequence, especially if there is a long lag between the exposure and the disease, subclinical disease, exposure (e.g., a treatment) brought on by an early manifestation of the disease.

Biological gradient (dose response)

A biological gradient exists when the risk of disease/outcome increases with increasing exposure to the suspected risk factor.

A linear relationship between dose and response supports causality. For example, the higher the number of cigarettes smoked, the greater the risk of lung cancer.

On the other hand, a lack of dose response does not exclude a causal link. Causal associations showing a single jump (threshold, saturation effects) rather than a monotonic trend have been described.

Sometimes associated non-causal factors may also increase in a similar way to the causal factor. As discussed earlier,

the risk of Downs syndrome seems to increase with birth rank, while birth rank increases with age of the mother. Although a gradient is observed, the cause of the increased risk of Down's syndrome is linked to age of the mother not birth rank [4].

Plausibility

This refers to the biological plausibility of the hypothesis i.e. its consistency with current biological knowledge about the disease (for example, oral contraceptives and *** cancer). Being largely based on prior beliefs, it remains a subjective judgement.

During an outbreak of psittacosis in Australia, reported in the Lancet, 16 cases had spent 17.5 hrs in the garden compared with controls who had only spent 5.2 hrs [8]. Cases were more likely to mow the lawn than their controls (OR 8.8, 95%CI 1.2 – 389). It was quite plausible that this was a causal relationship (and still is). However, the authors in their study, had not taken account of the gender of controls. Controls were evenly distributed between males and females yet nearly all the cases were male. Stratification by gender reduced the strength of association, as measured by Mantel Haenszel Weighted Odds Ratio, to 5.5 and its lower limit of the confidence interval to 0.6 ($p=0.19$). Although the OR was still raised increased after stratification, it was misleading to present the results without taking gender into account.

In this way, plausibility may sometimes mislead in drawing conclusions.

Coherence

We speak of coherence when the interpretation of cause-effect relationship does not conflict with what is known of the natural history and biology of disease. This is similar to plausibility. All observations are expected to fit with a hypothesized model to form a coherent picture.

Absence of coherence cannot be taken as evidence against causality, and vice versa. Many studies have shown that prevalence of meningococcal carriage in teenagers rises with age. This has been explained biologically as the result of changes in mucosal characteristics that occur with age. However, when prevalence was adjusted for social factors such as going to pubs and clubs, kissing and smoking, the increasing trend disappeared [9].

Experimental evidence

Ideally experimental evidence should be obtained if at all possible. Robert Koch proposed four postulates that establish an micro-organism as a cause of a disease.

Summary of Koch's postulates

- (i) the micro-organism must be consistently present in the diseased and not in the healthy individual
- (ii) the micro-organism must be isolated and grown
- (iii) pure culture of the micro-organisms should induce disease
- (iv) the micro-organism must be re-isolated and shown to be the same as in (i)

These four postulates are regarded as sufficient but not necessary to establish [causation](#). Certain types of study designs may provide more convincing evidence than others. In current systems for classification of evidence (eg. the Scottish Intercollegiate Guidelines Network), randomised controlled trials are considered to provide strong evidence of cause and effect [\[10\]](#) . The highest level is provided by systematic reviews of randomised controlled trials.

Since it is sometimes unethical and/or impractical to conduct controlled trials, a possible alternative is to remove the exposure and see if the disease decreases, unless the causal process is regarded as irreversible. This is observed in “natural experiments” when intervention leads to change in one direction, and removal of that intervention reverses direction in outcome. The introduction of pertussis vaccination in the UK for instance, led to a fall in the incidence of whooping cough. Unsubstantiated concerns about adverse effects later led to a fall in its uptake. This was followed by a rise in incidence of the disease. Uptake then rose again and incidence fell correspondingly [\[11\]](#).

Analogy

The existence of other cause-effect relationships analogous to the one under study supports a causal interpretation. This is a weak criterion for causality, but can be useful for speculating how a risk factor may operate in a different context.

References

1. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; 58:295-300.
- 2.. Killalea D, Ward LR, Roberts D, de Louvois J, Sufi F et al. International epidemiological and microbiological study of outbreak of *Salmonella agona* infection from a ready to eat savoury snack - I: England and Wales and the United States. *BMJ* 1996; 313:1105-7.
3. Lewis DA, Paramathanan R, White DG, Neil LS, Tanner AC et al. Marshmallows cause an outbreak of infection with *Salmonella enteritidis* phage type 4. *Commun Dis Rep Review* 1996; 6: 183-8.
4. Chan A, McCaul KA, Keane RJ, et al. Effect of parity, gravidity, previous miscarriage, and age on risk of Down's syndrome: population based study. *BMJ* 1998;317:923-4.
5. Taylor R, Najafi R, Dobson A. Meta-analysis of studies of passive smoking and lung cancer: effects of study type and continent. *Int J Epidemiol* 2007; 36: 1048-59
6. Purcell B, Samuelsson S, Hahne S, Ehrhard I, Heuberger S, Camaroni I, Charlett A, Stuart JM. Effectiveness of antibiotics in preventing meningococcal disease after a case: systematic review. *BMJ* 2004;328: 1339-42.
7. Taylor B et al. Autism and MMR vaccine: no epidemiological evidence for a causal association. *Lancet* 1999; 353: 2026-9.
8. Williams J, Tallis G, Dalton C, Ng S, Beaton S et al. Community outbreak of psittacosis in a rural Australian town. *Lancet* 1998; 351: 1697-9.

9. MacLennan J et al. Social behaviour and meningococcal carriage in British teenagers. *Emerg Infect Dis* 2006; **12**: 950-7.
10. SIGN 50: A guideline developers' handbook. Section 6: Forming guideline recommendations.
<http://www.sign.ac.uk/guidelines/fulltext/50/section6.html> (4 August 2006).
11. Gangarosa EJ et al. Impact of anti-vaccine movements on pertussis control: the untold story. *Lancet* 1998;351:356-61.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Causal Inference

Last modified at 9/15/2015 4:40 PM by Arnold Bosman

Association and Causation

Epidemiologists aim to draw conclusions on whether an observed association is one of cause and effect. Establishing this relation (causation/causality) is a difficult task. In fact the concept of cause itself continues to be debated as a philosophical matter in the scientific literature. In this section we explore what is meant by causation and encourage an open mind about causal inference.

As "cause" can be used to give undue weight to an association, it is important to consider and remember its meaning. One definition of cause is a "preceding event", condition or characteristic that leads to a given outcome at that time. The [mechanism behind a cause](#) can be divided into necessary and sufficient components

Bradford Hill set out nine [viewpoint on causality](#): strength of association, consistency, specificity, temporality, biological gradient, plausibility, coherence, experimental evidence, and analogy. While these viewpoints are helpful when considering cause and effect, he insisted that "none of [his] nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis". What they can do, with greater or lesser strength, is to help epidemiologists make up their minds on the fundamental question - Is there any other way of explaining the set of facts before them? Is there any other answer equally, or more, likely than cause and effect?

It is important to keep in mind that most judgments of cause in epidemiology are tentative and should remain open to change with new evidence. It is important to remain critical, to aim always for stronger evidence, and to keep an open mind. Checklists of causal criteria should not replace critical thinking.

"The world is richer in associations than meanings, and it is the part of wisdom to differentiate the two." John Barth, novelist.

The concept of causal inference is related to, yet differs from [statistical inference](#), which is described elsewhere.

EPIET Lectures:

Causal Inference



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Effect Modification and Confounding

Last modified at 9/15/2011 10:25 PM by Arnold Bosman

An epidemiological study can be conducted to investigate the cause of a disease in a certain population, attempting to quantify an [observed association](#) between exposures and disease outcomes [1]. To determine the effect of an exposure within a population on disease occurrence, requires ideally a comparison between disease occurrence in a certain 'exposed group' with the disease occurrence within that same group in absence of that same exposure. As this is impossible (once people are exposed, we can no longer study how disease would have occurred among those same individuals without the exposure), we usually compare the incidence amongst an exposed group to that amongst a similar, yet unexposed unexposed group. If the incidence amongst the unexposed is the same as that amongst the exposed had they *not been* exposed, then the straightforward comparison is justified. If not, then the comparison is confounded; bias is introduced.

If life was truly simple, then to *measure the effect* between exposure and outcome (expressed as relative risk, odds ratio, vaccine effectiveness etc) ideally it would be enough to measure the distribution of the exposure and outcome of interest in a population and present these variables in a single two-by-two table.

However, life is always more complex; there are 'third variables' that can distort (*confound*) or *modify* the effect in our study. In some studies there may be many of these third variables, which we call confounders.

Serious problems can arise if confounding and effect modification are not considered at all stages: designing a study, analysing the data, interpreting the findings [2].

Confounding

If a factor is known to be associated with **both** the exposure **as well as** with the outcome in a study, such a 'third variable' is considered to be a [confounder](#). Unless we correct for this confounding variable, our [measurement of association](#) (e.g. RR or OR) will be distorted, leading to over- or underestimation of the true effect. In some instances, it might reverse the direction of the effect.

There are two ways to account for confounding variables:

Stratification of data

An association may be seen between age at first birth and carcinoma of the breast. There is also a perceived association between the number of children a woman bears and carcinoma of the breast - however those who have their first child earlier will be more likely to have larger families. Therefore, the data must be separated to compare those who only have one child; and risks calculated according to age at first parity. Note that not all the data collected will be used if this is the plan of analysis.

Unless the association between the exposure and the disease outcome varies markedly between the strata, the evidence from the different strata can be combined to present a summary of the association, to create one RR value or one OR value. Some strata may include more individuals than others, and therefore will have a more accurate measurement of the association. Therefore the average of the associations observed across all strata is weighted towards the most accurate: the most widely used weighting scheme used is that proposed by [Mantel and Haenszel](#).

Multivariable analysis

[Logistic regression](#) can be used to simultaneously adjust for the effects of more than one confounding variables. Similar methods can be used for data from cohort studies.

Effect Modification

Using an adjusted/weighted odds ratio implies that the observed association between exposure and disease is really the same in each of the strata - once the strata are defined by the levels of the confounding variables. However, this is not always the case, and where it is not, it makes no sense to present a summary of the association. If the exposure causes the disease according to different levels of the confounding variable, then we say that the confounding variable is actually an [effect modifier](#). [Interaction](#) and "heterogeneity between strata" are frequently used as though synonymous with effect modification, though they do differ. In this event, it will be appropriate to present different measures of associations (RR or OR) as according to the different levels.

References

1. Richard Farmer, Ross Lawrenson, David Miller. *Epidemiology and Public Health Medicine*, Fifth edn. John Wiley & Sons Ltd, 2008,
2. Raj Bhopal. *Concepts of Epidemiology: an integrated introduction to the ideas, theories, principles and methods of*



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Selection Bias

Last modified at 5/26/2011 5:06 PM by CeRC

Selection **bias** is a **systematic error** in a study that occurs from the process used to identify (select) the study participants, allocate them to study groups and from factors that influence study participation [1][2].

It leads to preferential selection or participation of subjects into the study according to their exposure status or outcome/disease status, with resulting systematic differences in the characteristics of participants between the study and control groups, i.e. the groups differ from each other by factors which may affect the outcome of the study [2]. The measurement of the **association** between exposure and outcome will then differ between those who are included in the study and those who were eligible but not included.

Selection bias may be due to:

- Sampling bias
- Ascertainment bias
 - case ascertainment (surveillance) bias
 - referral / admission bias
 - diagnostic bias
- Participation bias
 - self-selection (volunteerism)
 - healthy worker effect
 - non-response / refusal bias
 - survival bias
 - loss to follow-up

Selection biases in case-control studies include among others: case ascertainment (surveillance) bias, referral bias, diagnostic bias, non-response bias, survival bias.

Selection biases in cohort studies include: healthy worker effect, diagnostic bias, non-response bias, loss to follow-up.

In epidemiological studies, all efforts should be made to avoid biasing the selection of study participants. By paying attention to a number of factors, it is possible to [minimise selection bias](#).

References

1. Rothman KJ. Epidemiology - An Introduction. New York: Oxford University Press; 2002.
2. Bailey L, Vardulaki K, Langham J, Chandramohan D. Introduction to Epidemiology. Black N, Raine R, editors. London: Open University Press in collaboration with LSHTM; 2006.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Information (Measurement) Bias

Last modified at 5/26/2011 5:23 PM by CeRC

Information (or measurement) [bias](#) refers to a [systematic error](#) in the measurement or classification of participants in a study [\[1\]](#). It occurs when the accuracy of information collected about or from study participants is not equal between cases and controls (i.e. differences in accuracy of exposure data), or, between exposed and unexposed (i.e. differences in accuracy of outcome data). Lack of accuracy could mean that study subjects are assigned into the wrong category of exposure (exposed/unexposed) or outcome (case/control), or both. All attempts should be made to minimise or [prevent information bias](#).

The term "misclassification" is frequently used to describe this [bias](#). Cases and controls can be misclassified. Exposed and unexposed as well e.g. a heavy smoker who is categorised as a light smoker is misclassified. Misclassification results in an incorrect estimation of the [association](#) between exposure and outcome, the size and direction of this depending on the type of misclassification of exposure or outcome. The mechanism of misclassification can be [differential](#) or [non-differential](#).

Differential (non-random or preferential) misclassification

This occurs when one group of study participants is more likely to be misclassified than the other [\[1\]](#). [Misclassification](#) of *exposure* is differential if it differs according to a person's disease status (e.g. if cases are more or less likely to be classified as being exposed than controls ([case-control study](#))). [Misclassification](#) of *outcome* (disease) is differential if it differs between exposed and unexposed (e.g. if a person's exposure status makes them more or less likely to be classified as having the disease ([cohort study](#))). Differential [biases](#) can either increase or decrease the [measured effect](#).

Non-differential (random) misclassification

Non-differential (random) misclassification occurs when there is an equal likelihood of both groups (cases or controls, exposed or unexposed) being misclassified [\[1\]](#). With this type of [misclassification](#), either exposure or outcome (or both) is misclassified [\[2\]](#), but the misclassification is independent of a person's status for the other variable. [Misclassification](#)

of *exposure* is non-differential if it is similar among cases and controls i.e. the exposure (mis)classification is not related to the person's disease status. *Misclassification* of *outcome* (disease) is non-differential if it is equal between exposed and unexposed i.e. the outcome (mis)classification is not related to the person's exposure status.

The consequence of *non-differential misclassification* of a *dichotomous exposure* (e.g. exposed/unexposed) is - if there is an *association* - a weakening/ dilution of the *measure of association* (e.g. decrease the true value of an OR or RR), even to the point where a significant difference becomes insignificant [2][3]. It produces an *estimate of effect* - if there is an effect - that is diluted or closer to the no-effect or null value than the actual effect i.e. a "*bias towards the null*". If there is no *association* to begin with, then *random misclassification* of the exposure will not bias the estimate of the *measure of association*, or create a *bias* that makes a factor seem significant for development of disease [2][3].

According to Rothman [2][4], if the exposure is not dichotomous, there may be *bias* away from or towards the null value; it depends on the categories to which individuals are misclassified. However, in general, random misclassification between two exposure categories will make the estimates of *measures of association* for those categories converge towards one another [2][4].

Differential misclassification	Non-differential misclassification
<ul style="list-style-type: none">• Systematic error (bias)• Misclassification of exposure DIFFERS between cases and controls• Misclassification of outcome DIFFERS between exposed and unexposed• Measure of association may be exaggerated or underestimated	<ul style="list-style-type: none">• Random error• Misclassification of exposure is SIMILAR between cases and controls• Misclassification of outcome is SIMILAR between exposed and unexposed• Weakness of the measure of association ("<i>bias towards the null</i>")

Misclassifications might be introduced by the *observer* (*interviewer bias*, *biased follow-up*), by the study participants (*recall bias*, *prevarication*), or by measurement tools such as *questionnaires* or instruments such as weighing scales or blood pressure cuffs.

Observer bias occurs when data gathering is influenced by knowledge of the exposure or outcome/disease status of the subject, or by the hypothesis under study [1].

Interviewer bias

Interviewer *bias* happens when interviewers ask questions differently about exposure to cases and controls in a *case-control study*, or, ask questions differently about outcome to exposed and unexposed in a *cohort study*. Knowledge of the patient's disease/outcome status may influence both the intensity and outcome of a search for exposure to the putative cause (Sackett described this as '*exposure suspicion bias*' [5]).

Example: in an EU-wide foodborne outbreak of listeriosis, British investigators in a *case-control study* may probe listeriosis cases about consumption of a suspected food item (French non-pasteurised milk soft cheese) more than controls. This can lead to an overestimation of 'a', falsely increasing the *odds ratio* (OR).

Exposure (risk factor)	Cases of listeriosis	Controls	OR
<i>Eats soft cheese</i>	a↑	b	OR↑
<i>Doesn't eat soft cheese</i>	c	d	reference
Total			

[Interviewer bias](#) may also happen when different interviewing techniques (e.g. [self-administered questionnaires](#) (postal or email or web-based) or [interviewer-administered questionnaires](#) (by phone interview or face-to-face) or proxy) are used for cases and controls. Different approaches can be taken to [prevent interviewer bias](#).

Biased follow-up

In this type of [differential misclassification](#), unexposed people are less likely to be diagnosed for disease than exposed people.

Example: in a study looking at risk factors for mesothelioma, which can be difficult to diagnose histologically, a histopathologist may be more likely to report on a biopsy specimen as mesothelioma if a history of asbestos exposure is reported. The diagnosis of mesothelioma might be less likely to be reported among those without a history of asbestos exposure, leading to a [differential misclassification](#) of disease.

Recall bias

Recall [bias](#) is a [systematic error](#) that occurs due to differences in accuracy or completeness of recall of past events/ exposures (e.g. between cases and controls), that is not independent of outcome/disease (or exposure) status [1], e.g. a person may be more likely to recall an exposure to a potential risk factor if they become ill (become a case). It is a [differential misclassification](#) because the information on exposure is misclassified differentially for those with and without disease [2]. It has also been described as *response bias* [3], and *responder bias* or *reporting bias* [1].

Example: in a [case-control study](#) to identify the vehicle of a foodborne outbreak of Salmonella, study participants are interviewed to obtain exposure information after (Salmonella) disease has already occurred. Cases may be more likely to remember exactly what they ate than controls, since they may already have suspected a particular food (e.g. eggs), and/or thought about the possible dishes that could be responsible. This would result in an increase in the measured [OR](#) for the suspected food item.

Exposure (risk factor)	Salmonella cases	Controls	OR
<i>Ate eggs</i>	a↑	b	OR↑
<i>Didn't eat eggs</i>	c	d	reference
Total			

Example: in a [case-control study](#) of babies born with birth defects/ malformations, mothers who have given birth to a baby with a malformation may be more likely to recall accurately many exposures/ events during early pregnancy e.g. taking non-prescription drugs, experiencing trauma, having a febrile rash etc. The adverse pregnancy outcome serves

as a stimulus for the mother to remember and consider potential exposures, a stimulus that mothers who give birth to normal babies don't have [2]. This particular type of [recall bias](#) has been described as *maternal recall bias* [2].

Exposure (risk factor)	Mothers whose pregnancy ended in foetal malformation (cases)	Mothers whose pregnancy ended normally (controls)	OR
Took non-prescription drugs	a↑	b	OR↑
Didn't take non-prescription drugs	c	d	reference
Total			

Example: [case-control studies](#) on self-reported sun exposure as a risk factor for melanoma have been described as having the potential for [recall bias](#) as there is a lot of public awareness about the relationship of melanoma with ultraviolet radiation [6][7].

Note: as described by Rothman [2], this type of [recall bias](#) ([a differential misclassification](#)) is distinct from the general problem - which to some extent affects all people - of remembering and reporting exposures accurately, which tends to result in a [non-differential misclassification](#). Different approaches can be taken to [prevent recall bias](#) and to [reduce maternal recall bias](#).

Prevarication

This happens when some subjects deliberately lie when responding to the interviewer. According to how the subjects respond, this could increase or decrease the [measure of effect](#).

Example: in a [case-control study](#) looking at risk factors for death among elderly people during a heatwave, interviewed relatives may deny all behaviour which would suggest isolation/ abandonment of their elderly relatives. As a result, 'isolation' as a risk factor for heatwave-related death may be under-reported by relatives of elderly people who have died. Underestimation of 'a' will result in an underestimation of the [measure of effect](#), in this case the [odds ratio](#) (OR).

Exposure (risk factor)	Cases (elderly dead)	Controls	OR
Isolation	a↓	b	OR↓
Non-isolation	c	d	reference
Total			

References

1. Bailey L, Vardulaki K, Langham J, Chandramohan D. Introduction to Epidemiology. Black N, Raine R, editors. London: Open University Press in collaboration with LSHTM; 2006.

2. Rothman KJ. Epidemiology - An Introduction. New York: Oxford University Press; 2002.

3. Giesecke J. Modern Infectious Disease Epidemiology. 2nd ed. London: Arnold; 2002.
4. Rothman KJ, Greenland S, Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
5. Sackett DL. Bias in analytic research. J Chronic Dis. 1979;32(1-2):51-63.
6. [Parr CL, Hjartåker A, Laake P, Lund E, Veierød MB. Recall bias in melanoma risk factors and measurement error effects: a nested case-control study within the Norwegian Women and Cancer Study. Am J Epidemiol. 2009 Feb 1;169\(3\):257-66. Epub 2008 Nov 14](#)
7. [Gefeller O. Invited commentary: Recall bias in melanoma - much ado about almost nothing? Am J Epidemiol. 2009 Feb 1;169\(3\):267-70; discussion 271-2. Epub 2008 Nov 14](#)

Selection bias and case-control studies

Last modified at 10/11/2013 4:38 PM by Patty Kostkova

[Selection bias](#) occurs in [case-control studies](#) when cases and/or controls are selected on criteria related to the exposure of interest, i.e. they are selected differentially on the basis of their exposure status or there may be differences in reporting of exposure status between cases and controls [1]. [Case-control studies](#) are susceptible to [selection bias](#), as both the exposure and disease/outcome have occurred by the time the patient is recruited into the study [1].

In [case-control studies](#), [selection bias](#) can occur in the selection of cases if they are not representative of all cases within the population, or in the [selection of controls](#) if they are not representative of the population that produced the cases [1].

Example: in a hospital-based [case-control study](#) looking at the relationship between alcohol consumption and development of liver cirrhosis, in the first instance we select our controls from patients hospitalised due to trauma (Controls A). We classify our exposure (alcohol consumption) into 'heavy alcohol use' and 'light / no alcohol use'.

Exposure	Cases (liver cirrhosis)	Controls A (trauma ward)	OR
<i>Heavy alcohol use</i>	80	40	6.0
<i>Light / no alcohol use</i>	20	60	reference
Total	100	100	

But, how representative are hospitalised trauma patients of the population which gave rise to the cases? In the trauma ward, where we have selected our controls, there may be a higher proportion of patients who report heavy alcohol use compared to those who report heavy drinking in the population which produced the cases, leading to an underestimation of the [odds ratio](#) (OR). Compare this to the situation if we select our controls from hospitalised patients in a non-trauma ward (Controls B).

Exposure	Cases (liver cirrhosis)	Controls A (trauma ward)	OR	Controls B (non-trauma ward)	OR
<i>Heavy alcohol use</i>	80	40	6.0	10	36.0
<i>Light / no alcohol use</i>	20	60	ref.	90	ref.
Total	100	100		100	

Selection biases in case-control studies include among others: [case ascertainment \(surveillance\) bias](#), [referral bias](#), [diagnostic bias](#), [non-response bias](#), [survival bias](#).

References

1. Bailey L, Vardulaki K, Langham J, Chandramohan D. Introduction to Epidemiology. Black N, Raine R, editors. London: Open University Press in collaboration with LSHTM; 2006.
2. Sackett DL. Bias in analytic research. J Chronic Dis. 1979; 32(1-2):51-63.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Selection bias and cohort studies

Last modified at 10/27/2014 11:21 AM by Vladimir Prikazsky

[Selection bias](#) may occur in [cohort studies](#) if the exposed and unexposed groups are not truly comparable [\[1\]](#), e.g. comparing an occupational [cohort](#) with the general population.

[Selection biases](#) in [cohort studies](#) include: [healthy worker effect](#), [diagnostic bias](#), [non-response bias](#) and [loss to follow-up](#).

Healthy worker effect

The [healthy worker effect](#) (HWE) bias is an example of a [selection bias](#) that underestimates the mortality/morbidity related to occupational exposures [\[2\]](#). This [bias](#) reflects the healthier status of the workforce compared to the general population (which includes people who are too sick to work), so that a direct comparison of the workforce with the general population will be [biased](#). It is a problem for those who study occupational [cohorts](#).

The [healthy worker effect](#) phenomenon often leads, paradoxically, to lower mortality/ morbidity [rates](#) observed in subjects exposed to workplace toxins compared to the general population. Any excess risk associated with an occupation will tend to be underestimated by a comparison with the general population [\[1\]](#), leading to an underestimation of relative risk (RR) for occupational exposure and disease.

The following table illustrates the [incidence rate](#) of disease X in an exposed group of workers compared with the [incidence rate](#) in the general population (see the 'Total' row in the table).

		Person-time (years)	Cases	Cases / 100 years
<i>Exposed workers</i>		50,000	500	1.0
General population	Total	500,000	7,000	1.4
	<i>Workers</i>	450,000	4,500	1.0

	Non-workers	50,000	2,500	5.0
--	-------------	--------	-------	-----

In this hypothetical example, the [incidence rate](#) observed among exposed workers is 1 case/100 years compared to 1.4 cases/ 100 years in the general population, suggesting that exposed workers have a lower rate of illness than the general population. The general population, however, is composed of two groups: people that are healthy enough to work (workers), and many people who cannot work because of ill-health (non-workers). The group that is too sick to work is included among the non-workers in the table, and results in non-workers having a higher incidence than the remainder of the general population that comprises current workers [\[2\]](#).

In the above example, we observe that the [incidence rate](#) among workers in the general population is the same as that of exposed workers at our study site. But, because the non-workers in the general population have a [rate](#) that is five times as great as workers, this results in the overall [rate](#) in the general population being greater than that of exposed workers.

As a consequence, any study comparing [rates](#) of disease X between exposed workers and the general population would give a [biased](#) estimate (with the exposed workers having a substantially lower [rate](#) of disease X than the general population), due to the 'healthy worker effect' [selection bias](#).

Two *components of HWE bias* have been suggested [\[3\]](#):

1. *healthy worker hire effect*: the selection of healthier workers at hire, either due to self-selection (e.g. perceived health status) or employer selection (e.g. healthier subjects at lower risk of disease being employed preferentially)
2. *healthy worker survivor effect*: once hired, less healthy workers are more likely than healthy co-workers to leave high-exposure jobs, either by ending employment or being transferred out. While this selection away from exposed jobs may reduce the impact of exposure in a given patient (protecting that person's health), it may lead to the false ([biased](#)) conclusion that the higher-exposure jobs are safe.

Factors that determine the *size of the HWE bias* [\[3\]\[4\]](#) have been identified for mortality studies (some of which may also affect this [bias](#) in morbidity studies), and include:

1. sociodemographic factors: gender, age at hire, ethnic group, community unemployment rate
2. employment factors: occupational class, length of employment, time since hire/length of follow-up, time since termination
3. outcome factors: cause of death

Efforts should be made to [avoid bias from the HWE](#).

Diagnostic bias

Diagnostic [bias](#) can also occur in [cohort studies](#) if the diagnosis depends on the knowledge of the exposure status.

Example: in a [cohort study](#) of risk factors for mesothelioma, understanding that identification of mesothelioma is based on a difficult histological diagnosis, histopathologists may be more likely to diagnose a biopsy as mesothelioma if a history of asbestos exposure is reported.

Non-response bias

In a [cohort study](#), non-response matters only if it is associated with both the exposure and the outcome/ disease (see also [non-response bias in case-control studies](#)). Efforts should be made to [prevent non-response bias](#).

Example: the table below illustrates the results of a hypothetical [cohort study](#) where the following scenarios occur:

1. all exposed and unexposed participate in the study (i.e. no non-response)
2. non-response is associated with outcome (being a case)
3. non-response is associated with exposure
4. non-response is associated with both exposure and outcome (being an exposed case)

All respond					
	Total	Cases	Non-cases	Rate / 1000	Rate ratio (RR)
<i>Exposed</i>	10,000	100	9,900	10	10
<i>Unexposed</i>	10,000	10	9,990	1	reference
Non-response among cases (only 10% respond)					
	Total	Cases	Non-cases	Rate / 1000	Rate ratio (RR)
<i>Exposed</i>	9,910	10	9,900	1	10
<i>Unexposed</i>	9,991	1	9,990	0.1	reference
Non-response among exposed (only 10% respond)					
	Total	Cases	Non-cases	Rate / 1000	Rate ratio (RR)
<i>Exposed</i>	1,000	10	990	10	10
<i>Unexposed</i>	10,000	10	9,990	1	reference
Non-response among exposed cases (only 10% respond)					
	Total	Cases	Non-cases	Rate / 1000	Rate ratio (RR)
<i>Exposed</i>	9,910	10	9,900	1	1
<i>Unexposed</i>	10,000	10	9,990	1	reference

Loss to follow-up

This [bias](#) reflects differences in completeness of follow-up between comparison (exposure) groups i.e. exposed and unexposed. It is a problem for [cohort studies](#) as the length of time a [cohort](#) needs to be followed up can make it

difficult to follow all subjects until the end of the study e.g. due to people moving, losing contact etc. If subjects are lost randomly (in both exposure groups), this does not create loss to follow-up [bias](#) [5] (we will just have a smaller sample size/ study population on which to base our RR calculation, and wider confidence intervals [5]).

[Loss to follow-up bias](#) occurs if the loss of follow-up is associated with both exposure and outcome e.g. associated with exposed cases. It behaves similarly to [non-response bias](#) in [cohort studies](#). Differences in loss to follow-up between exposure groups can lead to [bias](#) as the people who are lost to follow-up may be more (or less) likely to have developed the outcome of interest [1].

Example: in a [cohort study](#) looking at smoking as a risk factor for development of lung cancer, [loss to follow-up bias](#) occurs if smokers who have lung cancer are more likely to be lost to follow-up (e.g. if they are more likely to die from lung cancer) than non-smokers with lung cancer.

Loss to follow-up among exposed cases (50% smokers with lung cancer lost to follow-up)					
	Total	Cases	Non-cases	Rate / 1000	Rate ratio (RR)
<i>Exposed (smokers)</i>	955	45	910	47	4.7
<i>Unexposed (non-smokers)</i>	1,000	10	1,000	10	reference

References

1. Bailey L, Vardulaki K, Langham J, Chandramohan D. Introduction to Epidemiology. Black N, Raine R, editors. London: Open University Press in collaboration with LSHTM; 2006.
2. Rothman KJ. Epidemiology - An Introduction. New York: Oxford University Press; 2002.
3. Le Moual N, Kauffmann F, Eisen EA, Kennedy SM. The healthy worker effect in asthma: work may cause asthma, but asthma may also influence work. *Am J Respir Crit Care Med*. 2008 Jan 1; 177(1):4-10. Epub 2007 Sep 13.
4. Baillargeon J. Characteristics of the healthy worker effect. *Occup Med*. 2001 Apr-Jun;16(2):359-66.
5. Giesecke J. Modern Infectious Disease Epidemiology. 2nd ed. London: Arnold; 2002.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Antigen presenting cells (APC)

Last modified at 11/30/2013 12:40 PM by Arnold Bosman

[Most of the links from this page still refer to Wikipedia, where a good overview of immunology is present. FEMWIKI looks for editors for the immunology chapters, to align them better to the content of ECDC training content for junior, mid-career and senior experts in disease prevention & control]

The human innate immune system includes cellular defenses, of which the group of Antigen Presenting Cells (APC) are an important part.

T cells cannot recognize (and therefore cannot respond to) 'free' antigen. T cells can only 'see' an antigen that has been processed and presented by cells via carrier molecules like MHC and CD1 molecules. Most cells in the body can present antigen to CD8⁺ T cells via MHC class I molecules and, thus, act as "APCs"; however, the term is often limited to specialized cells that can prime T cells (i.e., activate a T cell that has not been exposed to antigen, termed a *naïve T cell*). These cells, in general, express MHC class II as well as MHC class I molecules, and can stimulate CD4⁺ ("helper") cells as well as CD8⁺ ("cytotoxic") T cells, respectively. (Almost all nucleated cells express MHC class I receptors, including professional APCs. If a virus infects a macrophage or dendritic cell, it will try to promote its own destruction through cytotoxic T cells. However, dendritic cells can ingest viruses through pinocytosis and therefore activate the adaptive immune response to create antibodies for the virus through class II MHC receptors.)

To help distinguish between the two types of APCs: '*professional*' and '*non-professional*', those that express MHC class II molecules are often called **professional antigen-presenting cells**.

Professional APC's

Professional APCs are very efficient at internalizing antigen, either by phagocytosis or by receptor-mediated endocytosis, and then displaying a fragment of the antigen, bound to a class II MHC molecule, on their membrane. The T cell recognizes and interacts with the antigen-class II MHC molecule complex on the membrane of the antigen-presenting cell. An additional co-stimulatory signal is then produced by the antigen-presenting cell,

leading to activation of the T cell. The expression of co-stimulatory molecules is a defining feature of professional APCs.

There are three main types of professional antigen-presenting cell:

- [Dendritic cells](#) (DCs), which have the broadest range of antigen presentation, and are probably the most important APC. Activated DCs are especially potent T_H cell activators because, as part of their composition, they express [co-stimulatory](#) molecules such as [B7](#).
- [Macrophages](#), which are also CD4⁺ cells and are therefore also susceptible to infection by [HIV](#) as HIV invades immune cells through CD4⁺ receptor interactions.
- Certain [B-cells](#), which express (as B cell receptor) and secrete a specific antibody, can internalize the antigen, which bind to its BCR and present it incorporated to MHC II molecule, but are inefficient APC for most other antigens.
- Certain activated [epithelial cells](#)

Non-professional

A non-professional APC does not constitutively express the Major Histocompatibility Complex class II ([MHC class II](#)) proteins required for interaction with naive T cells; these are expressed only upon stimulation of the non-professional APC by certain cytokines such as [IFN-γ](#). Non-professional APCs include:

- [Fibroblasts](#) (skin)
- [Thymic epithelial cells](#)
- [Thyroid epithelial cells](#)
- [Glial cells](#) (brain)
- Pancreatic [beta cells](#)
- Vascular [endothelial](#) cells

Interaction with T cells

After APCs have phagocytosed pathogens, they usually migrate to the vast networks of [lymph vessels](#) and are carried via lymph flow to the draining [lymph nodes](#) (this network is collectively known as the [Lymphatic system](#)). The lymph nodes become a collection point to which APCs such as dendritic cells (DCs) can interact with T cells. They do this by [chemotaxis](#), which involves interacting with [chemokines](#) that are expressed on the surface of cells (e.g., endothelial cells of the [high endothelial venules](#)) or have been released as chemical messengers to draw the APCs to the lymph nodes. During the migration, DCs undergo a process of maturation; in essence, they lose most of their ability to further engulf pathogens, and they develop an increased ability to communicate with T cells. Enzymes within the cell digest the swallowed pathogen into smaller pieces containing [epitopes](#), which are then presented to T cells using MHC.

Recent research indicates that only certain epitopes of a pathogen are presented because they are immunodominant, it seems as a function of their binding affinity to the MHC. The stronger binding affinity allows the complex to remain kinetically stable long enough to be recognized by T cells.

References:



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Antigens (Ag)

Last modified at 11/30/2013 12:50 PM by Arnold Bosman

Large molecules, anything that stimulate a specific immune response.

In immunology, an antigen is the substance that binds specifically to the respective antibody. The term antigen originally came from **ANTI**body **GEN**erator.

The antigen may originate from within the body or from the external environment. "Self" antigens are usually well tolerated by the [immune system](#), which has been educated to non-reactivity against the structures present inside the body under the physiological conditions. "Non-self" antigens can be identified as invaders from the outside world or modified/harmful substances

References:

Wikipedia: <http://en.wikipedia.org/wiki/Antigens>



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Antibodies

Last modified at 11/17/2013 9:21 AM by Arnold Bosman

Immunoglobulins

The antibodies that our [adaptive immune system](#) produces are part of a class of proteins that we call 'immunoglobulines'. These have 5 different classes:

IgG: small immunoglobulin, in a way the basic antibody group, that comprises the largest group of immunoglobulins. The IgG has a constant part and 2 arms with highly variable composition. The variable part is produced in such a way that it can exactly fit (bind) to the epitope of an antigen. Due to the small size, this class of antibodies can pass through the placenta, and hence transfer antibody protection to the fetus, which will protect the newborn child for the first 3 months of its life, against all the infections that the mother has developed immunoglobulins against.

IgM: large immunoglobulins, composed of 5 antibody molecules (=pentamer). They can bind many more antigens at the same time, since they have 10 arms with highly variable antigen binding sites. These immunoglobulins are usually associated with the primary stages of the infection. These IgM immunoglobulins are so large, that they cannot pass through the placenta.

IgA: small molecules, associated mainly with mucosal immunity, hence we find large quantities in the gut and respiratory tract. The molecules can present as a dimer (i.e. 2 antibody molecules bound together) or as a monomer, which resembles the basic IgG molecule.

IgE: a monomer immunoglobulin, related to allergic responses and to protection against parasites.

IgD: not so well known. Not to be found free in serum, but found on the plasma membrane of B cells.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Vaccination

Last modified at 9/14/2011 3:56 PM by CeRC

This article is a result of new content structure of the FEM Wiki.

You are invited to contribute suitable content such as definitions, scope, examples and other related material. Feel free to link to other external resources.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

How pathogens try to trick our defenses

Last modified at 9/23/2018 10:06 AM by Katharina Kopp

The microbiome

Microorganisms have several major advantages over us, such as their incredibly huge numbers and their ability to adapt very rapidly (bacteria can create offspring every 20 minutes). Imagine a bowl of pudding that is left outside the refrigerator: if it initially contains only 50 bacteria (which is usually considered a minute amount that will not easily make us sick), then after 3 hours at room temperature there may be already 25.000 bacteria teeming in the pudding and another 2 hours later over 1.5 million !

In order to get an impression of the staggering numbers of microorganisms around us, consider that each human carries a huge population around on and within our bodies. This population of microbes is called 'the microbiome', and consists of 10 times more individual organisms than the number of cells we have in our body and a gene pool of 100 times bigger than our human genes. The average human being carries less than a kilogram of microbes around. And to be fair to these organisms: without them we would not be able to survive, since they work closely together with us to process our food, to produce vitamins that we are unable to synthesize ourselves and without which we would die.

The microbiome also makes sure that most of the living space for microorganisms in our bodies is already occupied, so that invading pathogens will meet with a lot of competition from our 'friendly microbes'.

Tricking our defenses

In the chapter of [applied immunology](#) we already explained the major lines of defenses within our immune systems. At the start of this page we have seen how fast microbes can multiply and with each multiplication step, some genetic errors may occur, some of which will kill the microbe, yet in rare occasions the genetic change makes the microbe stronger against our defenses. Such offspring will have a better chance of surviving in our body, and pass their new genetic trait to millions of offspring.

One way to evade our defenses is to avoid being detected. Some microbes develop the ability to cover themselves with bits of our own natural [antigens](#), to make our defenses believe that they are not foreign, yet a regular part of our body. Others are able to hide in cells and tissues where our immune surveillance cannot reach them so well. Some even hide in the cells of the immune system themselves (such as HIV). Others do not worry so much of being detected, because they continuously change their coat with new antigens, so that our immune system has to build up new memory (such as influenza virus).

For example, take a look at worms that can live in our bodies for up to decades. They are not really small (several inches in length) and they are clearly foreign to our body. Some of them produce [chemicals](#) that render our defenses harmless.

Then we have pathogens that attack our weapons: some destroy our antibodies, some kill our T-cells or B-cells. Some produce decoys, that keep our immune response busy, while the real pathogen can work undisturbed.

Microbes that change their identity

Diseases such as influenza, malaria and sleeping sickness all are caused by pathogens that are quite able to change their antigenic fingerprint and fool our defense mechanism. Our immune system has to recognize 'self' from 'non-self', by identifying the antigens on the surface of pathogens. It was already discovered early that the parasite that causes sleeping sickness (*Trypanosoma brucei*) is able to change the antigens on its surface, hence evading a targeted immune response. This parasite is able to do that, by switching on and off certain parts of its genome that code for the various [antigens](#) and even by further changing their genetic code during the infection to create even more variant antigens. This is also the reason why we have not been able to develop an effective vaccine against sleeping sickness. The same is true for malaria parasites. This poses one of the great challenges in infectiology science.

The influenza virus approaches its identity changes differently. It does not have a collection of different genes for different antigens, yet instead it changes the antigens through gradual mutation from one generation to the next. This causes the influenza antigen to slowly 'drift' out of the reach of our immunological memory. This is also the reason why we need a new influenza vaccine against seasonal flu each year.

Then there is one more trick that Influenza virus has up its sleeve: it has the ability to exchange its entire set of genes that code for one particular class of antigens with another influenza virus. When that happens, the antigenic fingerprint of flu 'shifts' instantly to another. Such a shift could lead to an entirely new type of flu virus that the world has not yet seen: this could trigger a worldwide new epidemic, or pan-demic. Without such drastic shifts of the major antigenic proteins, the flu virus would have huge problems to keep circulating.

references:

- Epidemics - the Dynamics of Infectious Diseases by Dr. Marcel Salathé, Dr. Ottar N. Bjornstad, Dr. Rachel A. Smith, Dr. Mary L. Poss, Dr. David P. Hughes, Dr. Peter Hudson, Dr. Matthew Ferrari, Dr. Andrew Read.
[Coursera course by PennState university.](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Defining a Case

Last modified at 5/9/2014 10:32 AM by Arnold Bosman

Case definitions will allow an objective categorisation of individuals based on their disease status, which facilitates a comparison of disease occurrence between populations, in different location and at different times.

A typical case definition includes:

- **clinical** criteria and **laboratory** findings to characterise the disease,
- a clear **time** period within which we count cases,
- a precise identification (**personal** characteristics) of the population from which we count cases and its **location**.

Therefore: the **disease**, the **time**, the **place** and the **person**.

For example the case definition used in a school outbreak of Measles was: **"A case patient was a child of School A, aged 5 to 15 years, with an illness characterised by a generalised rash lasting greater than or equal to 3 days AND a temperature greater than or equal to 38.3°C AND one of cough, coryza, or conjunctivitis; with onset of symptoms between 15 February and 28 March 2001."**

The following table is a line listing of signs and symptoms of 11 children during the epidemic. It illustrates application of case definition criteria to count cases.

Child ID number	Age in years	Rash duration	Temperature in ° celsius	Cough	Coryza	Conjunctivitis	Onset date	Case
1	5	2	38.0	Yes	No	No	15/02	NO
2	12	0	39.0	Yes	Yes	No	17/02	NO

3	16	4	38.5	Yes	Yes	Yes	17/02	NO
4	14	5	38.7	Yes	No	Yes	18/02	YES
5	8	3	39.2	No	Yes	Yes	19/02	YES
6	9	2	38.0	Yes	Yes	Yes	19/02	NO
7	4	0	38.5	Yes	No	No	20/02	NO
8	13	0	37.0	Yes	Yes	No	21/02	NO
9	10	5	38.5	Yes	Yes	Yes	29/02	NO
10	11	3	39.0	No	No	Yes	27/02	YES
11	3	3	38.2	Yes	Yes	Yes	26/02	NO
Etc.								

The table also demonstrates the difference between 'case' and 'patient'. It is clear that there are many more 'patients' in this table: in fact all people have symptoms of some illness. However, for the particular case definition that we have chosen here, only 3 count as a case. This can sometimes lead to heated discussions with clinicians, who want to know why their patient is not included as a case in the study. Such question is valid. Consider child number 6 in the table. The only reason that this child is not a case, is because there is no temperature over 38.5; is that a good justification to exclude this child as a case? The answer is: 'it depends'. It depends on the purpose of your study: what is it that you want to achieve?

The epidemiologist always needs to be able to justify why a certain case definition is chosen and what is the rationale behind each of the criteria.

An important question that we have to keep asking ourselves is:

- How much relevant information do I lose if the case definition is more specific or more sensitive?

During the stage of an [analytical study](#), we want the case definition to be as [specific AND sensitive](#) as possible, to avoid [misclassification](#) (case / non-case), since misclassification [biases](#) the study results towards the [null-hypothesis](#). In that sense, take another look at the table above: is it really justified that patient nr 3 is not counted as a case? This patient does not have the right age (5-15 years), however has all other elements of the case definition. A solution may be, to allow [various levels of certainty in a case definition](#) (possible, probable, confirmed).

Depending on where we are in the investigation, we need to be willing to take the responsibility to modify the case definition, in order to minimise bias. Though this is the ultimate responsibility of the lead investigator, different experts in the [outbreak team](#) may have relevant contributions to the discussion: epidemiologist, microbiologist, clinician, etc.

Table format and Analysis

Last modified at 6/3/2014 11:31 PM by Vladimir Prikazsky

Table format for an individually matched study

The results of matched case control study performed during a Echovirus outbreak in Germany in 2001 [1] are outlined in the [table](#). The study hypothesis was that swimming in a specific pond (pond A) was suspected of increasing the risk of aseptic meningitis due to an Echovirus. In the study, each case was matched to a single control ([individual matching](#)), therefore constituting a matched pair.

Table. Cases of Echovirus meningitis and controls according to swimming in pond A, Hesse, Germany, 2001, according to a matched table showing pairs.

		Controls		Total
		Exposed to pond A	Unexposed	
Cases	Exposed to pond A	194	46	240
	Unexposed	6	29	35
	Total	200	75	275

The figures inside the four cells of the two-by-two table no longer count individuals but pairs. We have therefore 275 pairs in the study including 275 cases and 275 controls. A pair in which both case and control have the same exposure is called a concordant pair. Alternatively if exposure differs between a case and its matched control they constitute a discordant pair.

In the [table](#) pairs are distributed as follows:

- 194 concordant pairs in which both the case and the control are exposed
- 29 concordant pairs in which both the case and the control are unexposed
- 46 discordant pairs in which the case is exposed and the control is not
- 6 discordant pair in which the case is unexposed and the control is exposed

As a general rule, in order to distinguish an unmatched two-by-two table from a matched table showing pairs, the letters (e, f, g, and h) are used rather than (a, b, c, d) to identify the discordant and concordant pairs.

	Cases	Controls	Total
Exposed to pond A	e	f	e+f
Unexposed	g	h	g+h
Total	e+g	f+h	e+f+g+h

It is important to draw such a [table](#) showing the pairs in order to get familiar with our data. It is also important to draw such a table if we need to do an analysis by hand.

From the [table](#) showing pairs we can reconstruct the table we would have had if doing an unmatched analysis of the same data. We can see that the marginal totals of the matched table correspond to each of the 4 inner cells of the unmatched table. However only the matched table is appropriate when analysing a matched study.

Table. Cases of Echovirus meningitis and controls according to swimming in pond A, Hesse, Germany, 2001, according to an unmatched table

	Cases	Controls	Total
Exposed to pond A	240	200	240
Unexposed	35	75	110
Total	275	275	550

Analysis of an individually matched case control study

In matched case control study odds ratio are calculated, just as in any case control study. Matched odds ratio are calculated by using pairs instead of individuals. We have seen from the [table](#) that a pair is equivalent to a stratum and that a matched analysis is a stratified analysis. We will therefore conduct a stratified analysis in which there will be as many strata as pairs.

The [following tables](#) illustrates the four types of strata one can have in a matched pair analysis. Taking the example the [table](#) we would then have:

- 194 strata with a concordant pair of type e,
- 46 strata with a discordant pair of type f,
- 6 strata with a discordant pair of type g
- 29 strata with a concordant pair of type h.

Since we are doing a stratified analysis we will use the Mantel Haenszel method to calculate an odds ratio.

Formula:

$$OR_{MH} = \frac{\sum(\frac{a_i d_i}{T_i})}{\sum(\frac{b_i c_i}{T_i})}$$

Tables with situation e, f, g, h

Situation e (case and control are both exposed, concordant pairs)

	Case	Control	Total
Exposed	1	1	2
Unexposed	0	0	0
Total	1	1	2

In situation e, the calculation using the formula yields to 0:

$$ad = 1 \times 0 \quad bc = 1 \times 0 \quad T = 2 \quad ad/T = 0/2 \quad bc/T = 0/2$$

Situation f (case is exposed and control is unexposed, discordant pairs)

	Case	Control	Total
Exposed	1	0	1
Unexposed	0	1	1
Total	1	1	2

In situation f, the calculation using the formula yields to 1/2:

$$ad = 1 \times 1 \quad bc = 0 \times 0 \quad T = 2 \quad ad/T = 1/2 \quad bc/T = 0/2$$

Situation g (case is unexposed and control is exposed, discordant pairs)

	Case	Control	Total

Exposed	0	1	1
Unexposed	1	0	1
Total	1	1	2

In situation g, the calculation using the formula yields to 1/2:

$$ad = 0 \times 0 \quad bc = 1 \times 1 \quad T = 2 \quad ad/T = 0/2 \quad bc/T = 1/2$$

Situation h (case is unexposed and control is unexposed, concordant pairs)

	Case	Control	Total
Exposed	0	0	0
Unexposed	1	1	2
Total	1	1	2

In situation h, the calculation using the formula yields to 0:

$$ad = 0 \times 1 \quad bc = 0 \times 1 \quad T = 2 \quad ad/T = 0/2 \quad bc/T = 0/2$$

From the [above tables](#) and from the [Mantel Haenszel formula](#) for the odds ratio, we understand that concordant pairs (e and h) contribute neither to the numerator nor to the denominator of the OR_{MH} .

Each discordant pair of type f contributes for 1/2 to the numerator and each discordant pair of type g contributes for 1/2 to the denominator of the OR_{MH} . The OR_{MH} calculated from the [example](#) on the Echovirus matched case control study is therefore:

$$OR_{MH} = \frac{\sum(\frac{a_i d_i}{T_i})}{\sum(\frac{b_i c_i}{T_i})}$$

$$\text{which is equal to } = \frac{\frac{1}{2} \times 46}{\frac{1}{2} \times 6} = 46/6 = 7,7$$

In other words the OR_{MH} is the ratio of the number of discordant pairs in which the case is exposed (f) over the number of discordant pairs in which the case is not exposed (g).

$$OR_{MH} = f / g$$

When more than one control per case are selected, the same principle applies. Let's supposed we selected two controls per case. For the stratified analysis each stratum includes therefore 3 individuals (the case and its two

controls). This leads to the 6 following possibilities for that type of strata.

From a triplet (1 case and 2 controls) we in fact constitute two pairs, the case with the first control and the same case with the second control. They are discordant or not with respect to exposure. The OR_{MH} will here also be the ratio of the sum of the discordant pairs in which the case is exposed over the sum of the discordant pairs in which the case is not exposed. Concordant pairs (in which the case and a control are either both exposed or both unexposed) do not contribute to the numerator nor to the denominator of the OR_{MH} .

Statistical softwares help us in the calculation of the matched OR. However what we need to do as epidemiologists is to "tell" the software that we are "interested" in a matched OR, and not in an unmatched OR.

Consequences of breaking the match (conducting an unmatched instead of a matched analysis)

One of the question frequently asked regarding matching is: "Why do we need to do a matched analysis since the groups we have created (cases and matched controls) are already equal with respect to the distribution of the [confounding factor](#)?"

This is because, by [matching](#), we have superimposed on the original confounding a selection bias that acts as an additional confounding. It is of course a very special type of [selection bias](#), because the investigator is fully aware that this is occurring. We therefore need to control for that additional confounding by performing a stratified analysis. Another way to look at this bias is to consider that since the controls are not randomly selected (but selected according to the matching criteria), they may no longer be representative of the population giving rise to cases. We therefore need to control for that selection bias. Failing to do so (i.e. doing an unmatched analysis) would usually bring the OR towards one.

The [matched Mantel Haenszel odds ratio](#) is controlling for the confounding effect we have introduced (selection bias) with the matched design. This is the reason why a matched analysis is required when matching. If the matched OR differs from the unmatched OR, this means we had introduced confounding when matching. We need to use the matched OR since it controls for the confounding we have introduced. If the matched odds ratio is equal to the unmatched odds ratio, this means that matching did not introduce confounding. It does not mean that matching was unnecessary. It does not mean that the confounding factor on which we have matched is not a confounder.

From matched [table](#) we have seen that the OR_{MH} was equal to 7.7 (46/6). If we break the matched pairs and do an unmatched analysis with the unmatched table is the unmatched OR is equal to:

$$OR = \frac{240 \times 75}{200 \times 35} = 2.57$$

The unmatched OR (2.57) is different from the matched OR (7.7) strongly suggesting that confounding was introduced by matched and that the OR_{MH} should be used to describe the results.

How to analysis sets with unequal numbers of controls per case

A matched case control study may have one to several controls per case, in the same study. This means that in a particular study some cases might be matched to one control, while other might be matched to two or more controls. This may happen easily, because controls might be difficult to enroll. In the stratified analysis each stratum includes one case and one to several controls. If each stratum does not include the same number of controls per case we could end up with strata including 2 individuals (1 case and one control), and strata with 3 individuals (1 case and two controls), etc. The computation of the stratum specific ad/T or bc/T would then change accordingly (T being equal respectively to 2 or 3, etc.). This could be a tedious process. Fortunately some statistical packages permit the analysis of unequal number of controls per case.

References

1. Hauri AM, Schimmelpfenning M, Walter-Domes M et al. An outbreak of viral meningitis associated with a public swimming pond. *Epidemiol Infect.* 2005 Apr;133(2):291-8.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Reference group for cohort studies

Last modified at 10/28/2010 7:56 PM by Lisa Lazareck

For cohort studies, the field epidemiologist is likely to be involved in retrospective studies. In other words the investigation takes place after both exposure and disease have occurred. The commonest situation is an outbreak of food poisoning after a clearly defined event such as a party or wedding. Following the same principles as for the case control study, it is first essential to define the source population. This population then forms the cohort, usually defined as those who attended the function in question. Individuals within the cohort are then classified into exposed or unexposed, for example, according to whether they ate or did not eat specified items of food or drink. The unexposed constitute the reference group for each item.

Questions may arise about whether the unexposed should include those who did not eat any food. As for case control studies, this depends on your definition of the source population. Is the cohort defined as everyone who attended or everyone who attended AND who ate something? The answer will be influenced by the hypothesis that we wish to test. As the number who did not eat anything will probably be small, it may be sensible to include them. If we should discover a substantial proportion of cases among those who attended but did not eat any food, food may not be the source of the outbreak.

What happens if everyone ate the food in question i.e. there is no unexposed group? Luckily for the epidemiologist, our investigations involve human behaviour which usually offers a rich variety of exposures. In a food borne outbreak where everyone ate the delectable tiramisu, we then rely on trying to measure different levels of exposure (different amount of Tiramisu consumed). The reference group then becomes those with the lowest level of exposure.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Developing a control definition

Last modified at 10/28/2010 7:56 PM by Lisa Lazareck

Developing a control definition

Controls should have the following characteristics

1. be representative of the exposure distribution in the source population
2. have an equal chance of being identified as cases if they had the disease under study
3. have the same exclusion and restriction criteria as cases

One way of developing a control definition is first to consider the case definition proposed for the investigation. In a hypothetical outbreak of *E. coli* O157, mainly among young children, mainly in London, during June 2010, the following case definition might be used.

Case: Resident of London aged under 10 years with faecal isolate of *E. coli* O157 during June 2010.

Exclusion: Travel abroad in the week before onset of illness.

1. The source population for cases is residents of London in June 2010 aged under 10 years. Controls should be representative of this source population.
2. Since *E. coli* is a severe infection of children, we would expect all children in London to have a similarly high chance of being detected as cases if they had this infection. However there may be variations in proportion of cases diagnosed by geographical area through variation in factors such as health seeking behaviour, primary care sampling, diagnostic facilities. This may introduce a selection bias when we come to choose controls as it will be difficult to identify this same source population. This bias will not matter unless the proportion exposed differs between cases identified for our study and those cases who remain undetected.
3. In this definition cases have been excluded if they travelled abroad in the week before onset of illness. An equivalent suitable exclusion period for controls might be travel abroad in the week before interview. However, if cases

mostly arise during school term, and if controls are interviewed in the summer holidays, some controls may be excluded unnecessarily. Another option might be to exclude those who travelled abroad in June. Or, if individually matched on potential time of exposure, travel exclusion could be restricted to the dates of the week before onset of illness in the matched case.

A suitable control definition might be:

Control: Resident of London aged under 10 years during June 2010.

Exclusion: Travel abroad in the week before interview.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Control Selection

Last modified at 10/29/2010 5:42 PM by CeRC

Let us consider the important decision on selecting controls [\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#).

Two broad types of control groups can be considered, unmatched and matched.

In unmatched or **population controls**, the aim is to obtain a random sample of the population that gives rise to cases. One suitable method is to seek controls from a population register. A random sample of the source population should be achievable if the register has a high level of completeness, contains the cases (it should be possible to check that all the cases are identified in the register), can identify the parameters specified in the control definition (e.g. area of residence, age), and is accessible to the investigator

If a register is not available or is not suitable, other methods of population sampling can be considered. A commonly used method is **random digit dialling** [\[6\]](#). This involves phoning random numbers (cold calling), a system that has the advantage of speed and convenience but has important limitations. The source population is limited to those who have a phone and to those who are available to answer. It may be difficult to be sure that the relevant geographical area is covered, or alternatively one may find that such a large area is covered by the phone listings that it is difficult to find controls from the (smaller) source area. This is more of a problem if phone numbers are used that do not have an area code e.g. mobile phone numbers. Co-operation from those receiving such calls may be low.

Hospital controls are a type of population control that can be used if the cases have all been admitted to hospital. Controls are easily identified and available at low cost from the same dataset that contains the cases e.g. hospital episode statistics. Disadvantages may be that there are different catchment populations for different diseases so that the controls are not representative of the source population for the cases. More particularly the same causative factors can be responsible for the disease under study and other diseases that result in hospital admission. This will reduce the chances of showing a true association with the causative factor (bring the OR towards 1). In the study of any disease caused by smoking, selection of hospital controls would have a high chance of selecting people who were admitted with other conditions caused by smoking [\[7\]](#).

The above examples of controls are all attempts to draw a random sample of the source population. The following examples are not attempts to draw a random sample of the population. Controls are selected because they have one or more characteristic in common with the cases. This method of selection is called [matching](#).

Neighbourhood controls. This involves selecting controls from the same neighbourhood as the cases i.e. they are matched for neighbourhood [\[8\]](#). One advantage is that there is no need for a population register. Also, controls are likely to be similar to cases in respect of socio-economic factors. This may be helpful if we wish to control for such complex factors and if we cannot measure them sufficiently.

Disadvantages are low co-operation (selection bias), it may be time consuming and expensive (low efficiency), and if we wish to measure the risk associated with socio-economic factors, we may not be able to do so. In case control study of a disease that has a socio-economic gradient, e.g. invasive meningococcal disease, picking neighbourhood controls may not show any association between illness and level of income. People living in the same neighbourhood are likely to have the same or similar socio-economic characteristics.

Friend controls are another way of selecting matched controls. Where speed of investigation is of the essence, eg. in a suspected outbreak of E.coli O157, friends offer a rapid and convenient means of finding controls. Similarity of socio-economic characteristics and social behaviours have the same advantages and disadvantages as neighbourhood controls. In investigations of outbreaks of food borne infection, our aim is to identify a common source. Although friends may be more likely to share similar food habits as their corresponding case leading to an underestimate of the strength of association, the relative risk estimates can still be very high [\[9\]](#). More of a problem may be a reluctance on the part of the case to give the names of friends to be interviewed [\[10\]](#).

Family controls (relatives) are rarely used in field epidemiology as exposures in family controls are often so similar to those of the cases that the association of interest may not be shown at all.

[Cases as their own control.](#) Cases act as their own controls in the case cross over method. Selecting controls in this manner is useful for diseases that have short incubation times and has the advantage of being efficient, achieving perfect matching and controlling for confounding on subject characteristics that are stable over time. A disadvantage, particularly in outbreaks caused by transient contamination of food or water, is that cases that are exposed to the contaminated product during their time as a case will also be more exposed to the product during their time as a control when that exposure was not associated with any increased risk. This will bias the OR towards 1.

[Controls with the same disease as cases.](#) In case-case studies, controls have the same disease as cases but are infected with a different subtype of the organism [\[11\]](#). For example cases of *Campylobacter coli* have been compared to *Campylobacter jejuni* cases to investigate potential risk factors [\[12\]](#). Controls can be selected randomly or through systematic sampling. Both selection bias and recall bias that occurs when cases are compared to healthy controls are removed through this method. Disadvantages are that general risk factors cannot be explored because their distribution will be similar in the two groups and cases and controls will differ for the exposure that led to them being infected.

The above are given as examples of different types of controls. Other [questions on selection of controls](#) may arise when considering different study designs, deciding numbers of controls, and what might happen when asymptomatic cases and immune subjects are included in the control group.

References

1. Rothmann KJ. Epidemiology: an introduction. Oxford University Press 2002.
2. Hennekens CH, Epidemiology in Medicine. Lippincott-Williams and Wilkins 1987.
3. Gregg MB. Field epidemiology. Oxford University Press 1996.
4. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case control studies I-III. Am J Epidemiol 1992; 135: 1019-50.
5. Grimes DA and Schulz KF. Compared to what? Finding controls for case-control studies. [Lancet](#).2005;365:1429-33
6. **Hartge P, Brinton LA, Rosenthal JF, Cahill JI, Hoover RN, Waksberg J. Random digit dialing in selecting a population-based control group. *Am J Epidemiol* 1984; 120:825–33.**
7. Cartwright RA, Adib R, Appleyard I, Glashan RW, Gray B, et al. Cigarette smoking and bladder cancer: an epidemiological inquiry in West Yorkshire. J Epidemiol Community Health. 1983; 37:256-63.
8. Werber D, Lausević D, Mugosa B, Vratnica Z, Ivanović-Nikolić L, et al. Massive outbreak of viral gastroenteritis associated with consumption of municipal drinking water in a European capital city. Epidemiol Infect. 2009; 137:1713-20.
9. Killalea D, Ward LR, Roberts D, de Louvois J, Sufi F et al. International epidemiological and microbiological study of outbreak of Salmonella agona infection from a ready to eat savoury snack - I: England and Wales and the United States. BMJ 1996; 313:1105-7.
10. Boccia D, Charlett A, Bennett S, Orr H, Sarangi J, Stuart JM. Outbreak of a new Salmonella phage type in South West England: alternative epidemiological investigations are needed. Commun Dis Public Health 2004; 7:339-43.
11. McCarthy and Giesecke. Case-case comparisons to study causation of common infectious diseases. Int J Epidemiol 1999; 28:764-8.
12. Sopwith W, Birtles A, Matthews M, Fox A, Gee S, et al. Investigation of food and environmental exposures relating to the epidemiology of Campylobacter coli in humans in Northwest England. Appl Environ Microbiol. 2010; 76:129-35.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Analysis by time characteristics

Last modified at 9/6/2012 12:10 PM by Vladimir Prikazsky

Temporal analysis of routine surveillance data uses graphical representation, usually plotted by time of onset.

However, the time of onset may give a false impression of a decreasing trend in the most recent time units because of reporting delays. If delays are known to occur, time of notification should be preferred.

Appropriate graphs for time data includes histograms (or joint bars) and line charts. The use of histograms must be limited to enumerated data (count of cases) and not used for rates for which line charts are preferred.

Figure 1: Distribution of viral hepatitis in Lebanon, by weeks, as of week 2003-15

- Analysis of case-based surveillance time characteristics is subject to limitations due to its nature. Surveillance is an evolving process that can be affected by:
- Change in case definitions as new tests become available
- Enrolment of new sources of data
- Increase in the completeness of reporting following a sensitization campaign
- Enhancement of surveillance in the event of an outbreak,

Each of these approaches induces limitations in the use of historical data as a baseline against which alerts can be detected.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Analysis by place characteristics

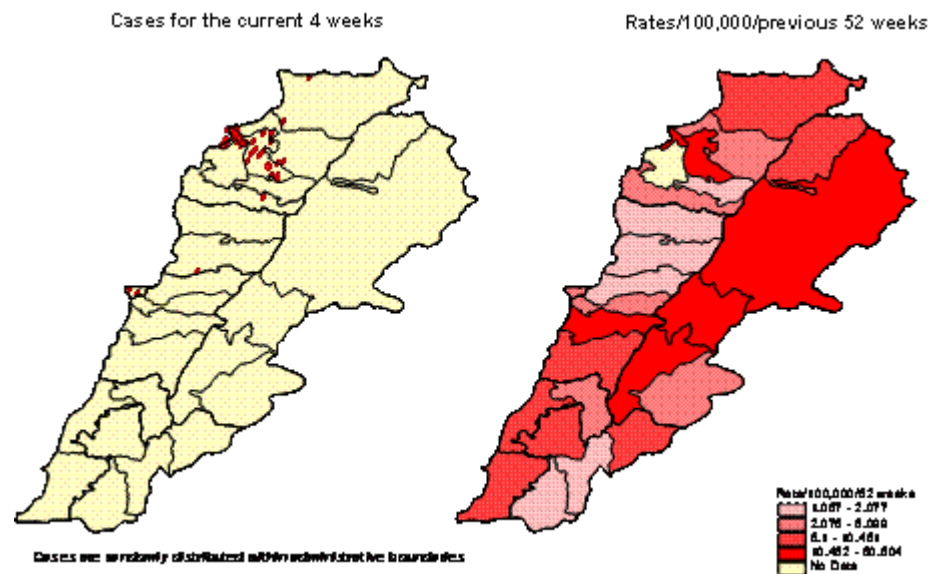
Last modified at 9/4/2015 7:31 AM by Vladimir Prikazsky

Plotting surveillance data by geographical location is an important step of the descriptive analysis. When done properly, this method enables the visualization of the geographical extension of a public health event, the identification of disease clusters, and the targeting of affected populations.

Data can be expressed in absolute numbers or [rates](#) (incidence, attack rates). Using rates allows for the comparison of distributions across geographic boundaries (based on places of disease notification). However, [standardization of rates](#) by age may sometimes be necessary in order to compare areas with different population structures (i.e. younger urban vs. older rural population).

The level of geographical analysis depends on the public health event under consideration. A cluster of a rare disease can only be described at the level of the affected community, while the national picture is of interest for diseases more commonly distributed.

Figure 1: Distribution of viral hepatitis in Lebanon, by district, as of week 2003-15.



Because they are available and denominators are known, administrative boundaries are often used for analyzing surveillance data. In addition, they reflect the structure through which the response to a public health crisis will be implemented. It is important, however, to remember that in some instances they may distort the true distribution of a disease in a population (e.g. when a city is split over several administrative boundaries, diluting any increase in reporting among several places).

Mapping data issued from sentinel surveillance system requires specific techniques in order to extrapolate the information from sentinel sites to the entire country[1].

See additional information on mapping surveillance data in chapters [Choosing an appropriate type of map](#) and [Which indicator to map](#)

Spatial distribution of diseases

The analysis of the spatial distribution of communicable disease is an important step in the routine analysis of surveillance data. In relation to early warning functions, it explores whether the spatial distribution of the disease follows some geographical pattern that may contribute to identify areas at increased risk of transmission. During outbreak investigations, mapping cases helps generate hypotheses about the mode of transmission.

The spatial distribution of communicable diseases relates to its mode of transmission:

- **Communicable diseases related to a point common source** are likely to show spatial clustering. For example, a cluster of infected people are likely to have been exposed to a common source such as a water supply system contaminated by enteric germs or a cooling tower contaminated with *legionella*.
- **Other communicable diseases transmitted from person to person** can also show spatial clustering when people are infected by being exposed to existing cases. Viral hepatitis and meningitis are examples of such diseases.

However, some communicable diseases are unlikely to show meaningful spatial patterns. This is the case for example for:

- **Communicable diseases associated with travel**, especially when the incubation period is long enough to allow the traveler to leave the area of exposure before becoming ill, e.g. viral hepatitis or travel associated Legionnaire disease.
- **Communicable diseases related to a widespread common source**, such as a food product having a broad distribution in a country.

Steps in mapping data

A map is an analog way to represent the spatial distribution of a disease. Making a meaningful map requires the following steps:

- Defining the [appropriate indicator](#) to represent the magnitude of the disease
- Selecting the appropriate variable to display the spatial disease pattern
- Choosing the [appropriate type of map](#)

References

1. Carrat F., Valleron A-J.: Epidemiologic Mapping Using the "Krigging" Method. Application to an Influenza-Like Illness epidemic in France. Am. J. Epidemiol., 1992; 135:1293-1300



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Analysis by person characteristics

Last modified at 3/15/2016 1:41 PM by Vladimir Prikazsky

Data on case characteristics, when available, are used to compare [incidence](#) or [rates](#) between various groups of cases, for example between different ages, between sexes or between different professions. This is useful to detect outbreaks which predominantly affect a particular group. The result of these comparisons can be expressed as a relative risk ([RR](#)) and its confidence interval.

When cases and deaths are notified, the case-fatality ([CF](#)) is often a useful indicator to compute. An increase in the case fatality may alert on an increase in the severity of the disease corresponding to a specific germ (e.g. cholera when diarrhoea is under surveillance as a syndrome).

Additional information on case characteristics, such as vaccine status, is not collected routinely since it is rarely necessary for triggering a public health alert. In most surveillance systems, this type of analysis is only performed at the "response" stage (investigation), and only on an *ad hoc* basis.

Analysis of case characteristics may be subject to [biases](#) as notified cases may not be representative of all cases in the population:

- Females and young children may be overrepresented since they are more likely than adult men to seek medical care in some settings, while in other male cases may be more likely to be attending health care facilities
- Milder cases of a condition may be underrepresented as such patients do not seek medical attention

These biases may affect the results of the analysis. Results must therefore be interpreted in the light of potential biases.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Significant probability to be different from the expec...

Last modified at 9/30/2015 12:12 PM by ikatsi

If we want to know if the results of a measurement are significantly different from the results that we should expect, then we first need to determine what we expect, and then define what we call significantly different.

What do we expect?

In our fictitious example of an outbreak of botulism among people eating in a restaurant, we want to investigate if eating home preserved green olives is significantly associated with developing botulism. Suppose we have performed a cohort study among 135 guests that ate in the restaurant where the outbreak of botulism occurred. Then the following 2x2 table shows the results of occurrence of botulism among the guests, for the exposed and the unexposed group:

Botulism outbreak in Restaurant X				
	Ill		Not Ill	
Ate Olives	9	43	52	
Did not eat olives	4	79	83	
	13	122	135	

How will we know what the probability is to find these results by chance / coincidence (so if really the null hypothesis is true and there is no association in reality)?

The first step is to determine what results we expected: we know that 13 (9.6%) of the 135 restaurant guests developed botulism. So if the olives are not the cause of the outbreak, then we expect that the occurrence of botulism will be the same (9.6%) among those who ate olives and those who did not eat olives. This means that we actually should expect the following table:

Expected occurrence of botulism

	in Restaurant X		
	Ill not Ill		
Ate Olives	5	47	52
Did Not Eat Olives	8	75	83
	13	122	135

The next thing we need to do is to quantify the difference between the observed results in our study and the expected values. For this we need the chi-square value: for each cell the expected number is subtracted from the observed number, this difference is squared and then divided by the expected number. The chi-square then sums the result for all cells. The formula is as follows:

$$\chi^2 = \sum \{(\text{observed num.} - \text{expected num.})^2 / \text{expected num.}\}$$

In our example, the $\chi^2 = 5.73$

What does this mean? The larger χ^2 is, the more the observed data deviate from the assumption of independence (no effect). Intuitively we assume that the larger the chi-square value, the lower the probability that our results differ due to chance and that our null hypothesis is **not true**. In that case we have evidence against H0 so that it can be rejected in favor of the alternative hypothesis. All we need to do now is to quantify the probability (p-value) that this chi-square value we observe is due to chance.

Is this difference significant?

In our example the p-value that corresponds to a $\chi^2 = 5.73$ with one degree of freedom is 1.6% (or: p=0.016). That sounds small, but is it small enough to be significant? Well, **significance is a convention**. And the most common convention is that a p-value equal or lower than 5% is considered significant. In other words we accept the decision to reject H0 if the probability that our results are due to chance, rather than to a true association is 5% or lower.

This means that in our example we assume that the probability that the difference in occurrence of botulism between those who ate olives and those who did not is mainly caused by chance (coincidence) is 1.6%. In other words highly unlikely that only chance can explain this difference.

Please note that this does not prove that the difference is caused by eating olives, just that we can likely rule out chance as an explanation. For further reading, see the chapter on [Causal Inference](#).



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Measures of effect in various study designs

Last modified at 1/26/2016 4:42 PM by Vladimir Prikazsky

Summary

Many epidemiologists consider that the studies they are conducting are measurement exercises. Simple studies include measurements of disease frequency which may be expressed as risks, rates, prevalence or odds. More advanced studies will aim at identifying the causes of diseases and the effect of specific exposures on disease occurrence. This is achieved by comparing disease frequency between sub groups of a population. This comparison can be expressed as a difference or a ratio, so called "effect measure".

A core function of epidemiologists is to measure the causal effect of an exposure on the occurrence of a disease. To measure a causal effect we would ideally have to compare occurrence of disease in exposed persons to what would have happened in the same persons, at the same time, in the absence of exposure. This is however theoretical since such two measurements, in the same group of persons under study, are not feasible during the same time period. In order to approach this theoretical situation as closely as possible, we will use as unexposed group a population similar to the exposed group but for the exposure. In these two populations (or in 2 subsets of the same population, exposed and unexposed), we will then measure and compare disease occurrence.

To compare disease occurrence between exposed and unexposed populations epidemiologists will have either to assign exposure or to observe populations naturally exposed. Assignment of exposure is only ethically feasible when exposure is potentially protective (treatment, vaccine, preventive measures). Observation of accidental or naturally assigned exposures will allow us to study the effect of potentially harmful exposures. To measure the effect of exposure several types of epidemiological studies are available. In [cohort studies](#) the frequency of disease is compared between a group of exposed and unexposed cohorts, while in [case control studies](#) the exposure status is compared between persons with and without disease. There are several [advantages and disadvantages of cohort and case control studies](#) which will be presented in this chapter. Furthermore, [case cross over studies](#) will be discussed in this chapter in which exposure information is obtained from the same case group but during different periods of time. The

[case to case study design](#), a type of case control study when the disease of interest can be subclassified in two or several groups that have specific risk factors, will be highlighted briefly.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Table measuring risk, rate and odds ratio

Last modified at 10/7/2010 11:20 PM by Masja Straetemans

Cohort studies allow to directly measuring risk or rate of disease occurrence and their related ratio in subgroups of a population (exposed and unexposed). Case control studies do not allow measurement of risk or rates. They however allow estimation of the risk ratio and the rate ratio. The selection of the control group is a crucial step of the study. The following table summarises the type of measures and controls selection as described in the above chapter.

Measuring risk, rate and odds ratios in a case control study, using various sampling methods for controls

(Source: *Rodrigues L et al. Int J Epidemiol. 1990;19:205-13*)

Measure	Definition	Case Control Formulation	Design	Controls sampled from
1-Risk ratio	$\frac{Ce}{Ne} \frac{Cu}{Nu}$	$\frac{Ce}{Cu} \frac{Ne}{Nu}$	Case cohort	Total study population present at beginning of study
2-Rate ratio (incidence density ratio)	$\frac{Ce}{PTe} \frac{Cu}{PTu}$	$\frac{Ce}{Cu} \frac{PTe}{PTu}$	Density case control	People at risk at time of case disease onset

3-Odds ratio	$\frac{\frac{Ce}{Ne-Ce}}{\frac{Cu}{Nu-Cu}}$	$\frac{\frac{Ce}{Cu}}{\frac{Ne-Ce}{Nu-Cu}}$	Traditional case control	People disease free throughout study period
--------------	---	---	--------------------------	---

Measuring risk

Last modified at 10/28/2010 8:41 PM by Lisa Lazareck

Cohort studies measuring risk (incidence proportion)

Cohort studies that measure risk compare occurrence of disease between exposed and unexposed cohorts. The risk (incidence proportion) of disease in those exposed (IPe) and unexposed (IPu) can be computed as follows:



In the above example IPe = Ce/Ne

$$IPu = \frac{\text{number of cases among unexposed}}{\text{unexposed population at risk at beginning of follow up}}$$

In the above example IPu = Cu/Nu

The absolute effect of exposure on disease occurrence is the risk difference (RD) between the exposed and unexposed cohorts.

$$\text{Absolute effect (RD)} = IPe - IPu \text{ (Also called : absolute risk reduction)}$$

The relative effect of the exposure on disease occurrence can be expressed as the risk difference between exposed and unexposed, divided by (relative to) the risk in unexposed.

$$\text{Relative effect} = \frac{\text{Risk difference}}{\text{Risk in unexposed}} = \frac{IPe - IPu}{IPu} = \frac{IPe}{IPu} - \frac{IPu}{IPu} = RR - 1$$

Where RR is the risk ratio defined as:

$$\text{Risk Ratio} = \frac{IP_e}{IP_u}$$

Example

Cases of gastroenteritis according to consumption of food X, nursing home A

Consumption of food X	Population at risk	Cases	IP	Risk Ratio	Relative effect
Yes	150	60	0.4	4	3
No	100	10	0.1		

One can express the result by saying that the relative effect of consuming food X is 3 which would suggest a 300% increased risk of gastroenteritis among exposed. One can also express the results by saying that the risk of disease is 4 times higher in the exposed cohort than in the unexposed cohort.

Thus the relative effect is the risk ratio minus 1. Since the relative effect is $RR - 1$, epidemiologists frequently refer to RR as a measure of relative effect without subtracting 1. The term "relative risk" is very popular among epidemiologists even if, as mentioned above, it is not a measure of relative effect but rather a risk ratio. When using the relative risk that way we have to remember that a value of 1 corresponds to an absence of effect.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Interaction

Last modified at 9/3/2014 5:08 PM by Arnold Bosman

Statistical and biological interaction

The term interaction is frequently used alternatively with effect modification. Statistical packages for multivariable analysis offer to test for interaction. They often are based on a multiplicative assumption. Interaction is present only if the joint effect of A and B is more than the multiplication of the respective effect of A without B by the effect of B without A. This is called "statistical interaction". It is assessed after logarithmic transformation.

Epidemiologists are interested in "biological interaction". Where there is no biological interaction between two exposures (A and B), the risks related to A and B are added to each other when both A and B are present. If biological interaction occurs, we expect the joint risk to be higher than the sum of A and B risks. The difference is attributable to the joint effect.

Formulae

In cohort studies the risk due to interaction between two exposures A and B can be calculated as follows [1]:

$$\text{Interaction} = R_{AB} - R_A - R_B + R_O$$

In which R_A is the risk when only exposed to A, R_B is the risk when only exposed to B, R_{AB} is the risk when exposed to both, and R_O the risk when exposed neither to A nor B.

If there is no interaction (i.e. the exposures are independent of one another) the expected risk when exposed to both factors can be computed as:

$$R_{AB} = R_A + R_B - R_O$$

Cohort Studies

During an outbreak of *Salmonella enteritidis* gastroenteritis two risk factors were suggested by the data, consumption

of undercooked chicken (exposure A) and taking anti-acid medications (exposure B). The risk of illness was respectively 5/1000 among those who were not exposed to any of the 2 risk factors, 10/1000 among those who took anti-acid medication but did not eat undercooked chicken, 20 among those who ate undercooked chicken but did not use anti-acid medication and 100/1000 among people eating undercooked chicken and taking anti-acid medication.

Exposures		Cases	Total	Risk
neither A nor B	No chicken, no antiacids	1	1000	0.001
A but not B	Chicken but not antiacids	20	1000	0.020
B but not A	Antiacids but not chicken	10	1000	0.010
A and B	Chicken and antiacids	100	1000	0.100

If there was no interaction between exposure to undercooked chicken and antiacids, the expected risk when exposed to both risk factors would be:

$R_{AB} = 0.02 + 0.01 - 0.001 = 0.029$

However, the observed risk is 0.100, suggesting a biological interaction between consumption of undercooked chicken and taking anti-acid medications. The joint risk is more than the simple addition of the two risks. The additional risk linked to the exposure to both undercooked chicken and antiacids is potentially responsible for an incidence of 71 cases per 1000, potentially explaining 71% of disease occurrence.

Case-control studies

In case control studies biological interaction can be measured using the following formula [2]. As risk cannot be computed in a case control study, the odds ratio (OR) is used instead:

$Interaction = (OR_{AB} - 1) - (OR_A - 1) - (OR_B - 1)$

Given no interaction (independence between the two exposures):

$OR_{AB} = (OR_A - 1) + (OR_B - 1) + 1$

In a national case control study looking at *Salmonella* enteritidis in children in France [3]

--	--	--	--	--

Season	Eggs	Cases	Controls	OR
Not Summer	< 2 weeks storage	32	36	ref
Not Summer	> 2 weeks storage	7	3	2.63
Summer	< 2 weeks storage	52.	64	0.91
Summer	> 2 weeks storage	12	2	6.75

The joint effect of A and B is therefore computed as:

$$OR_{AB} = (0,91-1) + (2,63-1) + 1 = 2,74$$

However, the observed value for OR_{AB} is 6,75.

Therefore, $6.75 - 2.74 = 4.01$ of the effect is due to biological interaction when both exposures are present. The biological interaction represents 60.7% ($4.01/6.75 * 100$) of the effect when both exposures are present.

In 1976, an outbreak of Ebola viral haemorrhagic fever occurred in the Bumba zone of Zaire (now Democratic Republic of Congo). The disease was amplified by exposure to a large, active hospital [4].

Hospital	Case	Cases	Controls	OR
Unexposed	Unexposed	41	266	ref
Exposed	Unexposed	85	22	25.1
Unexposed	Exposed	149	26	37.2
Exposed	Exposed	43	4	69.1

If there was no interaction between "exposure to hospital" and "exposure to a case", the OR associated with contact to both exposures:

$$OR_{AB} = (37.2 - 1) + (25.1-1) + 1 = 61.3$$

This is slightly different from 69.1 as shown in the table. We could conclude for the presence of very little additive interaction between the two risk factors. In addition we would need to explain the biological meaning of such an interaction, when it seems unlikely that cases in hospital would do much mixing - being too ill to socialize in that setting.

The above represents a simplified explanation of additive and multiplicative models when testing for interaction. Further explanation can be found in major text books [1].

Using a statistical package for multivariable analysis (based on a multiplicative assumption) could have lead to different conclusions. However it is possible to assess biological interaction with such package if using an additive model. For doing so we simply need to create four level of exposures (dummy variables) exposure to A and B, exposed to A and not B, exposed to B but not A, exposed to neither. The later exposure is used as reference for the 3 others and variables and biological interaction can be measures using additive assumptions as in the above examples.

References

1. K.J.Rothman, S.Greenland, T.L.Lash. Modern Epidemiology. Third ed. Philadelphia, USA: Lipincott Williams and Wilkins; 2008.
2. Kalilani L, Atashili J. Measuring additive interaction using odds ratios. *Epidemiologic Perspectives & Innovations* 2006;3(1):5.
3. Delarocque-Astagneau, E., Desenclos JC, Bouvet P, Grimont PA. Risk factors for the occurrence of sporadic *Salmonella enterica* serotype enteritidis infections in children in France: a national case-control study. *Epidemiol Infect* 1998;121(3):561-7.
4. Ebola haemorrhagic fever in Sudan, 1976. *Bulletin of the World Health Organisation* 1978; 56 (2): 271:93.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Causal mechanisms

Last modified at 9/15/2015 4:45 PM by Arnold Bosman

The mechanism behind a cause usually has several components. These can be subdivided into two categories: necessary and sufficient components [1]. Let us take the example of an infectious disease, invasive meningococcal disease. We say it is caused by *Neisseria meningitidis* (difficult to argue against, as this organism by definition must be present). So it is a "necessary" component of cause.

On the other hand, infection with *Neisseria meningitidis* does not always result in meningococcal disease. Indeed illness is a rare outcome of the infection. The infection by itself is not "sufficient" and other factors need to be present. Lack of antibodies against this infection, the breakdown of the mucosal barriers by respiratory infection, low humidity, and passive smoking may be other causal factors.

Thus USUALLY

- several components are needed to produce a given outcome,
- any one component is not sufficient on its own,
- different combinations of components can produce the same outcome

These concepts of necessary and sufficient component causes explain the apparent anomaly whereby attributable fractions in a population (AFpop) can add up to more than 100%. The AFpop for *Neisseria meningitidis* is 100%. If *Neisseria meningitidis* were to be eliminated as a coloniser of the human pharynx, there would be no more meningococcal disease. The AFpop for low immunity is probably also close to 100%. This means that vaccination may be highly effective at reducing disease rates even without reducing colonisation.

Various conceptual models aiming to simplify the representation of causal mechanisms have been developed in epidemiology. A well known model is that of infectious disease causation (the agent, host, environment pyramid). Another widely adopted model for chronic disease causation is Rothman's "sufficient component cause model" whereby the cause of any effect must consist of a constellation of components that act in concert [2].

References

1. Rothman KR. Epidemiology: an introduction. Oxford University Press, Oxford 2002.
2. Rothman K, Greenland S. Causation and causal inference in epidemiology. Am J Public Health. 2005;95:S144–S150.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

[Statistical Concepts](#)

Logistic Regression

Last modified at 6/7/2016 2:03 PM by testuser1

Summary

[Stratification](#) is one of the pillars of epidemiological analysis. It allows investigators to familiarise with the distribution of data according to the variables of interest, to estimate the effect of a variable adjusted by the effect of covariates or confounding factors and to study interaction or effect modification between two factors.

However, stratification is limited in the number of variables to be examined simultaneously because the number of subjects in each stratum may drop to 0 or 1 thus even statistical methods for dealing with sparse data may not be applicable.

Regression analysis overcomes this limitation by estimating regression models to approximate the function describing the relationship between dependent and independent variables.

The different regression analysis techniques are very efficient estimating the independent effect of several covariates and for the study of interactions. On the other hand, modelling data encompasses underlying assumptions. Researchers should be familiar with regression techniques and the interpretation of results to assure that underlying model assumptions are realistic. Researchers using regression analysis may lose track of patterns of data distribution and the process may not be well understood by the target audience.

A combination of both techniques, stratification and regression, is probably the best approach for the analysis of epidemiological data.

In this chapter, we will focus on "logistic regression models", a regression analysis technique suited for the analysis of case-control data.

Topics covered in this chapter include:

1. [Linear models](#)

2. The logistic model
3. Fitting logistic regression models
 1. Interpreting model coefficients
 2. Estimating Odds Ratios in the presence of interaction
4. Model building strategies



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Questionnaire Layout and Coding

Last modified at 5/26/2011 5:32 PM by CeRC

Questionnaire Layout

A questionnaire should have a clear consistent layout, leave adequate space to answer, bear a large font size and appropriate page breaks. Experimental layouts, fancy logos and disturbing features such as "printed on recycled paper/is an equal opportunity employer" should be avoided. In some cases, using colour or printing the questionnaire on coloured paper may help to increase the response. This depends again on the targeted audience and should be discreet and tasteful.

Filter questions may be of use if some questions or a group of questions are targeted to a subgroup of the respondents. This will enable interviewers a smooth flow through the questionnaire and shorten the time to fill in the questionnaire in self-administrated questionnaires. Similarly, it is important to give clear instructions to interviewers or respondents. This will reduce misunderstandings.

Questionnaire Coding

Closed questions on a paper questionnaire can be pre-coded. This means that a number is assigned in advance to each possible answer. Coding will enable a quicker and easier data entry. The ideal code numbers depend on the software which will be used for data analysis. Some software packages only accept 0/1 codes for dichotomous variables. In order to avoid time-consuming data cleaning, be careful to choose appropriate and consistent codes for all variables in advance.

Examples:

Male	X	1		Ill		1
Female		2		Not Ill	X	0
Don't		3		Don't		9

Know				Know		
------	--	--	--	------	--	--



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Seven Golden Rules to Design Questions

Last modified at 5/26/2011 5:32 PM by CeRC

It is very important to adjust the style of the questions to the target audience. The phrasing of the questions might vary substantially depending whether the target audience consists of medical professionals or the general public.

1. Ask one information at a time

Do you own a dog?

instead of Do you own a dog or have frequent contact with dogs?

2. Ask precise questions

How often did you touch a dog during the past 3 months?

instead of Do you often touch dogs?

3. Ask appropriate, non-judgemental questions

How often have you consumed alcoholic beverages during the past 6 months?

instead of Are you a drunk?

4. Avoid suggestive questions

Which beverage did you consume?

instead of Did you drink the strange pink drink?

5. Be as simple as possible

Did you smoke an average of 2 packages of cigarettes/week for the last 5 years?

instead of Did you smoke not less than a mean amount of 7 cigarettes/2 days from 1999 onwards?

6. Avoid jargon, abbreviations or slang

How often do you get up at night to pass urine?

instead of How often do you get up at night to PU?

Remember that key words need to be defined (example "PU", "fully vaccinated").

7. Use mutually exclusive and exhaustive answer options

Put the options in a vertical order. Do not forget the option "don't know, if applicable.

Yes		or	Every day		or	0-10 years	
No			Several times a week			11 - 20 years	
Don't Know			Less than once per week			21 years or older	



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Criteria for confounding

Last modified at 9/14/2011 1:23 PM by CeRC

Two conditions and two restrictions are necessary for a characteristic to be a confounding factor [\[1\]](#):

A confounding factor:

1. must be a proxy measure of a cause, in unexposed people
2. must be correlated (positively or negatively) with exposure in the study population. If the study population is stratified into exposed and unexposed groups, the confounding factor has a different distribution in the two groups
3. must not be an intermediate step in the causal pathway between exposure and disease
4. must not be an effect of the exposure

These four criteria must be verified whenever a characteristic is suspected of being a confounding factor. In the [previous example](#) the confounding factor (vaccination) is associated with both exposure (gender) and outcome (disease). Vaccination is not in any biological pathway between gender and disease and unvaccinated children have a higher risk of disease in both sexes. The two conditions and restrictions are met. The crude risk ratio was artificially increased by the unequal distribution of vaccinated among boys and girls and the fact that vaccination is a protective factor against disease.

To numerically identify a confounding factor the measure of the crude effect is compared to a summary measure of the effect. This is a weighted measure taking into account the stratum specific value of the effect (i.e. the RR in each stratum), attributes a weight to each (based on the size of the sample). If the weighted measure of effect differs from the crude measure, then the characteristic on which we have stratified our analysis (vaccination in this example) may be a confounding factor.

It must be considered, however, that there is no statistical method to compare a crude from a weighted measure of effect (a good rule of thumb: confounding may be considered if the difference between the two measures is more than 15-20%).

References

1. McNamee R. Confounding and confounders. Occupational and Environmental Medicine 2003 Mar 1;60(3):227-34.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Estimating Odds Ratios in the presence of interaction

Last modified at 9/19/2011 9:47 AM by Arnold Bosman

When interaction is present, the association between a risk factor and the outcome varies according to and depends upon the value of a covariate. Interaction between two variables can be positive (their joint role increases the effect) or negative (their joint role decreases the effect).

In logistic regression we will take interaction between two variables into account by adding to the model an interaction term. Let suppose we are studying the role of two exposures (tiramisu and beer) in the occurrence of gastroenteritis due to Salmonella.

The logit including an interaction between tiramisu and beer can be written as follows:

$$\text{Ln (P gastroenteritis / tiramisu, beer)} = \beta_0 + \beta_1 \text{ tiramisu} + \beta_2 \text{ beer} + \beta_3 (\text{tiramisu} * \text{beer})$$

The term $\beta_3 (\text{tiramisu} * \text{beer})$ reflects the interaction.

We have therefore 2 variables and four combinations of coefficients:

Table 1: Effects of different combination of exposures to tiramisu and beer

Tiramisu	Beer	Equations	Relative effect (RO)
0	0	β_0	Reference
1	0	$\beta_0 + \beta_1$	β_1
0	1	$\beta_0 + \beta_2$	β_2
1	1	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_2 + \beta_3$

The following table shows the results of the steps in the analysis of data when testing for interaction between consumption of Tiramisu and consumption of Beer on occurrence of gastroenteritis in our example.

Model	Constant (β_0)	Tiramisu	Beer	Tiramisu*beer	LRS	p-value
1	-2,9741	$\beta_1 = 4,3116$ OR = 74,56			180,3927	<0,001
2	-2,6740	$\beta_1 = 4,4097$ OR = 82,2419	$\beta_2 = -0,8895$ OR = 0,41		4,3210	0,0376
3	62,9704	$\beta_1 = 4,88$ OR =131,62	$\beta_2 = -0,0085$ OR = 0,99	$\beta_3 = -1,2079$ OR = 0,2988	1,6078	0,204

Model 1 tests the effect of consumption of tiramisu on the occurrence of gastroenteritis due to salmonella.

Model 2 suggests that beer plays a slight confounding effect ($p = 0,037$, OR changing from 74 to 82) for the association found in model 1. In model 3, the introduction of the interaction term (tiramisu*beer) suggest that there is interaction (negative) between consumption of tiramisu and consumption of beer. Beer seems to decrease the risk of illness due to tiramisu consumption. However this interaction is NOT statistically significant (LRS = 1,60 and $p = 0,2048$).

In the presence of interaction, the effect of the different combinations of exposures should be worked out as shown in table 1, using the coefficients ($\beta_0 + \beta_1 + \beta_2 + \beta_3$) estimated in the model including the interaction term (model 3).

The following table shows output of the logistic regression model including the interaction term (using a statistical package).

Number of terms	4			
Total Number of Observations	245			
Rejected as Invalid	0			
Number of valid Observations	245			
Summary Statistics	Value	DF	p=	value

Deviance	153,3200	241		
Likelihood ratio test	186,3215	4	< 0.001	
Parameter Estimates				95% C.I

Terms	Coefficient	Std.Error	p-value	Odds Ratio	Lower	Upper
%GM	-2,9704	0,5127	< 0.001	0,0513	0,0188	0,1401
TIRA_	4,8800	0,6374	< 0.001	131,6250	37,7339	459,1393
BEER	-0,0085	0,7830	0,9913	0,9915	0,2137	4,6006
BEER* TIRA_	-1,2079	0,9338	0,1958	0,2988	0,0479	1,8634

[<<Back to Logistic regression](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Preventing bias

Last modified at 5/26/2011 5:24 PM by CeRC

All epidemiological studies, even randomised clinical trials, are susceptible to [bias \(systematic error\)](#). The objective of the epidemiologist will be to minimise these [biases](#). This can be done by considering, at the different stages of development and execution of a study, where and how [bias](#) may occur: the design stage (protocol writing), subject selection (case/control, exposed/unexposed, intervention/control group etc), data collection, data analysis and interpretation of results.

At the design stage, [bias](#) should be considered at the time of protocol writing. A lot of care should be given, at this stage of development of the study, to forecasting all potential [selection](#) and [information biases](#) that may be encountered. Despite all precautions taken, some [biases](#) will persist. They then need to be taken into account in the interpretation of the results of the study.

When writing the report or manuscript, sources of potential [bias](#) in the study absolutely need to be openly discussed. Particularly, the first part of the discussion section of a scientific paper should include a detailed paragraph in which authors discuss all potential [biases](#) which could have falsely led to the study results. If possible, the *direction of the bias* (overestimation or underestimation) and the *magnitude of the bias* also should be discussed.

While [case-control](#) and [cohort studies](#) are both susceptible to [bias](#), the [case-control study](#) is affected by more sources of [bias](#). Through our study design, we can try to [minimise selection bias](#) and [prevent information bias](#) in [cohort](#) and [case-control studies](#).

How to minimise selection bias

In epidemiological studies, all efforts should be made to avoid [biasing](#) the selection of study participants. [Selection bias](#) can be reduced by paying attention to the following:

1. The study population should be clearly identified i.e. clear definition of study population.
2. The [choice of the right comparison/ reference group](#) (unexposed or controls) is crucial

- for example, in an *occupational cohort study*, rather than comparing workers with the general population (which includes people who are too ill to work), ensure all subjects in the comparison are workers, and avoid bias from the *Healthy Worker Effect* (HWE) [1]. Compare workers in a specific job with those in jobs that differ in occupational exposures or hazards e.g.
 - select an external comparison group from another workforce e.g. in a situation where all workers of an occupational *cohort* had some degree of exposure [2]
 - select an internal comparison group within the same workforce e.g. if some workers had exposure while others did not [2].
- 3. In a *cohort study*:
 - exposed and unexposed groups should be identical but for the exposure
 - in a *retrospective cohort study*, the selection of exposed and unexposed groups should be done without knowing the outcome (disease status).
- 4. In a *case-control study*:
 - the control group should reflect the exposure of the population which gave rise to the cases
 - controls should be selected independently of the exposure status
 - for example, *non-response bias* happens when participation into a study is related to the exposure status
 - precise *case definition* and exposure definition should be used by all investigators.
- 5. In an *intervention study*, select participants through randomisation, so that they have an equal chance of receiving the intervention.
 - this allocation to intervention and control groups should rely on a mechanism that is not within the control of the study participant or the investigator, termed 'allocation concealment' [2], thus avoiding a situation where the investigator might be more inclined to allocate sicker patients to the intervention/treatment arm of the study, and less ill patients to the control arm
 - whether the randomisation has been successful or not can be checked by comparing baseline factors between the intervention and control groups afterwards, and seeing if the groups are similar in all other respects apart from receiving the intervention [2].

Preventing non-response bias

Non-response bias can be prevented by achieving high *response rates* ($\geq 80\%$ by convention) [3]. High *response rates* may be facilitated by:

- offering incentives to participate in the study e.g. entry into a raffle for a prize
- making it easy to contribute e.g. by using questionnaires that are not too long and don't take too much time to complete (see the chapter on *Questionnaire Design* for further hints on creating a well-designed questionnaire)
- setting aside protected time for the study e.g. in a school-based questionnaire study, asking teachers to allow pupils to complete the questionnaire during a class period rather than giving them the questionnaire to take home with them
- sending reminders e.g. a first reminder by post at 1 week and a second reminder at 2 weeks after the initial questionnaire.

Information on characteristics of the non-responders should be obtained if possible e.g. by getting a subset of non-participants to complete a *non-response questionnaire* (NRQ), or by getting some demographic information on *non-respondents*, if this is possible (so that they can be compared with *respondents*). This can give important insights into the extent of *selection bias*. However, it should be noted that obtaining this information on *non-respondents* is time-

consuming and not always successful.

- For example, in a [case-control study](#) by Vrijheid et al of mobile phone use and development of brain tumour [4], [selection bias](#) factors were estimated based on the prevalence of mobile phone use reported by non-participants from [NRQ](#) data. In this particular example, non-participation in the study seemed to relate to less prevalent use of mobile phones, and the investigators estimated that this could result in an underestimation of the odds ratio for 'regular mobile phone use' by about 10% [4].

Preventing information bias

[Information \(measurement\) biases](#) can be easier to prevent and measure than [selection biases](#) [3].

They can be prevented by:

1. Using standard measurement instruments e.g. [questionnaires](#), automated measuring devices (for measurement of blood pressure etc)
2. Collecting information similarly from the groups that are compared
 - cases/ controls, exposed/ unexposed
 - several sources of information can be used to validate each other, but all sources should be used for each subject
3. Use multiple sources of information
 - [questionnaires](#) (e.g. postal/ online/ face-to-face via interview)
 - should favour closed, precise questions and avoid open-ended questions
 - test the same hypothesis using different questions
 - field-testing / [piloting of questionnaire](#) in order to improve and refine it
 - standardise interviewers' techniques through training (with the questionnaire) to ask questions the same way
 - direct measurements
 - registeries (e.g. cancer registeries etc)
 - case records (e.g. from GPs, hospital notes etc)

Preventing interviewer/ observer bias

- 'Blinding' of investigator / interviewer to the study participant's outcome/ exposure status
 - in [case-control studies](#) those who are determining the exposure status of a study participant should be unaware of whether the participant is a case or a control
 - collecting information about exposure prior to definitive diagnosis / knowledge of outcome
 - e.g. in a *nested* [case-control study](#), information on exposures is likely to have been collected at baseline, before cases were diagnosed, rather than data on exposure and outcome being recorded at the same time, thus reducing [observer](#) (and [recall](#)) [bias](#) [2]
 - in [cohort studies](#), data on outcomes should be collected without knowledge of exposure status of a participant i.e. 'blinding' of interviewer to exposure status
- 'Blinding' of study participant (more difficult) by not revealing the exact research question in a study [2]
- 'Blinding' the interviewers to the study hypothesis
- Establishing explicit, objective criteria for exposures and outcomes [3]
- Using standard questionnaires, with good [questionnaire design](#); the questionnaire should be [valid](#) and reliable [2]
- Using a small number of interviewers to prevent too much variation between observers [2]

- Training interviewers to ask questions the same way

Preventing recall bias

Approaches taken to prevent [recall bias](#) include:

- improving timeliness of information gathering, so that the interval between the event/ illness of interest and the study (the recall period) is as short as possible, thus reducing [non-differential recall bias](#); data on exposures should be collected as near as possible to the time of exposure
- framing questions to aid accurate recall [\[1\]](#), so that inaccurate recall is limited among controls as well as among cases, thus reducing [differential recall bias](#)
- taking a different control group that will not be subject to the same incomplete recall i.e. using as controls individuals with a disease considered to have a similar impact on recall to the one being studied
 - e.g. *case-other disease* approach: to reduce [maternal recall bias](#) in a [case-control study](#), select as controls mothers of babies born with birth defects other than the one under study who may have recall of early pregnancy exposures similar to case mothers [\[1\]\[5\]](#); however, McCarthy suggests this approach should be treated cautiously, as knowledge by mothers of different hypotheses regarding causes for different birth defects would mean that recall could still be [differentially biased](#), and exposures relevant to the birth defect of the control group mean that these cases don't represent the real exposure experience of the population under study [\[5\]](#)
 - e.g. *case-case study design*: in analysing an outbreak of a salmonella strain, we could use exposure data from a recent outbreak of another salmonella strain as a 'control', instead of looking for controls in the present outbreak [\[5\]](#); the notified cases from a previous outbreak are more representative of the background population of the diagnosed salmonella cases in our outbreak, namely the subpopulation of people who would present to a doctor when they have gastroenteritis and have a specimen taken - they may have a different quality of recall than individuals who don't do this despite having similar symptoms [\[5\]](#).
- using information from medical records/ other independent sources recorded before the diagnosis/ disease outcome was known rather than information from questionnaires collected after the outcome [\[1\]](#), i.e. use objective records rather than relying on recall; see [\[6\]](#) for an example of a *nested case-control study* where symptoms were recorded at the time they were reported rather than being recalled retrospectively.

References

-
1. Rothman KJ. Epidemiology - An Introduction. New York: Oxford University Press; 2002
 2. Bailey L, Vardulaki K, Langham J, Chandramohan D. Introduction to Epidemiology. Black N, Raine R, editors. London: Open University Press in collaboration with LSHTM; 2006.
 3. Sackett DL. Bias in analytic research. J Chronic Dis. 1979; 32(1-2):51-63
 4. Vrijheid M, Richardson L, Armstrong BK, Auvinen A, Berg G, Carroll M, et al. Quantifying the impact of selection bias caused by nonparticipation in a case-control study of mobile phone use. Ann Epidemiol. 2009 Jan;19(1):33-41.
 5. McCarthy N, Giesecke J. Case-case comparisons to study causation of common infectious diseases. Int J Epidemiol. 1999 Aug;28(4):764-8.

6. Black C, Kaye J, Jick H. Relation of childhood gastrointestinal disorders to autism: nested case-control study using data from the UK General Practice Research Database. *BMJ*. 2002 Aug 24;325(7361):419-21.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Types of variables and line listing

Last modified at 8/12/2014 2:00 PM by Arnold Bosman

Types of variables

Methods for describing epidemiological measurements (by person, place and time) depend on the type of data or variables used [1, 2, 3]. A variable is a characteristic of the data under consideration. Types of variables can be classified in a number of ways. One common way to classify variables is by measurement scale which distinguishes four scale types: nominal, ordinal, interval and ratio scales [4]. Another common classification system contains two main classes: categorical (qualitative) and numerical (quantitative) variables. In general categorical variables cover the first two types of measurement scale, while numerical variables cover the second two scale types.

A **categorical variable** (also known as **qualitative variable**) is one for which each response can be put in a specific category. Categorical variables can be either nominal or ordinal.

A **nominal variable** is one that describes a name or category, e.g. occupation, place of birth, diagnosis. There is no inherent order in the set of possible names or categories. Nominal data is called **dichotomous** when it is characterised by only two classes e.g. sex (male/female), exposure history (yes/no).

An **ordinal variable** is a categorical variable for which the possible categories can be placed in a specific order or in some natural way that gives additional information, e.g. severity of illness may be categorised and ordered as "mild", "moderate" or "severe".

A **numerical variable** (also known as **quantitative variable**) is one that can assume a number of real values, units of measurements are used. Not all variables described by numbers are considered numerical. When the person is asked to assign a value from 1 to 5 to express the severity of his/her disease, numbers are used, but the variable itself (severity) is an ordinal variable.

A **discrete variable** can only take a finite number of real values, usually whole numbers. This variable often relates to counted items, e.g. number of new cases of salmonellosis in a given year, number of people in a household. Discrete

variables may also be grouped.

A **continuous variable** assumes an infinite number of real values, though necessarily recorded to a predetermined degree of precision. It often relate to measured items such as age, weight, temperature. To make them easier to handle, continuous variables are usually grouped into "class intervals" (e.g. age groups).

Line listing

When individual records are collected, they are typically entered and organised in a spreadsheet on a computer (or, if computer is not at one's disposal, on paper) where each row represents a case and each column represents a variable of interest (e.g. demographic information, clinical details, epidemiological information such as risk and exposure factors etc), creating a line listing. Such a list can be useful to view the entire database as time progresses, to fill in gaps of information, to share results with others on the team, an simply to "eyeball" for obvious errors, outliers, and trends [5]. It is a working document that also makes it easier to regroup and count cases by their characteristics, for example by using pivot tables [6].

New cases should be added to the list as they are identified, and all cases should be updated throughout the study or investigation as new information is obtained. Line listings that contain only the basic critical information have the advantage of providing a quick visual assessment of different aspects. However, a line listing with additional information may be more useful for assessing and characterizing the event of interest. All line listings should include the components of the case definition. In situations where more than one person enters data in the database, it is recommended to include initials of those who enter data in the database, should questions arise about the data entered [7].

A line listing enables the investigator to quickly summarize, visualize and analyze the key components of the data. See an example below.

Table: Partial line listing of a gastroenteritis outbreak

Initials (1)	Age (2)	Sex (3)	Date of onset (4*)	Presenting symptoms (5)				No. of diarrhoeal episodes per day (6)	Duration of illness in days (7)	Severity of illness (8)	Pathogen (9)
				Diarrhoea	Vomiting	Fever	Other				
N.L.	34	F	May 4	1	1	0	0	3	5	severe	Salmonella
G.D.	52	F	May 5	1	0	0	nausea	2	4	mild	pending
I.P.	26	M	May 5	1	0	1	0	2	3	moderate	pending
F.R.	40	F	May 8	1	1	0	nausea	2	1	mild	Norovirus
D.A.	37	F	May 5	1	0	1	abdominal cramps	3	6	severe	Salmonella
E.J.	61	M	May 9	1	0	1	headache	3	4	severe	Salmonella

Nominal variables: 1, 3, 5, 9; Ordinal variable: 4, 8; Discrete variable: 6; Continuous variable: 2, 7 (*Though time is continuous, date is ordinal.)

As line listings will contain individual patient data, including identifiers, disease outcomes and risk factors, these files

need to be considered as individual patient data and have to be treated with the same confidentiality and care as regular medical files. Proper data protection procedures need to be in place and monitored.

References

1. Merrill RM, Timmreck TC. An introduction to epidemiology. 4th ed. Sudbury, Massachusetts: Jones and Bartlett Publishers; 2005.
2. McLennan W. 1331.0 Statistics - a powerful edge! 2nd ed. Australian Bureau of Statistics; 1998. p. 103.
3. Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991. p. 10-13.
4. http://en.wikipedia.org/wiki/Level_of_measurement
5. Gregg MB. Appendix - A walk through exercise: a food-borne epidemic in Oswego county, New York. In Field Epidemiology. Ed Gregg MB. Oxford University Press, New York, 2002, p. 420
6. Fontaine RE, Goodman RA, Describing the findings. In Field Epidemiology. Ed Gregg MB. Oxford University Press, New York, 2002, p. 79
7. Torok M. Case finding and line listing: a guide for investigators. Focus on Field Epidemiology. Volume 1, Issue 4. The North Carolina Institute for Public Health



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Ascertainment Bias

Last modified at 9/15/2011 11:03 PM by Arnold Bosman

Case ascertainment (surveillance) bias

This happens when there is more intense surveillance/ screening for the outcome among exposed than among unexposed.

Example: let's assume that the objective of a [case-control study](#) is to assess if a history of past trauma is a risk factor for AS (ankylosing spondylitis). Cases of AS are compared with a random sample of the general population with regard to a history of past trauma. Having a history of trauma, which increases the likelihood of having X-ray investigations, will lead to a higher likelihood of diagnosis of AS in persons with this trauma history than in the general population. Therefore the proportion of AS cases with a history of past trauma is higher among cases, and the related [odds ratio](#) (OR) will be overestimated.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Referral bias

Last modified at 9/15/2011 11:10 PM by Arnold Bosman

Referral [bias](#) (admission rate [bias](#)) refers to a situation where the chance of exposed *cases* being admitted to the study is different to exposed *controls*. This happens frequently when cases are selected in a hospital whose activity is linked to the studied exposure. The [admission rate bias](#) may be due to a number of factors e.g. access to care, popularity of certain hospitals/ doctors etc [\[2\]](#).

In hospital-based studies, if the admission rates to hospital differ for different disease / exposure groups (e.g. admission rates of exposed and unexposed cases and controls differ), the [association](#) between exposure and disease will be distorted, and the relative [odds](#) of exposure to the putative cause may be spuriously increased or reduced. This [bias](#) is also known as *Berkson's bias* [\[2\]](#).

Example: in a study of risk factors for lung cancer, cases were compared to controls with regard to history of exposure to asbestos. Cases were recruited in the respiratory department of a hospital which is the National Reference centre for asbestosis. Controls were selected in the surgical wards of the same hospital. In that situation, it is likely that lung cancer cases of this respiratory department do not represent other cases with regard to history of asbestos exposure. Here, the selection of cases is linked to exposure. Selected cases are more likely to have been exposed to asbestos (than other lung cancer cases in the population), with an overestimation of 'a', resulting in an overestimation of the [odds ratio](#).

Exposure	Cases of lung cancer	Controls from surgical wards	OR
<i>Contact with asbestos</i>	a↑	b	OR↑
<i>No contact with asbestos</i>	c	d	reference
Total			

Diagnostic bias

Last modified at 9/15/2011 11:13 PM by Arnold Bosman

This happens when the diagnostic approach is related to knowledge of the subject's prior exposure to a putative cause (e.g. taking a certain drug, being exposed in an outbreak etc [\[2\]](#)).

Example: let's suppose that a [case-control study](#) is conducted to test if oral contraceptives (OC) are a risk factor for endometrial cancer. A group of cases and an equal number of controls are selected. Cases are selected at GP (family doctor) surgeries. Cases who use OC may be more likely to be offered screening for endometrial cancer either systematically or because of a side-effect of OC (breakthrough bleeding). The chance of undertaking detection of endometrial cancer is therefore higher among OC users than among other cases i.e. the use of OC may cause the search for endometrial cancer (by causing symptomless patients to bleed) rather than causing the cancer itself. The result is that a higher proportion of cases report using OC, with an overestimation of 'a', leading to an overestimation of the [odds ratio](#).

Exposure	Cases of endometrial cancer	Controls	OR
Uses OC	a↑	b	OR↑
Doesn't use OC	c	d	reference
Total			

Sackett [\[2\]](#) describes this example, where an innocent exposure may become a suspect, if, rather than causing a disease, it causes a sign or symptom which precipitates a search for a disease, as '*unmasking (detection signal) bias*'.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Non-response bias

Last modified at 9/15/2011 11:16 PM by Arnold Bosman

This is a [systematic error](#) due to the differences in response rates of participants in a study [\[1\]](#), and happens when participation in the study is related to the exposure status.

In a [case-control study](#) it is sometimes difficult to identify controls. Some don't respond either because they refuse, because they cannot be contacted, or because their exposure cannot be documented. The assumption is then that controls not included in the study (non-respondents) have the same history of exposure as controls who respond. However, if this is not true - and [non-respondents](#) exhibit exposures or outcomes which differ from those of respondents - the exposure among controls may be either overestimated or underestimated, leading to a lower or higher [odds ratio](#). Efforts must be made to achieve high [response rates](#) (i.e. a low 'non-response rate') and [prevent non-response bias](#).

The antithetical [bias](#) is called '[volunteer bias](#)' (i.e. volunteers from a specified sample may exhibit exposures or outcomes (e.g. be healthier) different to those of non-volunteers e.g. volunteers for screening [\[2\]](#)).

Example: the following example illustrates the consequences of [non-response](#) linked to exposure in a case-control study ([non-response](#) occurs among *controls*).

All cases and controls are responding			
Exposure	Cases	Controls	OR
Yes	150	50	9.0
No	50	150	reference
Total	200	200	

30% of controls are not responding (respectively 30% of exposed and 30% of unexposed controls)			
Exposure	Cases	Controls	OR
Yes	150	35	9.0
No	50	105	reference
Total	200	140	

i.e. if the *proportion of non-response* is equal among exposed and unexposed controls, the OR is unchanged.

This second example illustrates the effect on the estimation of the OR when the proportion of *non-response* differs among exposed and unexposed controls, although the overall *non-response* rate among controls is still 30%, as in the first example.

30% of controls are not responding (respectively 70% of exposed and 17% of unexposed controls)			
Exposure	Cases	Controls	OR
Yes	150	15	25
No	50	125	reference
Total	200	140	
30% of controls are not responding (respectively 10% of exposed and 37% of unexposed controls)			
Exposure	Cases	Controls	OR
Yes	150	45	6.3
No	50	95	reference
Total	200	140	

i.e. if the *proportion of non-response* is not equal among exposed and unexposed controls, the estimated OR is biased.

The same consequence can be observed if *non-response* occurs among cases.

Example: a *case-control study* to assess the association between smoking and myocardial infarction (MI) was done using a postal questionnaire. *Non-response* was higher among exposed than unexposed MI cases, leading to an underestimation of the strength of *association* between smoking and MI.

Survival bias

Last modified at 9/15/2011 11:18 PM by Arnold Bosman

This **bias** occurs when survivors of a highly lethal disease are more likely to enter a study than other cases.

Example: let's suppose we study the role of age as a potential risk factor for viral haemorrhagic fever (VHF), and that the study includes only those who are still alive at the time of the study. If older age is associated with VHF death, this will decrease the proportion of cases over a certain age in the study, and consequently underestimate the [odds ratio](#).

This is illustrated in the following table:

[illegible]



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Helminthic mechanisms of immune evasion

Last modified at 11/30/2013 12:43 PM by Arnold Bosman

Schistosomes are known to survive intravascularly for many years despite the continuing antiparasite immune response by the infected host. The living schistosomes are capable to reduce surface antigenicity and develop a tegument resistant to immune damage. In addition to this 'camouflage-behavior', these parasites influence cellular immune response using "chemical defenses":

Epidermal Langerhans cells (LCs) are a specific kind of [antigen presenting cells](#) and play a key role in immune defense mechanisms and in numerous immunological disorders. Studies have shown that percutaneous infection of mice with the helminth parasite *Schistosoma mansoni* leads to the activation of LCs but also to their retention in the epidermis. Moreover parasites transiently impair the departure of LCs from the epidermis and their subsequent accumulation as dendritic cells in the draining lymph nodes. The inhibitory effect is mediated by soluble lipophilic factors released by the parasites and not by host-derived antiinflammatory cytokines, such as interleukin-10.

References:

- Pearce, E. J., and A. Sher. 1987. Mechanisms of immune evasion in schistosomiasis. *Contrib. Microbiol. Immunol.* 8:219-232
- Angeli, V., C. Favéuw, O. Roye, J. Fontaine, E. Teissier, A. Capron, I. Wolowczuk M. Capron, and F. Trottein F. 2001. Role of the parasite-derived prostaglandin D2 in the inhibition of epidermal Langerhans cell migration during schistosomiasis infection. *J. Exp. Med.* 193:1135-1147.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Principles of matching

Last modified at 9/3/2014 5:04 PM by Arnold Bosman

Confounding implies that the confounding factor (which is one of the exposures) is not evenly distributed between cases and controls (or between exposed and unexposed). Therefore in order to prevent confounding the simplest solution would be to design a study in which cases and controls (or exposed and unexposed) would have an equal distribution of the confounding factor. This process is called matching.

Matching is most often applied in to case controls studies, however matching may be performed also in cohort studies [1].

We usually identify two types of matching process, **individual matching** and **frequency matching**. Both individual and frequency matching have the same consequence: matching will have to be taken into account during the analysis.

Individual matching

In this first method, matching is performed subject by subject. This is called individual matching. For example, if age is a confounding factor, for each case age 30 years, one control of the same age will be selected, and so far and so on for all cases included in the study. The results are *pairs* of individuals belonging to the same study population and sharing one common characteristic (in this example, a specific age).

In individual matching, we may also consider to select more than one control per case. Then two or more controls have then the same characteristic of the case. We have then constituted *triplets* (one case and 2 controls), *quadruplets* (one case and 3 controls), etc.

Frequency matching

In a second type of matching process, matching is no longer done individually but for groups of subjects. In such instance a group of controls is matched to a group of cases with respect to a particular characteristic (the confounding factor). For example if in a case control study with 50 cases there are 20 men and 30 women, we would select a

control group having the same gender distribution. We would first select 20 men from the male study population and then 30 women from the female study population.

Why matching?

Matching controls to cases is nothing more than stratifying in advance of analysis. Instead of constituting strata at the time of the study analysis we prepare them before the study is done, at the time of controls selection. When we select one control per case, each stratum will include one case and one control. We will therefore have as many strata as pairs in the study. The objective of matching is to prepare the analysis. Matching optimizes the number of cases and controls per stratum. It avoids having no case or no control in a stratum, as could happen when doing a stratified analysis afterwards (The biggest inefficiency in a stratified analysis done afterward would occur when in a stratum there is either no case or no control). This is why matching is frequently mentioned as a way to improve the efficiency of an analysis by better distributing cases and controls between strata.

References

1. Rothman KJ, Greenland S; Modern Epidemiology, second edition. Lippincott Williams & Wilkins 1998, 147-161.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Matching

Last modified at 9/19/2011 9:55 AM by Arnold Bosman

Studying the effect of an exposure (risk factor, behaviour, intervention etc) on a health outcome within a population is a key part of epidemiology. If life was truly simple, then measuring the distributions of the exposure and outcome of interest in a population and presenting these variables in a single two-by-two table would be enough to determine this effect (relative risk, odds ratio, vaccine effectiveness etc).

However, life is always more complex; there are 'third variables' that can distort (confound) our observation of the effect of interest. In some studies there may be many of these third variables, which we therefore call confounders.

In epidemiology there are different ways to address confounding.

- Matching is a way to *prevent* confounding during the stage of the study design.
- [Restriction](#) is another way to *prevent* confounding, which is also planned for during the stage of the study design.
- Performing a multivariate or a [stratified analysis](#) is a way to *control* confounding during the analysis, and not during the design of a study

Matching is most often used in a case control design, but it is also possible to use it with a cohort study design.

A [confounding factor](#) is a factor associated with the outcome (independently from exposure) and also associated with exposure (without being in the biological pathway between exposure and outcome). The confounding factor distorts the [measurement of the effect](#) (RR or OR) between the exposure and the outcome. Matching is the process that leads to have the same distribution of the confounding factor among cases and controls.

If the study was not planned with a matched design, an alternative solution to *control* confounding will be to perform a stratified analysis or to use multivariate models (for example a [logistic regression](#) model).

If matching was performed during the study design, it will need to be taken into account during the analysis. In this event, the formula used to calculate the OR will be different, and a special type of logistic regression should be used

(conditional logistic regression). Therefore the [table format and the analysis](#) to be used in a matched case control study are different than those be be used in an unmatched case control study.

During the study design, matching can be performed according to different [principles of matching](#), called frequency matching and individual matching,

Matching has [disadvantages](#). Therefore, the decision on whether to do a matched design must be carefully thought, especially nowadays where epidemiologists are not performing calculations by hand and multivariate models like logistic regression are available from many softwares. The greatest advantage is that by doing a matched design, we will be sure that no strata contains few or none observations, therefore increasing the efficiency of the analysis, with a reduced sample size and a higher amount of information per subject.

Matching is often used for convenience e.g. when it is difficult to obtain a random sample of the source population as controls. However there is no need to match since there are many limitations and traps when using a matching strategy. If resources are available a larger sample and an "a posteriori" stratified analysis may be easier to design and conduct, especially if we are confident that we can collect data on the main confounding variables. If we decide to match, we should make sure that the matching factor is a confounder, that we do not need to further study that factor, and that identification of matched controls will be logistically feasible and easier than an unmatched selection of more controls.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Special Considerations in Control Selection

Last modified at 9/14/2011 2:03 PM by CeRC

Controls in different types of case control studies: [case cohort](#), [traditional case control](#), [density case control](#).

Lets come back to the one of the characteristics of the control population, that they should be representative of exposures in the source population. In selecting controls for a case cohort study, a random sample of the source population should, if done correctly, be representative of the exposure distribution in the population that gives rise to the cases. In a traditional case control study, where cases are excluded from the control selection, a bias has been introduced as the exposure distribution in potential controls is no longer representative of the source population. If the attack rate is low, this bias will also be low, but if attack rate is high, the potential for bias will also be high. In a density case control study where cases occur over a long time period, controls should be selected from the source population still free of disease at the time the case occurs. In this way they should be representative of the person time experience of the source population [\[1\]](#) .

Asymptomatic cases. Does failure to identify those with mild or asymptomatic infection as cases introduce bias? This situation is analogous to non- response among cases. If the exposures among symptomatic and asymptomatic cases are the same, then no bias is introduced. There is only a reduction in statistical power. There is no difference in control selection if the controls are representative of the source population.

Example: In a hypothetical case control study with 40 cases and 40 controls, and 50% exposure among cases, Odds Ratio = $600 / 200 = 3.0$

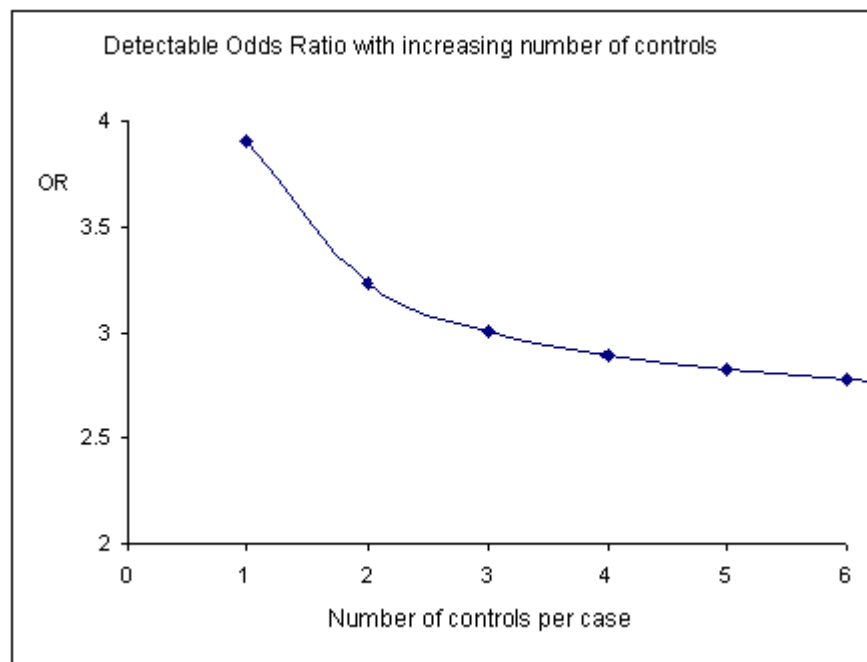
	Cases	Controls
Exposed	20	10
Not exposed	20	30

If we only detect 20 cases with the same number of controls , the Odds Ratio is unchanged ($300/100 = 3.0$) as long as % exposure is the same in detected and undetected cases.

	Cases	Controls
Exposed	10	10
Not exposed	10	30

Immune subjects. If some of the population are immune at the start of the study, then they are not eligible to be cases. They should then also be excluded as controls as they are not part of the source population. In practice we do not usually know who is immune. Again this may not matter if % exposed is the same in immune and non-immune cases. However it may be that subjects are immune because they have already been cases in the past and that they have a similar level of exposure to the risk factor that caused the cases in the outbreak under study . This introduces bias that reduces the OR towards 1 and may result in a failure to detect a true association, especially if the proportion immune is high. For example, the inclusion of immune subjects in the control group is thought to explain the results of some case control studies that fail to show an association between contaminated drinking water and cryptosporidiosis [2] .

Power and sample size in case control studies. A question often arises about the number of controls given a limited number of cases. Statistical programmes like Epi-Info can be used to estimate the sample size required to detect a specified odds ratio. It is unusual to select more than 3 or 4 controls per case as little statistical advantage is gained beyond this number (Figure) [3] . Alternatively we could show that power increases and plateaus with an increasing number of controls per case. The graph would then have the same shape but inverted.



References

1. Rothmann KJ. Epidemiology: an introduction. Oxford University Press 2002.
2. Hunter P. Modelling the impact of prior immunity, case misclassification and bias on case-control studies in the investigation of outbreaks of cryptosporidiosis. *Epidemiol Infect* 2000;125:713-8.
3. Kirkwood, B. R., Sterne, J.A.C. Essential Medical Statistics (2nd Ed). Blackwell Science 2003.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Methods for setting thresholds in time series analysis

Last modified at 10/28/2010 7:42 PM by Lisa Lazareck

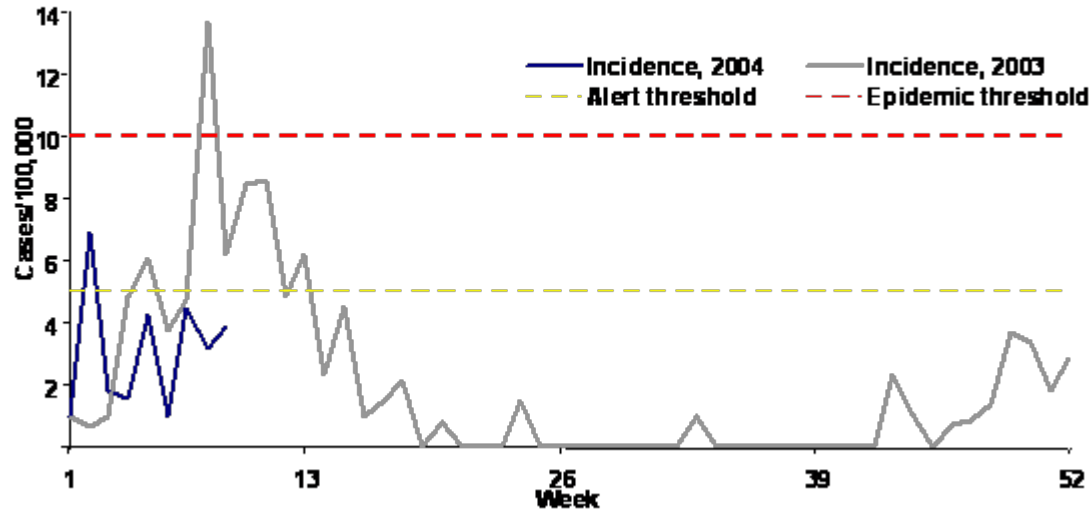
Temporal analysis consists in identifying abnormal events in the temporal distribution of a disease. This can be straightforward for rare diseases requiring immediate notification, but often requires the use of statistical methods to differentiate abnormal events from the expected fluctuation in notifications for diseases occurring at a baseline level in the community.

Such statistical methods include, from the simplest to the most sophisticated:

- **Crossing a predefined threshold**, as for meningitis in the meningitis African belt (> 5 cases/100,000/week, Figure 1)

Figure 1: Using a pre-defined threshold to detect unexpected changes

**Incidence of meningitis by week,
Boussé district, Burkina-Faso, 2003 - 2004**

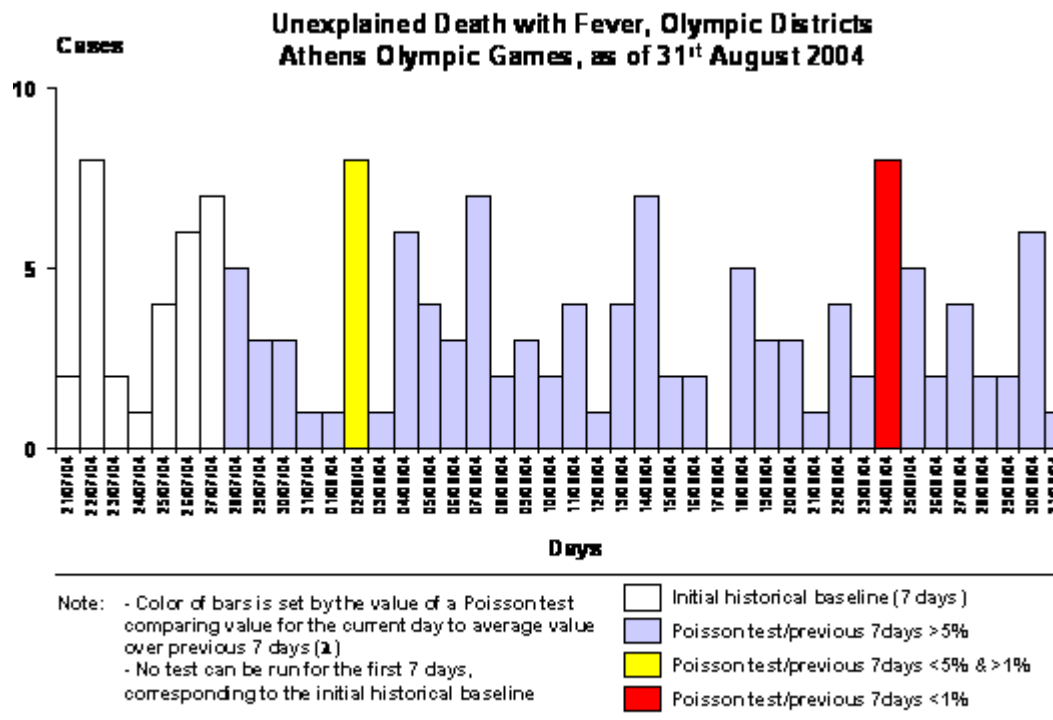


Source: Multi disease surveillance Centre, WHO, Ouagadougou, Burkina-Faso

- **Doubling or tripling of the absolute number of cases over a given time period.**

This method, although simple, is crude because it does not take into consideration the number of cases on which the increase is calculated. Doubling the number of cases from 10 to 20 carries a greater statistical significance than doubling the number of cases from 3 to 6. As an alternative, to enhance the method, a Poisson test can be applied to express the departure from previous period values (figure 2). For example, the averaged value observed in the previous 5 weeks is used as the expected value parameter in the Poisson test. The test returns the probability of observing the number of cases or more, for the week to test, assuming the averaged value is expected. For example if three cases, on an average, have been notified weekly in the past five weeks, the probability of notifying six cases or more is 0.084 (8.4%) using a Poisson test. This means that such an observation may occur by chance once every 12 weeks. Applying same approach to 20 cases when only 10 are expected yields a probability of 0.0035 (0.35%), potentially occurring by chance only once every 6 years.

Figure 2: Using a Poisson test to detect unexpected changes

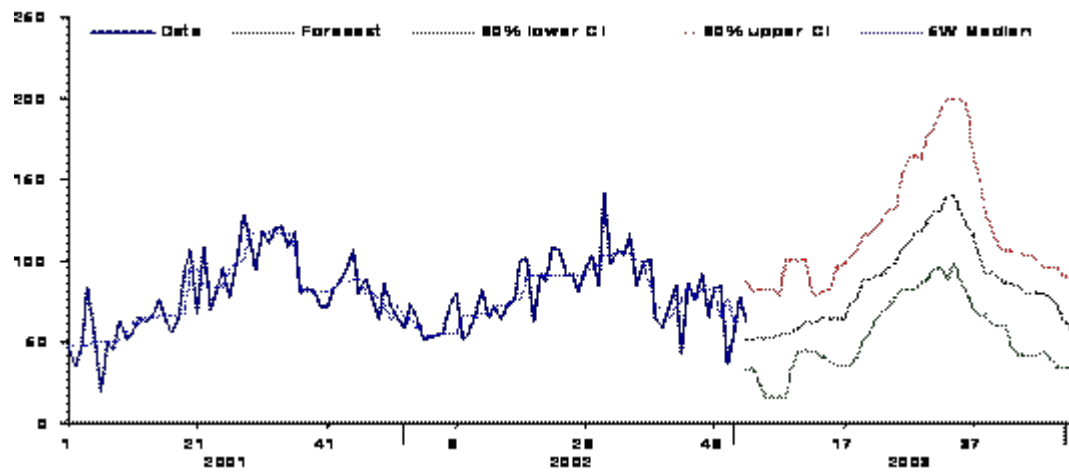


Source: *syndromic surveillance, Hellenic centre for communicable disease control*

- **Comparing number, proportional morbidity or rates with similar periods in the past.**

The mean or median of the number of cases observed on weeks from previous years, centered around the current week, are used as a basis for comparison (for example five weeks for the past five years resulting in an historical distribution of 25 weeks). The departure from the historical mean is measured by comparing it to the standard deviation of the historical distribution. When the median is used, thresholds for unusual pattern are defined by the value of the 90% or 95% percentile of the historical distribution (figure 3). This approach takes into account the seasonal variations by restricting the analysis to similar periods from past years but should not be applied to series presenting with a significant trend.

Figure 3: Using historical median and percentiles to detect unexpected changes



- Modeling historical data by time series analysis techniques that will account for trends and seasonality in the data.

Regression techniques are used to account for trend and seasons by fitting lines and sine curves to the data (figure 4). The confidence interval for detecting unusual events relies on calculating the dispersion of the residuals and applying a statistical threshold (95% confidence interval). More advanced techniques using seasonal autoregressive and moving average models (SARIMA) on differentiated data are indicated for diseases with unstable historical patterns but are rarely used in routine surveillance (figure 5).

Figure 4: Using periodic regression modeling to detect unexpected changes

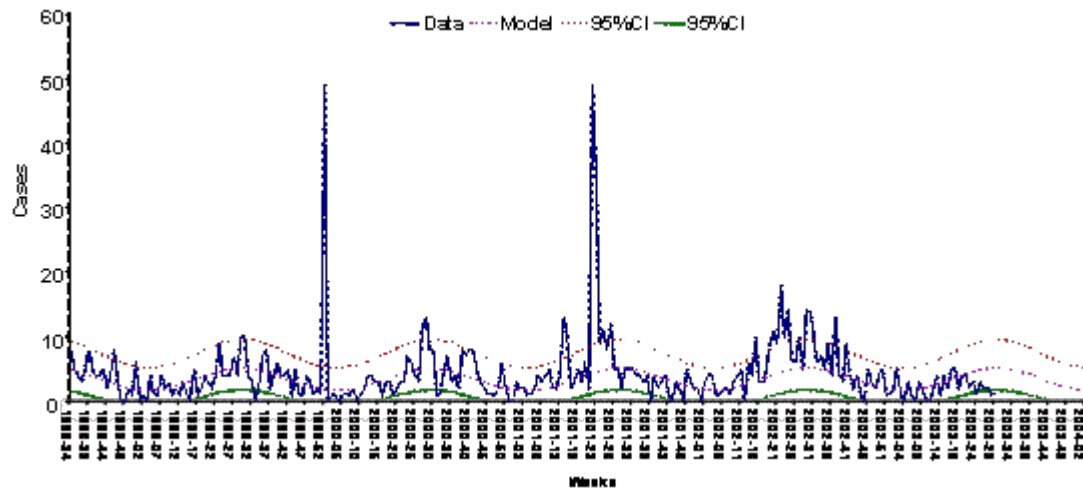
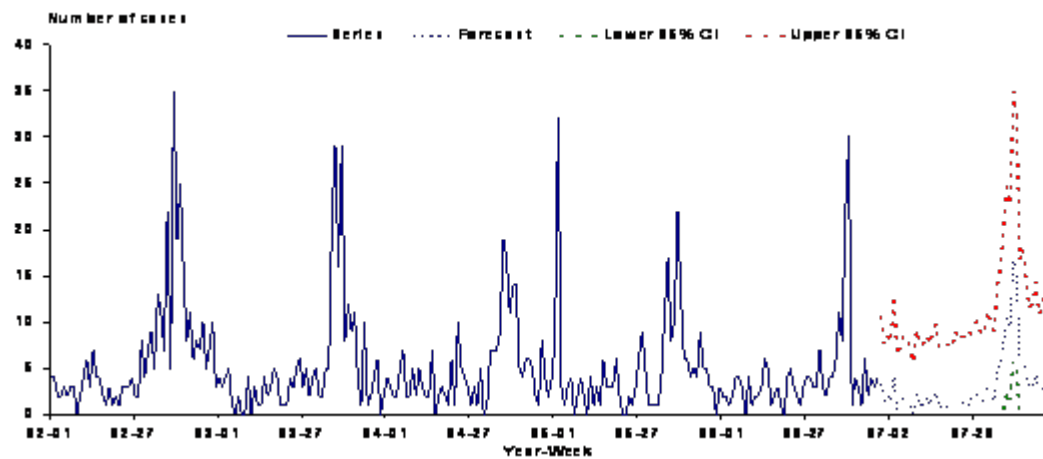


Figure 5: Using SARIMA modeling to detect unexpected changes



- **Weekly notification of typhoid and paratyphoid fever in France, 1992-1996**

The selection of the most appropriate method for analysis of time characteristic must take into account:

- The nature of the alert to detect, in relation with the type of transmission
- The availability of historical data. In a new system, historical comparisons can only be performed on the most recent periods, and may not account for seasonality.
- The existing statistical resources and software

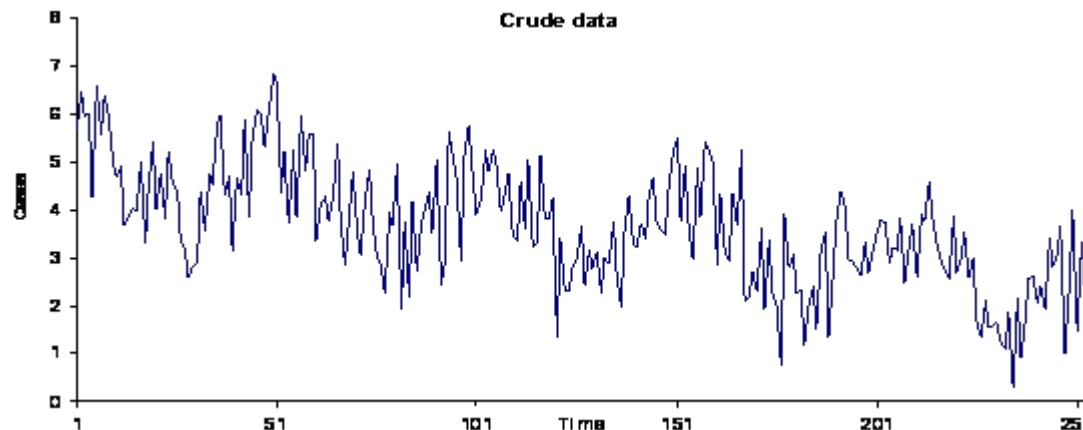
Smoothing techniques for describing time series

Last modified at 11/9/2010 11:51 PM by CeRC

Plotting crude data

Visual exploration of a time series uses a graphical presentation of the data, plotted along a time axis. The graph must be as simple as possible to allow a good visual inspection of the data.

Figure 1: Crude data signal

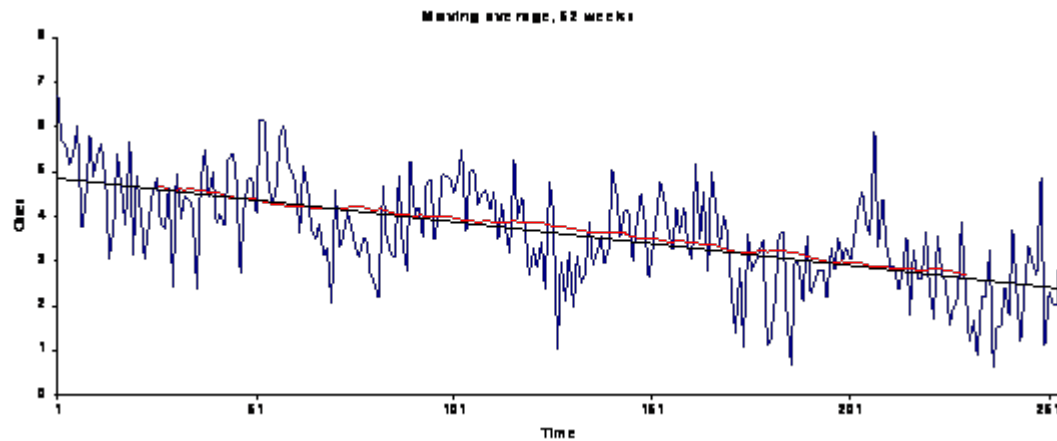


The aim of the descriptive analysis of a time series is to characterize it in term of trend and seasons.

Describing a trend

Describing a trend requires that a moving average window encompassing a full year be used (i.e. 52 weeks for weekly data). The shape of the resulting trend line is visually compared to the linear regression line.

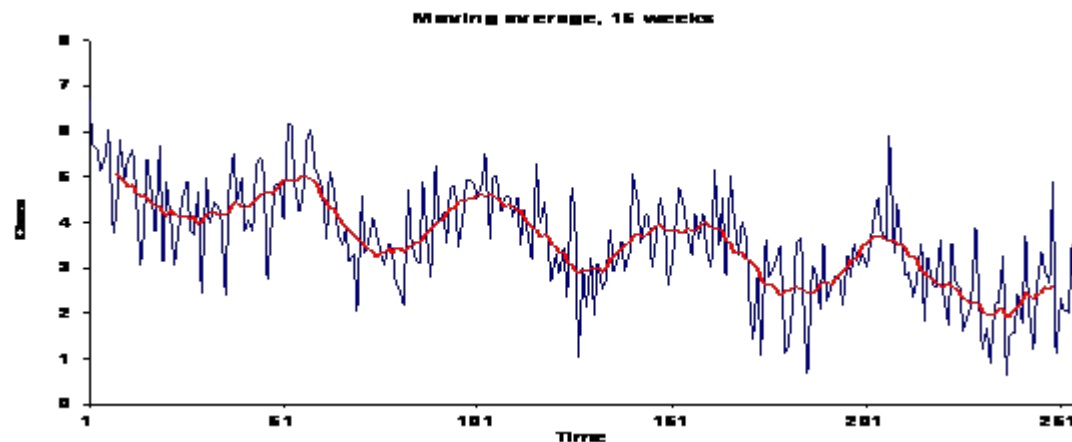
Figure 2: Crude data, 52 weeks moving average, and regression line



Describing seasonal variations

A moving average window of 15 weeks smooths the crude series, highlighting the seasonal pattern of the series. The size of the moving average windows depends upon the variance of the series, corresponding to the amount of variability to be smoothed. In general, windows from 5 to 15 weeks result in appropriate smoothing, but visual inspection of the result is required.

Figure 3: Crude data, and 15 weeks moving average



Principles for calculating moving averages

Spreadsheet software such as Microsoft Excel © allows to easily display time series filtered by a moving average window. For description purposes, it is important to display the averaged value in the center of the moving average window. The example below used a five-week moving average window. The first value equals the sum of the first 5 weeks, divided by 5 (5.73). This value is plotted against week 3, which represents the middle week for this first window.

Table 1: Crude series, formulas and smoothed series for calculating moving averages in Excel (5 week window)

	A	B	C	D
1	Time	Crude data	Formulas	Smoothed series
2	1	6.70		
3	2	5.73		
4	3	5.59	= AVERAGE(B2:B6)	5.73
5	4	5.15	= AVERAGE(B3:B7)	5.59
6	5	5.47	= AVERAGE(B4:B8)	5.20
7	6	6.03	= AVERAGE(B5:B9)	4.93
8	7	3.75	= AVERAGE(B6:B10)	5.07

Using the median value over the window rather than the mean generates a moving median which is more robust to outliers than the mean. In such case, the Excel © formula should be =MEDIAN(B2:B6) in cell C4 and copied over the range. Note that functions in Excel © may vary according to languages.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Which indicator to map

Last modified at 10/29/2010 10:39 PM by Vladimir Prikazsky

Count of cases

Counts are used to display the **burden** of the disease in the population. This helps policy makers and control programme managers to target programmes and allocate resources to areas most affected. However, expressing indicators as count of cases does not allow identifying areas with increased risk of transmission as the population varies across geographical areas.

Crude rates

Crude [rates](#) are a summary measure of the incidence of a disease in a population. They are calculated by dividing the number of cases (or deaths) of a disease having occurred in a certain period (often one year) by the average population in the area during the period. Rates are expressed per 1000 or 100,000 inhabitants, according to the frequency of the disease. Rates allow comparison between geographical areas by accounting for varying population size.

In outbreak investigations, rates are usually expressed for the epidemic period and referred to as attack rates.

Age and/or sex specific rates

Crude rates may be confounded by age and/or sex if the distribution of the disease is known to be associated with age and/or sex and if the population structure by age and/or sex varies across geographical areas. In some countries for instance, tuberculosis is known to occur at an increased rate among elderly people and elderly people are more represented in rural areas than in urban areas. Summarizing the incidence of tuberculosis using a crude rate will tend to over-represent rural areas with large elderly population while the risk of being infected, at a specific age, is not necessarily higher.

Mapping age and/or sex-specific rates controls for these potential confounders. However, maps cannot easily represent several age and/or sex specific rates in a single display and need to be repeated to reflect all age and/or sex

groups.

Standardized rates

Visual inspection of age and/or sex specific rates across geographical areas is a pre-requisite to mapping data. Whenever there are large variations of rates between age and/or sex categories, summarizing the incidence through standardized rates may not be indicated. However, there are instances where such a summary incidence is useful to assess risks of transmission across geographical areas, after controlling for age and/or sex potential confounders. This is achieved by a method called standardization of rates.

The use of crude rates when age-specific incidence and population structure differs, as in [Table 1](#), can result in the overall crude rate in district B being greater than that in district A (5.0 vs. 4.8) while age-specific rates in district B are both smaller than in district A (6.9 vs. 7.0 and 2.5 vs. 3.1). This paradox, called Simpson paradox, results from the confusion induced by age.

Table 1: Distribution of cases, population and rates of a disease by age group, in 2 hypothetical districts

District A	Cases	Population	Rate*		District B	Cases	Population	Rate*
0-39 years	42	600,000	7.0		0-39 years	55	800,000	6.9
40 years & +	25	800,000	3.1		40 years & +	15	600,000	2.5
Total	67	1 400,000	4.8		Total	70	1 400,000	5.0
* cases/100,000								

In these instances, standardization of rates is the technique required to control this confounder if a single summary incidence value is desired.

Direct standardization

Direct standardization consists of weighing age-specific rates by applying them to a reference population. Age-specific rates from district A and B are applied to a reference population for calculating age and/or sex-standardized rates. Controlling for age confounder by direct standardization as presented in [Table 2](#) shows that district B has an age-standardized rate smaller than district A, as expected when inspecting age-specific rates for both districts. The reference population can be an external population used at country level, such as the country population, for standardizing several indicators, or some international reference populations to allow for international comparisons. It can be the average population in the 2 districts, as in our example, if the objective is simply to compare the 2 areas.

Table 2: Calculation of age-standardized rates in 2 hypothetical districts by direct standardization

--	--	--	--	--

		District A		District B	
Age group	Reference population	Observed rate	Expected cases	Observed rate	Expected cases
0-39 years	1400000	7,0	98	6,9	96
40 years & +	1400000	3,1	44	2,5	35
Total	2800000		142		131
Age-standardized rate		5,1		4,7	

Indirect standardization

When the age distribution of the cases is not available in district A and B, or if age-specific rates are unstable in relation with small figures, indirect standardization is indicated. It consists of applying reference age-specific rates to the populations of study. This yields the expected number of cases in each district, if incidence had been in accordance with the reference model. The age-standardized incidence ratio is calculated by dividing the number of observed deaths over the number of expected. It is sometimes multiplied by 100 and expressed as a percentage. [Table 3](#) shows, in our theoretical example, that the incidence in district A is 1.02 times the reference incidence and 0.95 times in district B, which shows that the incidence is lower after standardizing on age.

Table 3: Calculation of age-standardized rate ratios in 2 hypothetical districts by indirect standardization

			District A			District B	
Age group	Reference rates		Population	Expected cases		Population	Expected cases
0-39 years	7,0		600,000	42,0		800,000	56,0
40 years & +	3,0		800,000	24,0		600,000	18,0
Total			1 400,000	66,0		1 400,000	74,0
Observed cases				67			70
Age-standardized rate ratio (SRR)				1,02			0,95

--	--	--	--	--

Strategy for standardization

When considering whether standardization is indicated, the first step is to consider whether the mapping of the data can be confounded by variables such as age and/or sex. If the disease is not associated with age or sex, standardization on these variables is not required. Similarly, if the age and/or sex structure of the population is identical across geographical areas, standardization on age and/or sex is not required. In other instances, standardization is required if a summary value of the incidence of the disease is desired, in order to control for the induced confounding effect.

When mapping the data can potentially be confounded by age and/or sex, using age and/or sex-specific rates allow accurate comparisons of the geographical distribution of the disease. However, whenever summary incidence information is preferred, age and/or sex standardized rates are indicated.

Direct standardization allows better comparability across geographical areas but may be unreliable if age-specific rates are based on small numbers. In addition, age-standardized rates represent hypothetical values which have no base in reality. Indirect standardization requires less detailed information on cases. It is expressed as percents of a reference situation, which is easily understood. However, indirect standardization of rates is less robust for comparing different geographical areas when the population structure is very heterogeneous.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Choosing an appropriate type of map

Last modified at 10/28/2010 7:38 PM by Lisa Lazareck

The choice of the appropriate type of map relies on the aim of the representation and the nature of the indicator to map.

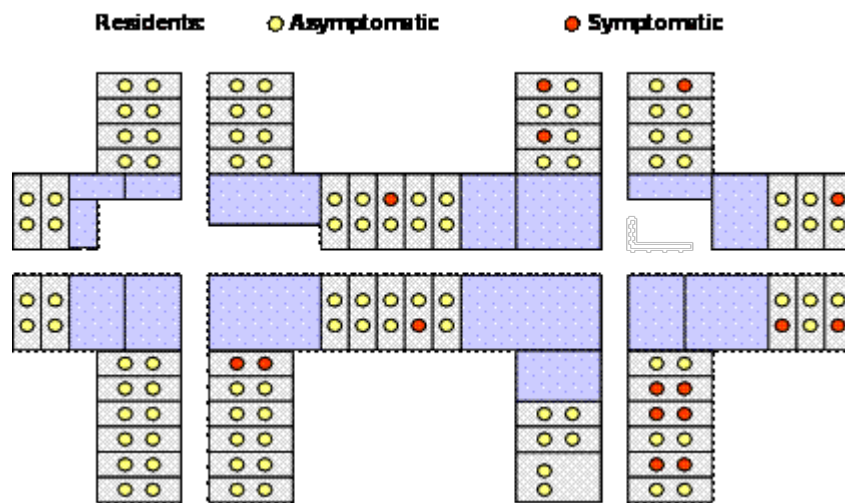
Dot-density maps

In a dot-density map, each case which occurred in an area is represented as a dot on the map. Dot-density maps are best at representing the **burden of disease** by geographical areas, expressed as count of cases. For rare diseases, dot maps are effective at detecting clustered cases. Dot-density maps are not indicated for representing rates or other composite indicators. They do not require classifying values in ranges and therefore, no information is lost in mapping.

Very few surveillance systems accurately record the exact address of residence of cases, allowing for a precise positioning of case dots. Usually, the smallest geographical area available in the surveillance data is used for mapping, and dots are randomly located within these administrative boundaries. Using randomly located dots on large geographical areas can be particularly misleading, as cases seem to have occurred homogeneously throughout the area, even in locations known to be sparsely populated. An indication should be placed on the map to indicate the random location of the dots within area boundaries. The size of the dots should be such that dots do not overlap because of their number. When the disease is frequent, a dot can be assigned to a number of cases, e.g. 1 dot = 10 cases. An example of a dot-density map is presented on figure 1 in chapter [Analysis by place characteristics](#).

During outbreak investigations, dot maps can be used to represent cases of the disease which occurred during the outbreak, plotted by place of exposure. If the outbreak occurs in a closed environment, such as a cruise-ship, a prison or a nursing home, cases can be plotted on a floor plan of the area (figure 1).

Figure 1: Distribution of resident by clinical status, Nursing home X, Delaware, USA, 1992.



Choropleth maps

In choropleth maps, geographical areas, usually defined by administrative boundaries, are filled with color or pattern to reflect the magnitude of the indicator to be mapped. Choropleth maps are popular because of their visual and intuitive appeal. However, careful thoughts should be given to designing the map as the level of aggregation, the mode of classification and number of categories, as well as the choice of colors may hide meaningful spatial patterns or create artifactual ones.

Choice of the level of geographical aggregation

Whenever crude rates, age and/or sex specific rates or standardized rates are plotted, numerator and denominator data needs to be aggregated by geographical areas. Administrative boundaries are generally used since denominators are usually available with sufficient details. However, the choice of the administrative level should avoid resulting in too few cases reported in each area that would result in wide variations of rates. It is not advisable to calculate rates when less than 20 cases are contributing to the numerator. Therefore, a balance needs to be struck between the size of the geographical area and the period during which cases are accumulated, in order to avoid such small figures.

Classification

Affecting a color or a pattern to a geographical area requires that data are organized in categories. Four to eight categories are used in general. Too few categories results in loss of information while too many may result in difficulties in perception. In addition, there should not be too many categories when the map includes relatively few geographical units.

Most mapping software offer several approaches in classifying data. Selecting the appropriate classification method requires inspecting the distribution of values across geographical areas ([figure 2](#)). The most commonly used classification methods include:

Equal count classification

This type of classification attempts to assign an equal number of observations in each category. Quartiles of the

distribution are used for four categories, quintiles for five. While this method is popular and commonly used, it may classify areas with close values in different groups and areas with dissimilar values in same categories based solely on ranks. It is best used for evenly- or normally- distributed data.

Equal interval classification

This classification applies equal amplitude ranges to data. Therefore, it is indicated when values of the geographical areas are evenly distributed, resulting in a straight line on the distribution plot ([figure 2](#)) and an equal number of areas in each category. It should not be used when the distribution is skewed or includes outliers since it may result in categories not being represented on the map (as in [figure 3](#), map 2).

Equal area classification

Equal area classification attempts to classify data so that each class is represented by an equal area on the map. Caution should be used with this method when population density varies greatly across areas: large under populated areas may appear in two categories in relation to their large size although their values may be very similar.

Mean and standard deviation classification

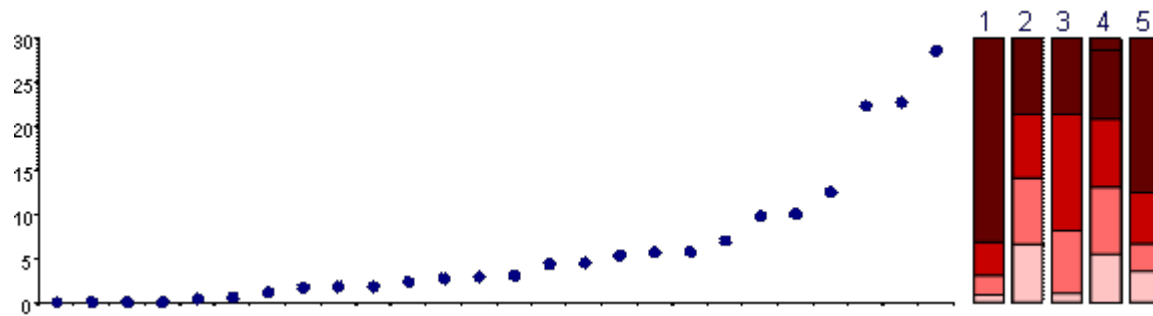
This classification method assigns categories on both sides of the mean using standard deviations (e.g. -1, -2, -3 standard deviations under the mean, +1, +2, +3 standard deviations above the mean). This is a statistical classification method best indicated if the distribution of values follows a normal distribution. It shows how area values differ from the average value. It is not indicated when the distribution of values is skewed and may result in categories not being represented on the map (as in [figure 3](#), map 4).

Natural breaks classification

This method attempts to regroup values by minimizing the variance within categories and maximizing the variance between categories. It tends to regroup similar values together to best represent the distribution.

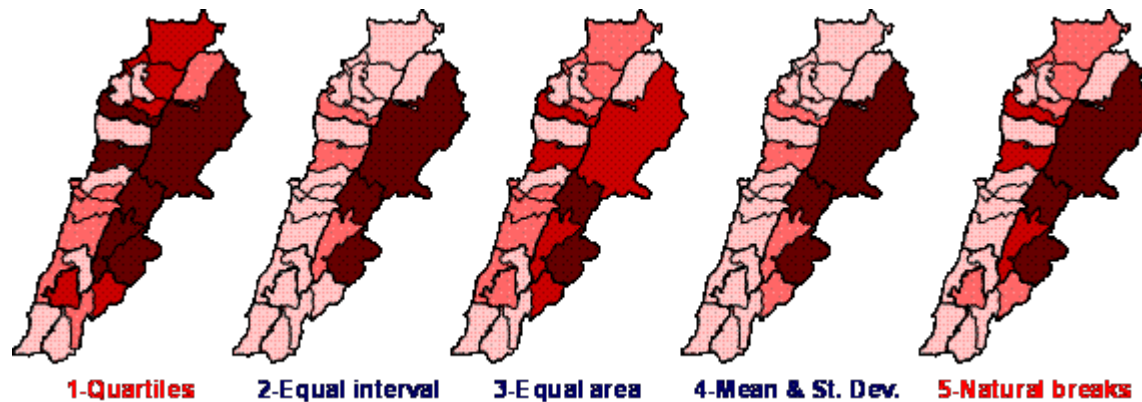
In conclusion, whenever values are evenly distributed, equal count or equal interval classifications can be used. When values are normally distributed, mean and standard deviation classification should be used. In other instances, natural breaks should be preferred, as this method makes no assumptions about the shape of the distribution. [Figure 3](#) shows example of the effect of the classification method selected.

Figure 2: Distribution of values and ranges of the classification methods, brucellosis notification in Lebanon, by district, 52 weeks rate, as of week 15 of 2003



1: quartiles, 2: equal interval, 3: equal area, 4: mean and standard deviation, 5: natural breaks

Figure 3: Distribution of brucellosis notification in Lebanon, by district, 52 weeks rate, as of week 15 of 2003

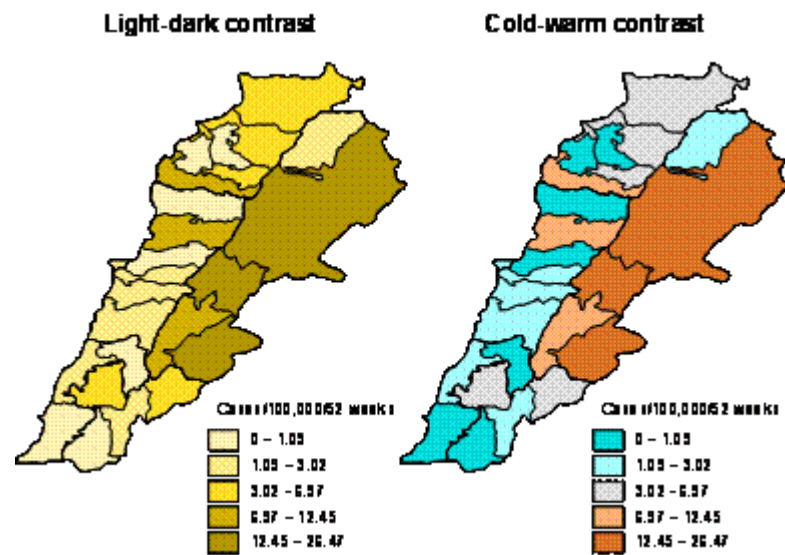


Grey shades and colors

Choropleth maps use analogy to represent the data. Caution should be used in assigning shades and colors to classification areas. Risks of transmission, expressed as crude, specific or standardized rates are quantitative in nature. Therefore, a light-dark color contrast should be preferred to represent it. A gradient of grey (black and white display) or of color (color display) such as yellow or red is indicated.

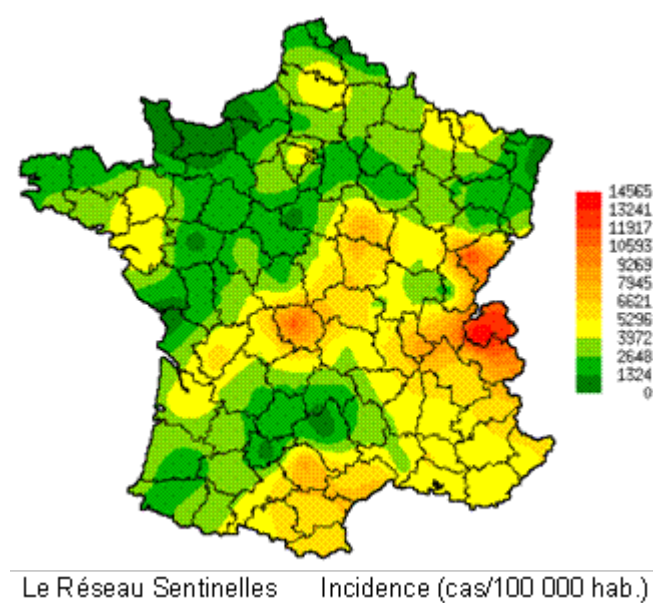
The cold-warm contrast ranging from cyan to orange can be used effectively to represent values below and over a mean as exemplified in [figure 4](#). While the cold-warm contrast is very effective in representing the opposition between high and low values, it does not appropriately represent the range of values when printed or duplicated in black and white. In addition, it may pose problems to person with color-blindness.

Figure 4: Distribution of brucellosis in Lebanon, by district, 52 weeks rate, as of week 15 of 2003



Isopleth maps

Figure 5: Incidence of flu-like illness in France, 2002



Source: INSERM Unit U 444, Epidemiology and information sciences

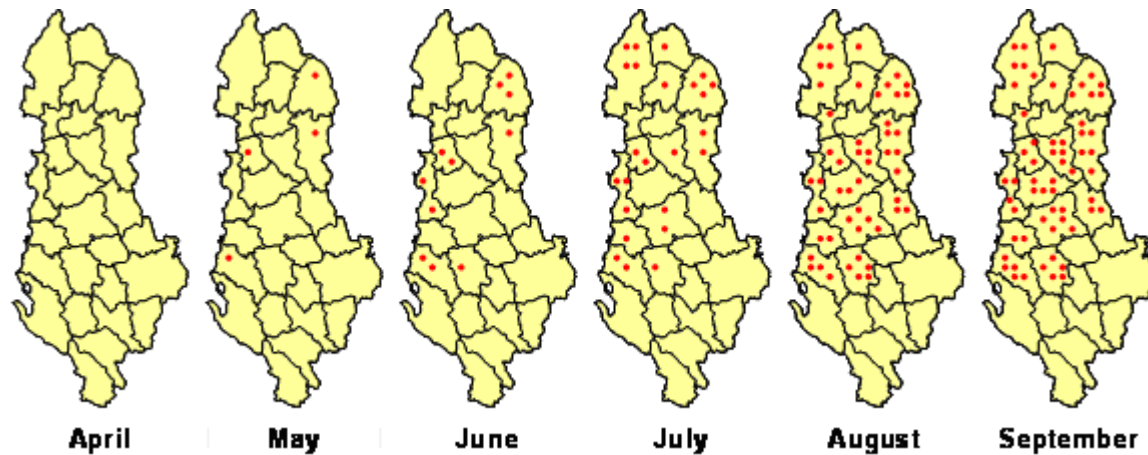
Isopleth maps ([figure 5](#)) do not require aggregating values by geographical areas but instead use the data itself to define geographical areas with similar values. It is indicated to represent continuous data that varies progressively over geographical locations, e.g. temperature, elevation. In surveillance, isopleth maps are indicated to represent incidence

when surveillance is based on a sample of sentinel sites. Shaded or colored areas boundaries are derived from the data using statistical methods such as kriging [1].

Mapping place and time

Displaying place and time characteristics of the distribution of a disease is a very effective way to grasp the dynamic of the disease transmission. This is best achieved by plotting subsequent maps next to each other as exemplified on figure 6.

Figure 6: Distribution of polio cases by district, Albania, April to September 1996



Additional information can be found in the "Handbook on geographic information system and digital mapping" from [UNSTAT](#).

References

1. CARRAT F, VALLERON A-J Epidemiologic Mapping Using the "Kriging" Method. Application to an Influenza-Like Illness epidemic in France. Am. J. Epidemiol., 1992; 135:1293-1300.
2. http://unstats.un.org/unsd/publication/SeriesF/SeriesF_79E.pdf
3. See <http://www.who.int/docstore/wer/pdf/2000/wer7538.pdf> for detailed information about meningitis thresholds in Africa.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Use of computers

Last modified at 9/14/2011 2:33 PM by CeRC

Graphs are now almost exclusively computer generated and epidemiologists become more and more dependent on the possibilities and limits of computer graphic software packages.

Whatever the software used some principles should probably be respected. We should avoid three dimensional graphs. They do not improve communication and they are not easier to interpret, rather the opposite. Colours should be selected according to complementary colour criteria. Many software programmes do not allow the user to construct histograms. What is frequently called a histogram is in fact a bar graph. Units and scales on the x and y axes are not always clear in some of the software packages. Most software does not allow users to create a standard epidemic curve with one square representing one case. Particularly when the x-axis illustrates time, most of the software programmes are not flexible enough to comply with what epidemiologists need.

Personal productivity tools are commercial products designed to handle standard computing tasks such as word processing, numerical analysis, data manipulation and storage, and data presentation. Typical products include [\[1\]](#):

- *Word processing software* is designed to create documents such as letters, reports, and manuscripts. MS Word, WordPerfect, and Word Pro are some examples of word processing products.
- *Spreadsheets*, used to organize information, are very useful for handling tabular data. Addition, subtraction, division, multiplication and totaling can be done very quickly, and all results can be automatically recalculated later if new data are inserted. Formatting and graphing facilities are used to aid analysis and presentation. Well-known spreadsheet packages are MS Excel and Lotus 123.
- *Database packages* make it easy to organize and store data in a uniform fashion. Data can be quickly and systematically searched, sorted and presented. MS Access and Lotus Approach are database packages.
- *Presentation packages* are used to illustrate discussions and lectures. MS PowerPoint and Lotus Freelance are examples of presentation packages.

References

1. Statistics Canada, Power from data! - The computer industry

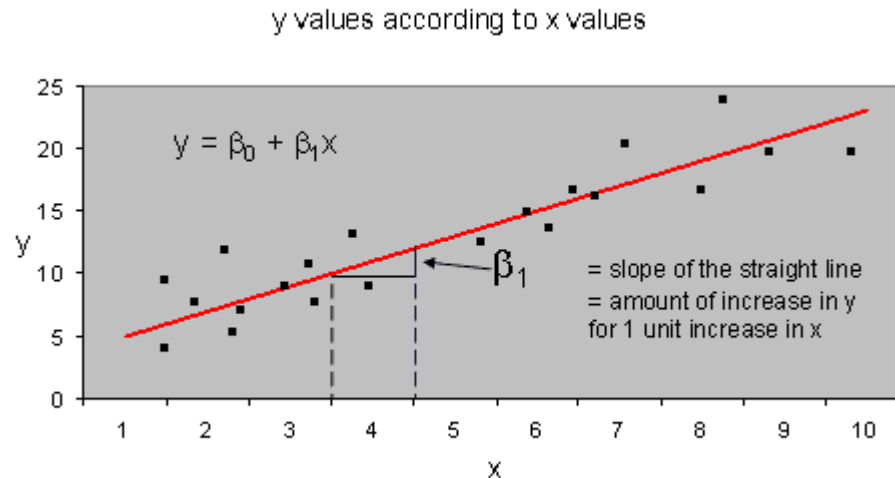
Linear models

Last modified at 9/19/2011 9:50 AM by Arnold Bosman

The straight line

Mathematical models can be as simple as a straight line. In a linear model the straight line is used to describe the relation between two variables. We express y according to the value of x . We predict y according to x . Therefore x is called the predictor or **independent variable** and y the predicted or **dependent variable**.

Figure 1 shows the relation between y and x .



The straight line represents the average values of y for different values of x . It is a regression line. It was obtained by fitting a straight line equation to the data. A simple way to understand how the straight line is fitted on the dot plot is to visually guess where it would need to be placed in order to minimise the various distances between each dot and

the line.

The equation of a straight line is:

$$y = \beta_0 + \beta_1 x$$

In which

β_0 is the intercept (value of y when $x = 0$)

$\beta_1 x$ is the coefficient of x. It describes the slope of the line. It represents the number of units of change in y when x increases by 1 unit.

The general linear model

Let's suppose that in the above example we want to predict y not only according to x_1 but also according to x_2 . We would have then two predictors. The relation between x_1 , x_2 and y is still a straight line. The equation is now:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

To locate the straight line in the dots we need to imagine a 3 dimensional rectangle coordinates with y expressed according to x_1 and x_2 .

The coefficients β_1 and β_2 respectively provide estimate of the effect of x_1 and x_2 which are mutually un-confounded. Mathematically there are no limits to the number of variables to be included in a model.

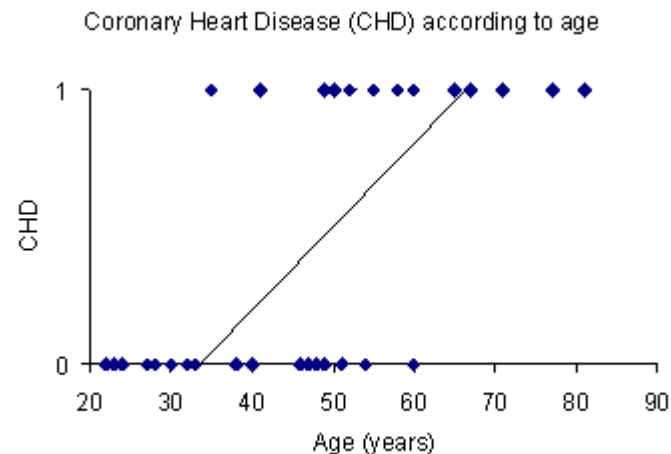
[<<Back to Logistic regression](#)

The logistic model

Last modified at 9/19/2011 9:50 AM by Arnold Bosman

In the linear model y can take all possible values from $-\infty$ to $+\infty$. However in epidemiology we are mainly interested in binary outcomes (ill or not, dead or not, etc.). They are frequently noted as 0 and 1.

Figure 1 shows the hypothetical distribution of cases of coronary heart disease (CHD) according to age.



From the above graph it seems that CHD cases may be older than others. A regression line would not really reflect the relation. In addition y , being a straight line, could vary between $-\infty$ and $+\infty$ which is not what we expect for disease occurrence.

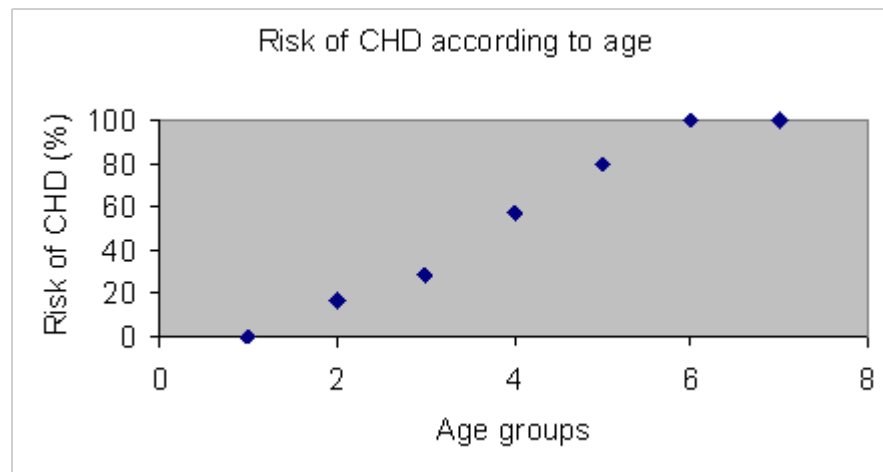
In the following table and figure, the relation between age and CHD is expressed as the proportion of persons with

CHD (risk) by 10 years age groups. The increase of risk of CHD with age is clearer and risk goes from 0 to 1 (here expressed as a %).

Table 1: Proportion of persons at risk of CHD by age group

Age group	Age group in years	Number in group	Disease	Proportion %
1	20-29	5	0	0
2	30-39	6	1	17
3	40-49	7	2	29
4	50-59	7	4	57
5	60-69	5	4	80
6	70-79	2	2	100
7	80+	1	1	100

Figure 2: Proportion of persons at risk of CHD by age group



Therefore we would be interested in identifying a transformation of the linear model which would limit the value of y between 0 and 1 in order to avoid getting impossible values for y .

The **logistic function** which is "S" shaped satisfies those constraints (figure 3).

Figure 3: The logistic function



The logistic function that we will use in logistic regression can be written as follows:

$$R = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

R (the risk) is also frequently noted as P (y/x) which is the probability of the outcome given x. In that case the above formula is:

$$P(y / x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

The logistic function needs to be transformed to become a user friendly tool. The transformation will help us keeping the values in the appropriate range. The logistic transformation includes two steps. The first is to use the odds of disease (P(y/x) / (1- P(y/x))) instead of the risk (P(y/x)). The second transformation is to take the natural logarithm of the odds of disease, ln [(P(y/x) / (1- P(y/x)))]. The result of these transformations is called the logit. The logit (ln [(P(y/x) / (1- P(y/x)))]) is the predicted value of a straight line:

$$\text{Ln} [(P(y/x) / (1-P(y/x)))] = \beta_0 + \beta_1 x_1$$

The interesting aspect of the transformation is that the exponential of the coefficient (e^{β_1}) is the ratio of the odds of disease among exposed (Oe) to the odds of disease among unexposed (Ou).

$$\beta_1 = \ln (Oe/Ou)$$

$$e^{\beta_1} = O_e/O_u = OR$$

Therefore the logistic regression is an interesting model to analyse case-control studies in which the measure of association is the odds ratio.

One of the major advantages of multivariable analysis is that it will allow controlling of confounding simultaneously in all variables included in a model. Variables would be then mutually unconfounded.

[<<Back to Logistic regression](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Fitting logistic regression models

Last modified at 9/19/2011 9:48 AM by Arnold Bosman

Once we have a model (the [logistic regression model](#)) we need to fit it to a set of data in order to estimate the parameters β_0 and β_1 .

In a linear regression we mentioned that the straight line fitting the data can be obtained by minimizing the distance between each dot of a plot and the regression line. In fact we minimize the sum of the squares of the distance between dots and the regression line (squared in order to avoid negative differences). This is called the **least sum of square** method. We identify b_0 and b which minimise the sum of squares.

In logistic regression the method is more complicated. It is called the **maximum likelihood** method. Maximum likelihood will provide values of β_0 and β_1 which maximise the probability of obtaining the data set. It requires iterative computing and is easily done with most computer software.

We use the **likelihood function** to estimate the probability of observing the data, given the unknown parameters (β_0 and β_1). A "likelihood" is a probability, specifically the probability that the observed values of the dependent variable may be predicted from the observed values of the independent variables. Like any probability, the likelihood varies from 0 to 1.

Practically, it is easier to work with the logarithm of the likelihood function. This function is known as the **log-likelihood**, and will be used for inference testing when comparing several models. The log likelihood varies from 0 to minus infinity (it is negative because the natural log of any number less than 1 is negative).

The log likelihood is defined as:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n \{y_i \ln[P(y/x)] + (1 - y_i) \ln[1 - (P(y/x))]\}$$

In which

<!--[endif]-->

$$P(y/x) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

Estimating the parameters β_0 and β_1 is done using the first derivatives of *log-likelihood* (these are called the likelihood equations), and solving them for β_0 and β_1 . Iterative computing is used. An arbitrary value for the coefficients (usually 0) is first chosen. Then log-likelihood is computed and variation of coefficients values observed. Reiteration is then performed until maximisation (plateau). The results are the **maximum likelihood estimates** of β_0 and β_1 .

Inference testing

Now that we have estimates for β_0 and β_1 , the next step is inference testing.

It responds to the question: "**Does the model including a given independent variable provide more information about occurrence of disease than the model without this variable?**" The response is obtained by comparing the observed values of the dependent variable to values predicted by two models, one with the independent variable of interest and one without. If the predicted values of the model with the independent variable is better then this variable significantly contributes to the outcome. To do so we will use a statistical test.

Three tests are frequently used:

- Likelihood ratio statistic (LRS)
- Wald test
- Score test

The **Likelihood ratio statistic** (*LRS*) can be directly computed from likelihood functions of both models.

$$LRS = -2 \ln \left[\frac{l(\beta)^-}{l(\beta)^+} \right]$$

Probabilities are always less than one, so log likelihoods are always negative; we then work with **negative log likelihoods** for convenience.

The likelihood ratio statistic (*LRS*) is a test of the significance of the difference (the ratio if expressed in log) between the likelihood for the researcher's model minus the likelihood for a reduced model (the models with and without a given variable).

The *LRS* can be used to test the significance of a full model (several independent variables in the model versus no

variable = only the constant). In that situation it tests the probability (the null hypothesis) that all β are equal to 0 (all slopes corresponding to each variable are equal to 0). This implies that none of the independent variables are linearly related to the log odds of the dependent variable.

The *LRS* does not tell us if a particular independent variable is more important than others. This can be done, however, by comparing the likelihood of the overall model with a reduced model which drops one of the independent variables.

In that case the *LRS* tests if the logistic regression coefficient for the dropped variable equals 0. If so it would justify dropping the variable from the model. A non significant *LRS* indicates no difference between the full and the reduced models.

Alternatively *LRS* can be computed from deviances.

Computations from **deviances**

$$LRS = D^- - D^+$$

In which D^- and D^+ are respectively the deviances of the models without and with the variable of interest.

The deviance can be computed as follows:

$$D^- = -2\ln \left[\frac{\text{Likelihood of model without variable}}{\text{Likelihood of saturated model}} \right]$$

$$D^+ = -2\ln \left[\frac{\text{Likelihood of model with variable}}{\text{Likelihood of saturated model}} \right]$$

(A saturated model being a model in which there are as many parameters as data points.)

Under the hypothesis that $\beta_1 = 0$, *LRS* follows a chi-square distribution with 1 degree of freedom. The derived p-value can be computed.

The following table illustrates the result of the analysis (using a logistic regression package) of a study assessing risk factors for myocardial infarction. The *LRS* equals 138,7821 ($p < 0,001$) suggesting that oral contraceptive (OC) use is a significant predictor of the outcome.

Table 1: Risk factors for myocardial infarction. Logistic regression model including a single independent variable (OC)

Model Fit Results	Value	DF	p-value
Likelihood ratio statistic	138,7821	2	< 0.001

Parameter Estimates 95% C.I.

Terms	Coefficient	Std.Error	p-value	Odds Ratio	Lower	Upper
%GM	-1,7457	0,1782	< 0.001	0,1745	0,1231	0,2475
OC	1,9868	0,2281	< 0.001	7,2924	4,6633	11,4037

In model 2, model 1 was expended and another variable was added (the age in years). Here again the addition of the second variable contributes significantly to the model. The LRS (LRS = 16,7253, $p < 0,001$) expresses the difference in likelihood between the two models.

Table 2: Risk factors for myocardial infarction. Logistic regression model including two independent variable (OC and AGE)

Number of valid Observations 449

Model Fit Results	Value	DF	p-value
Likelihood ratio statistic	16,7253	1	< 0.001

Parameter Estimates 95% C.I.

Terms	Coefficient	Std.Error	p-value	Odds Ratio	Lower	Upper
%GM	-3,3191	0,4511	< 0.001	0,0362	0,0149	0,0876
OC	2,3294	0,2573	< 0.001	10,2717	6,2032	17,0086
AGE	0,0302	0,0075	< 0.001	1,0306	1,0156	1,0459

[<<Back to Logistic regression](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Interpreting model coefficients

Last modified at 9/19/2011 9:48 AM by Arnold Bosman

The coefficients are the maximum likelihood estimates of β_0 and β_1 . After [fitting the model](#) the next step is to interpret those coefficients.

This time the question we need to answer is: **"What do the estimates of the coefficients tell us about the research question?"**

For the logistic regression model, from the logit which we will call $g(x)$, we can write the following:

$$g(x) = \beta_0 + \beta_1 x_1$$

where

$$\beta_1 = g(x + 1) - g(x)$$

It is easy to understand that the coefficient β_1 (the slope) represents the amount of change in the logit ($g(x)$) for a change of one unit in the independent variable ($x+1$ versus x).

The interpretation of this change will depend upon the measurement scales used for the independent variable.

Dichotomous variable

Let's first assess what happens when the independent variable is a dichotomous variable (e.g. yes, no).

From the formula of the logit,

$$\text{Ln} [(P(y/x) / (1-P(y/x)))] = \beta_0 + \beta_1 x_1$$

we had deduced that the exponential of the coefficient (e^{β_1}) is the ratio of the odds of disease among exposed to the

odds of disease among unexposed (see [Logistic model](#)).

$$\beta_1 = \ln(\text{Odds ratio})$$

$$e^{\beta_1} = \text{Odds ratio}$$

In case of a dichotomous variable the odds ratio is the ratio of the odds for $x = 1$ to the odds of $x = 0$. The log of the odds ratio corresponds to the difference of the two logits with respectively $x = 1$ and $x = 0$.

The odds ratio gives us an idea of how much more likely (or less likely) it is for the outcome (e.g. disease) to occur among those with $x = 1$ (e.g. exposed) as compared to those with $x = 0$ (e.g. unexposed).

The confidence interval around the odds ratio can be computed as follows:

$$e^{\left[\beta \pm Z_{1-\alpha/2} \times SE(\beta)\right]}$$

The following example shows the results of a logistic regression analysis of a study done during the investigation of a Salmonella outbreak in which the consumption of Tiramisu was suspected to be the vehicle of the epidemic. The data set includes 245 individuals including cases and controls.

Number of terms	2			
Total Number of Observations	245			
Rejected as Invalid	0			
Number of valid Observations	245			
Summary Statistics	Value	DF	p=value	
Deviance	159,2489	243		
Likelihood ratio test	180,3927	2	< 0.001	
Parameter Estimates				95% C.I

Terms	Coefficient	Std.Error	p-value	Odds Ratio	Lower	Upper
%GM	-2,9741	0,3875	< 0.001	0,0511	0,0239	0,1092
TIRA_	4,3116	0,4586	< 0.001	74,5578	30,3501	183,1579

In this output, $\beta_0 = -2,9741$ and $\beta_1 = 4,3116$

The related OR is therefore equal to $e^{\beta_1} = e^{4,3116} = 74,5578$. It expresses that the odds of gastroenteritis are 74,5578

times higher among Tiramisu consumers than not.

The confidence interval, applying the above mentioned formula is [30,3501 - 183,1579].

Polytomous variable

The independent variable may have more than 2 categories. Let suppose that we want to assess the role of the amount of Tiramisu consumed in the outcome (gastroenteritis due to Salmonella) of the above example. In this example the independent variable has four categories (no consumption, small amount, medium and large). The categories are mutually exclusive. In the data set this variable was coded 0 for no consumption, 1 for small amount, 2 for medium and 3 for large amount. If an analysis is performed using this coding we obtain the following result.

Number of terms	2			
Total Number of Observations	245			
Rejected as Invalid	0			
Number of valid Observations	245			
Summary Statistics	Value	DF	p=value	
Deviance	161,3985	243		
Likelihood ratio test	178,2431	2	< 0.001	
Parameter Estimates				95% C.I

Terms	Coefficient	Std.Error	p-value	Odds Ratio	Lower	Upper
%GM	-2,5463	0,3048	< 0.001	0,0784	0,0431	0,1424
TIRA_	2,8479	0,3440	< 0.001	17,2518	8,7904	33,8580

In the above result, *Tportion* coded as 0,1,2 or 3 is interpreted as a continuous variable. Under the assumption that the logit is linear in the continuous variable *Tportion*, the value of the OR represents the amount by which the OR is multiplied by for each increase of one unit of *Tportion*.

In our example when we go from no exposure to small (0 to 1) the OR is 17,2518. When the exposure increases from 1 to 2 (small to medium amount) the OR is also 17,2518. This is meaning that moving from 0 to 2 the OR would be $17,2518 \times 17,2518 = 297,62$. Similarly moving from 0 to 3 (no consumption versus large amount) the OR would be $17,2518^3 = 5134,56$. This obviously does not represent the relation between dose and outcome. The values 0 to 3 have no numerical meaning. They are in fact a code for categories.

In such a situation we will create design variables (also called dummy variables or factor variables). The principle is that for n categories we need to create n-1 design variables.

Using the example of amount of Tiramisu the three design variables (D1, D2 and D3) take the following values. For no consumption D1, D2 and D3 will be assigned a 0 value. For small amount D1 = 1 and D2 and D3 = 0. For average consumption D1 = 0, D2 = 1 and D3 = 0. For large amount only D3 equals 1. This is summarised in the following table. Using a common reference group for coding design variables is the most frequent method. Most logistic regression software packages will generate design variables using this method.

	Design (dummy) variables		
Tiramisu consumption	D1	D2	D3
None	0	0	0
Small	1	0	0
Medium	0	1	0
Large	0	0	1

Using the design variables in the above example, we obtain the following result.

Summary Statistics	Value	DF	p=value	
Deviance	150,6160	241		
Likelihood ratio test	189,0256	4	< 0.001	
Parameter Estimates				95% C.I

Terms	Coefficient	Std.Error	p-value	Odds Ratio	Lower	Upper
%GM	-2,9741	0,3875	< 0.001	0,0511	0,0239	0,1092
TPORTION2 ='1'	3,7518	0,4858	< 0.001	42,5966	16,4380	110,3828
TPORTION2 ='2'	5,3720	0,7168	< 0.001	215,2858	52,8285	877,3287
TPORTION2 ='3'	5,2767	1,1181	< 0.001	195,7142	21,8712	1751,3447

In the above table the odds ratio for eating a small amount (as compared to no consumption) is 42,5966. The odds ratio for eating a medium amount as compared to no consumption is 215,2858, for large amount it is 195,7142 (similar to the OR for medium portion). For the 3 categories the reference group is non consumers allowing therefore to compare odds ratios between categories of consumption.

Some important considerations on design variables merit to be noted:

- All dummy variables should be considered as a single variable.
- Each dummy variable corresponds to a degree of freedom (important in modelling).

- Dummy variables can be created to indicate different levels of exposure (dose-response analysis).
- Dummy variables can be created to indicate different levels of a quantitative variable (especially when doubting about linearity, see below).

Continuous independent variable

Some of the variables we will include in logistic models are continuous (e.g. age in years, weight in grams, height in cm, etc.). In such a case the interpretation of the coefficient will depend upon the unit chosen for the independent variable and the assumption that the logit is continuous in the dependent variable.

The logit can be expressed as:

$$g(x) = \beta_0 + \beta_1 x_1$$

Here also the coefficient β_1 gives the amount of change in the log odds (logit) for each unit of change in the independent variable.

The following table illustrates the relationship between age and occurrence of gastroenteritis due to Salmonella.

Number of terms	2			
Total Number of Observations	245			
Rejected as Invalid	6			
Number of valid Observations	239			
Summary Statistics	Value	DF	p=	value
Deviance	309,9103	237		
Likelihood ratio test	21,4135	2	< 0.001	
Parameter Estimates				95% C.I

Terms	Coefficient	Std.Error	p-value	Odds Ratio	Lower	Upper
%GM	-0,6164	0,2914	0,0344	0,5399	0,3050	0,9556
AGE	0,0001	0,0100	0,9882	1,0001	0,9808	1,0199

The logit including age as a continuous variable is:

$$g(\text{age}) = -0,6164 + (0,0001 \times \text{AGE})$$

For one year increase of age the OR is multiplied by 1,0001 assuming linearity between age and log odds of gastroenteritis. For 10 years increase the associated OR would be:

$$\text{OR (10years)} = \exp (10 \times 0,0001) = 2,72$$

In the above example age seems to increase slightly the risk of illness.

However if the logit is not linear then we should choose another way of analysis, e.g. creating categories of age (age groups) and develop the related design (dummy) variables or using other regression analysis techniques.

[<<Back to Logistic regression](#)



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Restriction

Last modified at 9/3/2014 5:07 PM by Arnold Bosman

Restriction is an *alternative* to matching, which is used to *prevent* [confounding](#) in a study during the stage of the study design [1]. Restriction consists in limiting the entrance to the study on a restricted number of subjects, on the basis of a possible confounder. For example, if we think that gender is a confounder, we may enroll in our study only males. In this way we will be able to study the association, for example, on the risk of cancer X and drinking whisky, without the need to account for gender as a potential confounder.

The disadvantage of using restriction is that we will not be able to quantify the absolute effect of drinking whisky and cancer X in the population, but our estimate will just represent the association in the male group.

It is also possible to perform a "restricted analysis", by conducting an analysis in one specific stratum, once a study has been conducted with a unmatched or matched design. This could mean to enroll males and female in the study and then to perform a sub-analysis only with one gender.

Obviously the consequence is that the subsequent results would only apply to that stratum of the population and could not be generalised. Sample size would also have to be calculated accordingly in order to maintain a sufficient power in the study.

References

-
1. Rothman KJ; Epidemiology: an introduction. Oxford University Press 2002, p.108-111.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Advantages and disadvantages of matching

Last modified at 10/28/2010 8:26 PM by Lisa Lazareck

Advantages of matching

Matching is a useful method to optimize resources in a case control study.

Matching on a factor linked to other factors may automatically control for the confounding role of those factors (e.g. matching on neighborhood may control for socio-economic factors).

Matching allows to use a smaller sample size, by preparing the stratified analysis "a priori" (before the study, at the time of cases and control selection), with smaller sample sizes as compared to an unmatched sample with stratified analysis made "a posteriori".

Matching avoids a stratified analysis with too many strata, with potentially no case or control, done to control several confounding factors at the same time. Indeed, in an unmatched case control study, while we perform [logistic regression](#), or even more simply a [stratified analysis](#), we might end up with empty strata (no cases or no control in some strata). Matching avoids this situation.

Disadvantages of matching

The efficiency in data analysis that matching provides is limited by several disadvantages.

The greatest disadvantage of matching is that the effect of matching factor on the occurrence of the disease of interest cannot be studied anymore. *One should therefore limit matching to factors that are already known to be risk factors for the studied outcome.*

If statistical softwares with [logistic regression](#) are available, it is possible to *control* for many confounding factors during the analysis of the study, and therefore *preventing* confounding by matching during the design of the study might not be needed, especially if the study is including a large population and there are few chances that we will end up with empty strata.

If matching is performed, it must also be taken into account in the statistical analysis, because a matched OR needs to be calculated, and conditional logistic regression need to be used.

However the study of the matching factor as an [effect modifier](#) is still possible if doing a stratified analysis over several categories of the matching factor. For example when matching on age, analysis is still feasible within each age stratum created. However to use different age categories than those used for matching would require a multivariable analysis. Trying to identify a [dose response](#) involving a matching factor would also require a multivariable model of analysis.

Matching on criteria that are only associated with exposure and not with outcome further biases the [measurement of the effect](#). In this situation the matching factor is not a confounding factor and matching would bring the OR towards 1.

Another difficulty occurs when matching on several factors. It then becomes difficult (time and energy) to logistically identify and recruit controls due the high number of matching factors (e.g. same age, sex, socio economic status, occupation, etc.). Matching on several criteria may improve the efficiency of statistical analysis with a reduced sample size but the difficulties to recruit controls may jeopardize that efficiency. It may also exclude cases for which no matched controls can be identified. In addition, matching on many criteria increases the risk of matching on exposure (therefore bringing the OR closer to one). This is sometimes called *overmatching*.

One major challenge when matching is to properly define the various strata of the matching variable. For example when frequency matching on age, we need to make sure that, within each of the age group created, age is no longer a confounding factor. This is sometimes called [residual confounding](#). Several analysis with several width of age strata may be tested. For example, let's suppose we stratify on several age groups 20 years wide (0-19, 20-39, 40-59, 60-79, 80+). To assess if age is still a confounder within one age group we could further stratify (by five years age group) and test if age is still a confounding factor inside a 20 years wide age group. So it may still be important to take account of age as a potential confounder in a multivariable analysis.

Dose Effect

Last modified at 9/14/2011 1:26 PM by CeRC

So far we have only considered that persons could be either exposed or not, but exposure can frequently be quantified and a gradient established. Examples may include the number of glasses (or volume) of water consumed per day, the number of pizza slices eaten, or the degree of completion of a vaccination course etc. The risk of illness can then be calculated according to amount drunk, eaten or completed.

Amount drunk	Exposed	Cases	Risk (%)	RR
None	100	5	5	ref
1 glass	100	11	11	2.2
2 glasses	100	16	16	3.2
3 glasses	100	22	22	4.4
≥4 glasses	100	28	28	5.7
	500	82	82	

The X^2 test for trend across these strata = 23.24, $p \leq 0.001$ - indicating that there is a trend across the exposure strata.

If instead a case control study had been conducted (note the similarity between the OR and the RR above):

Amount drunk	Exposed	Cases	OR
None	5	30	ref
1 glass	11	30	2.2

2 glasses	16	30	3.2
3 glasses	22	30	4.4
≥4 glasses	28	30	5.6
	82	150	

For which the X^2 test for trend = 12.77, $p \leq 0.001$

In the above examples, the risk of disease occurring increases with the amount of water consumed (i.e. with the dose of water). This is frequently called a dose response or a dose effect relationship.

To compute the risk ratio or the odds ratio the same reference level of exposure is used. Here the non exposed group is used as a reference. Alternatively the group with the lowest incidence could be used.

Dose response is particularly interesting when in an outbreak (often in water borne or a food borne outbreak) everybody is exposed to the same factor but the level of exposure differs among individuals. Choosing the lowest category of exposure as a reference still allows identification of a food item or a beverage as the vehicle for the outbreak. To test that the trend observed in risk or odds ratio differs from no trend, we can use a chi square for trend [1-3]. Dose response is also an important criterion to consider when providing evidence for [causality](#) proposed by Bradford Hill [4].

References

1. J.J.Schlesselman. Case-control studies: design, conduct, analysis. New York: Oxford University Press; 1982.
2. B.R.Kirkwood, J.A.C.Sterne. Medical Statistics. Second ed. Massachusetts: Blackwell Science Ltd; 2003.
3. K.J.Rothman, S.Greenland, T.L.Lash. Modern Epidemiology. Third ed. Philadelphia, USA: Lipincott Williams and Wilkins; 2008.
4. A.Bradford Hill. The Environment and Disease: Association or Causation? Proceedings of the Royal Society of Medicine 1965;58:295-300.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Residual Confounding

Last modified at 10/28/2010 8:04 PM by Lisa Lazareck

Residual confounding

Ideally, in order to account for confounding between exposure and disease, we stratify according to the confounding variable. When the confounding variable lends itself to few strata (i.e. gender, which has two strata: male or female), there tends to be sufficient data for results of the stratification to have sufficient power: lending to an overall strength of inference.

Strata can have a biological meaning or a quantitative rationale. Whatever the choice, the variable of interest is no longer a confounder within the stratum. If the dataset is stratified according to 10 year age groups, then it must be verified that age is no longer a confounder within those 10 year age group strata. If we stratified further to examine for confounding, there would be cells containing little or no data: which would need to be compensated for using modeling assumptions.

Typically these assumptions form a regression model - which will never be entirely correct. The bias remaining is 'residual confounding' - that which remains after confounding has been adjusted for as much as possible [\[1\]](#).

Cancelling out

Factors may not appear to be confounders if they cancel each other out.

In a (hypothetical) study examining the risk of lung cancer amongst those exposed to silica dust; exposure is found to be a risk factor for disease. However, if 50% of those exposed to silica dust were also tobacco smokers, while only 30% of the unexposed were tobacco smokers, then smoking appears to be confounding the relationship between exposure to silica dust and development of lung cancer.

If the exposed are also younger than the unexposed, but the young are less likely to develop lung cancer (regardless of smoking); then the two confounders (age and smoking) are affecting the original association (between exposure to

silica dust and lung cancer) in opposite directions. Neither may appear to be a confounding variable; they have cancelled each other out.

It is hoped that situations such as these are rare, but while an exact cancellation of two confounding factors may be rare, they will both have an effect on the total confounding within the model [2].

References

1. K.J.Rothman, S.Greenland, T.L.Lash. Modern Epidemiology. Third ed. Philadelphia, USA: Lipincott Williams and Wilkins; 2008.
2. McNamee R. Confounding and confounders. *Occupational and Environmental Medicine* 2003 Mar 1;60(3):227-34.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Updated
Pages

User
manual
for
editors

Legal
notice

Help and
Support

About

User
manual
for
approvers

Legal notice

PRIVACY STATEMENT

1. PURPOSE OF THE PROCESSING OPERATION

ECDC processes the personal data collected in accordance with Regulation (EU) 2018/1725. The purpose of the processing is the following:

FEM Wiki is an open information sharing platform for public health experts, hosted and funded by ECDC. The content of FEM Wiki is provided by users of the platform and does not necessarily represent opinion of ECDC. By contributing content to FEMWIKI, users agree to the conditions as described under Creative Commons Licence and to FEM Wiki users' Code of Conduct.

2. IDENTITY OF THE DATA CONTROLLER

European Centre for Disease Prevention and Control (ECDC), Gustav III:S Boulevard 40, 16973 Solna, Sweden
PHC, Public Health Training, FEMWIKI@ecdc.europa.eu

3. LEGAL BASIS FOR THE PROCESSING

The legal basis of the processing operation is:

- Article 5 (d) the processing is based on consent
- ECDC Founding regulation 851/2004, specifically article: 11 and 3 (c)

4. CATEGORIES OF PERSONAL DATA COLLECTED

The categories of data collected and used for the processing operations are the following:

Name (first name and surname); E-mail; Phone Number; Address; Unit; Work Position/Occupation; Picture;

The provision of the personal data is not mandatory.

The processing of your data will not be used for an automated decision making, including profiling.

5. WHO HAS ACCESS TO YOUR INFORMATION AND TO WHOM IS IT DISCLOSED?

The recipients of the data are any users of FEM Wiki - the general public.

6. HOW LONG DO WE KEEP YOUR DATA?

ECDC will retain the data as long as necessary.

7. HOW DO WE PROTECT AND SAFEGUARD YOUR INFORMATION?

In order to protect your personal data, a number of technical and organisational measures have been put in place. Technical measures include appropriate actions to address online security, risk of data loss, alteration of data or unauthorised access, taking into consideration the risk presented by the processing and the nature of the data being processed. Organisational measures include restricting access to the data to authorised persons with a legitimate need to know for the purposes of this processing operation.

measures include appropriate actions to address online security, risk of data loss, alteration of data or unauthorised access, taking into consideration the risk presented by the processing and the nature of the data being processed. Organisational measures include restricting access to the data to authorised persons with a legitimate need to know for the purposes of this processing operation.

8. WHAT ARE YOUR RIGHTS AND HOW YOU CAN EXERCISE THEM?

The Controller may be contacted at any time by the data subjects for exercising the right of access, to rectify, to block, to erase, to transmit or to object to the processing of the data. Where the legal basis to the processing is consent, this consent can be withdrawn at any time. The Controller can be contacted on: FEMWIKI@ecdc.europa.eu

Data subjects can request the deletion of their personal data by the data controller, who will do so within 45 working days.

Data subjects can also contact the ECDC Data Protection Officer (DPO) in case of any difficulties or for any questions relating to the processing of their personal data at the following email address: dpo@ecdc.europa.eu. The data subject has the right of recourse at any time to the European Data Protection Supervisor: www.edps.europa.eu and at edps@edps.europa.eu

Exceptions based on Regulation (EU) 2018/1725 may apply.

DISCLAIMER

ECDC maintains FEM Wiki to enhance open information sharing between public health experts. The goal is to keep this information and materials accurate and up to date. If errors are brought to our attention, any necessary corrections will be made as soon as possible.

Please note that the information available is:

- not intended to address the specific circumstances of any particular individual or entity;
- sometimes linked to external sites over which ECDC has no control and for which ECDC assumes no responsibility;
- not legal advice (if you need specific advice, legal or otherwise, you should always consult a suitably qualified professional.)

Please note that it cannot be guaranteed that a document available online exactly reproduces an officially adopted text. Thus, in case of disagreement, the official adopted text will always prevail.

It is our goal to minimise disruption caused by technical errors. However, some data or information may be in files or formats that are not error-free and we cannot guarantee that our service will not be interrupted or otherwise affected by such problems. ECDC accepts no responsibility for such problems incurred as a result of using this website.

The ECDC logo may not be used without ECDC's prior written consent.

This disclaimer is not intended to limit the liability of ECDC in contravention of any requirements laid down in applicable national, EU or international law, nor to exclude its liability for matters which may not be excluded under the relevant law.

COPYRIGHT AND LICENSING

3.1. Licenses to use the material

The material contained on the website have a Creative Commons BY license. This means that reuse is allowed, provided appropriate credit is given and any changes made to the material are indicated. For more information, please go to:

<https://creativecommons.org/licenses/by/4.0/> .

3.2. License applicable to the content provided by contributors

ECDC does not claim ownership of any content you may submit or make available for inclusion on the website or courses. However, by submitting or making available any content, you agree that such content shall be covered by a Creative Commons BY 4.0 license.

ECDC reserves the right to remove any content without prior notice at any time and for any reason, in particular if any of the terms of the FEM Wiki Code of Conduct are violated.

ECDC reserves any other rights over content provided by users it may have, for example under other licenses granted in ECDC's favour.

To the extent that you provide any content, you represent and warrant that:

(a) you have all necessary rights, licences and/or clearances to provide such content and permit ECDC to use and publish such content as provided above;

(b) such content is accurate and complete to the best of your knowledge;

(c) you are responsible for the payment of any third party fees related to the provision, publication and use of such content; and

(d) such use and/or publication of your content does not and will not infringe or misappropriate any third party rights or constitute a fraudulent statement or misrepresentation.

Any content included on the website or within the courses that infringes the intellectual property rights of any third party will be removed as soon as possible after ECDC is made aware of such infringement.

3.3. How to address the infringement of your intellectual property rights?

If you are the owner of intellectual property rights, or otherwise authorised to exercise these rights, you may report any alleged infringements of these rights by emailing the administrators at FEMWIKI@edc.europa.eu.

In order that ECDC may react as quickly as possible, please include the following:

- a. details of the alleged infringement of your intellectual property rights, including the title of the content concerned and the full URL for access to that content ;
- b. which country your intellectual property rights apply to;
- c. a statement explaining how the content allegedly infringes your intellectual property rights;
- e. your mailing address, telephone number and email address so that ECDC can contact you;
- f. a statement that the information contained in the notice is accurate and that you are the owner of the intellectual property rights or have an exclusive legal right to bring infringement proceedings in respect of its use; and
- g. your signature (an electronic signature is sufficient).

ECDC has discretion to take remedial action if required, which may include the removal of the challenged content.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Updated
Pages

User
manual
for
editors

Legal
notice

Help and
Support

About

User
manual
for
approvers

User manual for editors

Wiki Editing (authenticated user, aka contributor)

How to behave - Code of conduct

FEM Wiki is an open information-sharing platform for public health experts, hosted and funded by ECDC. By accessing FEMWIKI, users agree to act in accordance with this Code of Conduct.

Users of FEM Wiki shall comply with the provisions of the [Legal Notice](#). Exchanges of information, views and dialogue via FEM wiki shall be undertaken in a respectful and courteous manner. Misuse of FEM Wiki may result in the deactivation of user accounts and is considered to include the following:

- Breaches of the Legal Notice
- Disrespectful, improper behaviour towards other users
- Misuse of other users' personal data or site content
- Advertising commercial products or services or recruitment offers
- Any form of technical attack on the platform

Should non-compliance with this Code of Conduct or any other misuse be identified, ECDC reserves the right to deactivate accounts at its discretion.

If you have any questions or require further guidance on permitted use of the platform, please contact femwiki@ecdc.europa.eu.

Manage your account

Sign in

In order to contribute to FEM Wiki you need to sign in on the right-upper corner. If you don't have an account there is an option to create one.

Forgotten password

If you have forgotten your password click on the Sign in page from where a password reset function is available.

Other

If you encounter any problem you cannot solve yourself contact s at FEMWIKI@ecdc.europa.eu

Creating a new Wiki Page (embed you page in the right spot in the taxonomy)

There are two main ways to create a new page in your wiki library:

Create a link to a page which does not exist and then click on it to create the page:

This is the recommended way to create a page because it is easier for people to find the page when another page links to it. Links to pages that do not yet exist have a dashed underline.

Create a page that is not linked to any other:

In the Settings menu, click Add a page. This will ask you for a name and then create that page in the current wiki library.

Frequency of crawling

Changes usually take place after 24 hours so don't be surprised if your page or edits are not visible immediately.

Deleting a page

To prevents accidental page deletions this permission is reserved to administrators only. Please write an e-mail to FEMWIKI@ecdc.europa.eu if you need a page deleted.

Updating existing wiki pages (Editing)

To edit this page, click on the Edit Page icon at the top of the page. To see all available commands, click on the Page tab. When you are editing, you can type text onto the page or insert tables and pictures. To stop editing, click the Save button at the top of the page. If you leave the page while editing, you will be prompted to save your changes.

All authenticated users see the edit button under the page ribbon.

Version control including compare versions and restore an old version.

If you need to restore a previous version of a page, click Previous Versions in the Page tab at the top of the page. You can then click on any of the versions in the quick launch to view the page as it existed at that time. If you want to restore to a version, select it and click "Restore" this version in the toolbar.

Style Guidelines for Editors

Adding Images

When you are uploading images, please check that you have uploaded a image that is clearly visible. For example, if you have

uploaded a graph, all axis labels, axis tics, titles, etc. should be legible.

There are two locations that you can insert Images from:

- From your local computer
- From Address: by giving the Internet URL

Hyperlinks

You can link to another page in this wiki library by enclosing the name of the page in double brackets. When you type `<[the wiki` will suggest page names that start with what you've typed. You can select one of these pages using the arrow keys or the mouse, or you can type in a new name. If you type in a new name, it will make a link to a page which has not yet been created. Links are finished by typing `>`. If you are linking to a page in a subfolder or an item in another list, you can press tab to add the selected suggested item to your link without closing the link.

You can link to many objects in SharePoint, not just pages. Here are some examples of links:

`<[Dogs]` : A link to a page named Dogs in the same folder.

`<[Dogs]` : A link to a page named Dogs in a subfolder called Animals.

`<[Welcome]` : A link to the item called Welcome in the Announcements list on this site.

To create a link to a page and have the link display different text than the page name, type a pipe character (|) after the page name, and then type the display text. For example, type `<[Home Page]` to create the link labeled Home Page that points to the page named Home.

To display double opening or closing brackets without making a link, type a backslash before the two brackets. For example, `<[or]`.

References

Rather than a conventional book or article format where references appear together at the end of the document, the FEM Wiki will use footnotes for references, much like those on Wikipedia. If the content of a page requires a reference, that reference should appear at the bottom of the citing page.

We will adhere broadly to the Vancouver style of referencing (useful article here), i.e:

Author of article AA, Author of article BB, Author of article CC. Title of article. Abbreviated Title of Journal. year; vol(issue):page number(s).

Equations

The FEM Wiki does not include a built-in equation editor. The solution is to create an image of the equation you want to include.

How to upload files

Files should be uploaded into the "Documents" section of FEM Wiki.

Forums

What are Forums?

Forums are places where users can publicly communicate around certain themes, e.g. Manual content or teaching structure.

Freeform discussion

The standard forum type which supports general discussion, e.g. the EPIET forum.

Q & A

The 'Ask the Expert' forum is currently the only example of this. It lets the user post a question and then mark it as answered when they judge that it has been.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Updated
Pages

User
manual
for
editors

Legal
notice

Help and
Support

About

User
manual
for
approvers

User manual for approvers

Wiki revisions

All the wikis on the FEM Wiki site follow the wiki principle of storing the entire history of revisions for each of its pages, including when the revision was made and by whom. You can see the history of all versions via the History tab where you can also compare versions. This allows anyone to change the page since it makes it easy to spot when a page has taken a turn for the worse and recover it.

Approved versions

Given the importance of the FEM Wiki Manual content, we have also added the facility to mark a specific page version as approved to provide confidence in the reliability of its content.

Approval can be performed by any one of a designated subset of group members recognised as domain experts with a single button click using the approval widget described below. The most recent version that has been approved is then featured.

Approval widget

The approval widget is only shown to moderators (and owners) of the FEM group and sits alongside Manual pages. For new pages, the widget will appear as below:

Once a page has been approved, it cannot be approved again so the button is removed and the following displayed:

If the page is revised further then the widget again displays the approval button with an appropriate message:

Finally, if an old version is visited then the following link to the latest approved version is displayed:

Notes

Since we are only featuring the most recent approved version, it makes sense to restrict approvals to versions that are more recent than the current most recently approved version.

The date and time of the approval is recorded and the most recently approved version is featured. The contributing editors of a page are listed alongside it under the header 'Approvers'.

There is currently no 'unapprove' button. Rather than attempt to 'retract' approvals, recommended practice is to revise content then approve that.

Validate Taxonomy and Tagging fit

The FEM Wiki content is divided into a number of chapters. Pages in each of the chapters have a latest community version (corresponding to the version of the page that was most recently edited by a member of the FEM), and an expert reviewed version (the last version that was reviewed by a domain expert).

On this page, we give a simple overview of the concepts behind the taxonomy and tagging before digging into the detail of how this is implemented on the FEM Wiki site.

Taxonomy

The contents of the FEM Wiki Manual have a tree-like structure of concepts, sub-concepts and so on - implementing a taxonomy. This lets users 'zoom in' on topics of interest whilst also highlighting topics from the same family.

Tagging

All pages can be assigned any number of tags - or keywords - which provide a way to bundle pages together around a theme or topic. They can be applied freely wherever a user feels a tagging would be useful via the Tag link shown below:

Tagging is generally better way to add structure to the Manual without the significance of restructuring the taxonomy.

A simple example

An example using a simple biological taxonomy is given below with the carnivore, colourful and upright tags.

Site usage

We now describe how the above abstract models are implemented on the FEM Wiki.

Tagging

Tagging is very simple, lightweight and reversible. Any number of tags can be added to a page (or discussion) as it is being created or any time after. They can be typed as a comma-separated list or selected from a list of previously-used tags. It is good practice to select existing tags, where possible, rather than creating arbitrary synonyms.

Taxonomy

The position of a page in the taxonomic structure is set entirely by the Parent Page field for the page during editing. Since all registered users can generally edit any page, they can also edit the taxonomic structure. However, this should only be done after gaining a consensus on the forum since it is a significant change.

Adding a page is a more straightforward process. If the page could be better placed, then a confident user can move it (via its parent) to a more appropriate location at an early point.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Updated Pages

User manual
for editors

Legal notice

Help and
Support

About

User manual
for
approvers

Help and Support

User Manuals

The help and support area of [FEM Wiki](#) is grouped according to the type of user.

Click a user role to see instructions for specific actions for each role.

- [User manual for FEM Wiki readers](#)
- [User manual for FEM Wiki editors](#)
- [User manual for FEM Wiki approvers](#)

FAQs

Can't edit a page?

If you want to enter any content you must ensure that:

- you are logged in (right upper corner)
- you have subscribed the FEM Wiki application in the [ECDC subscriptions page](#).

Can't delete a page?

To prevent accidental page deletions this permission is reserved to administrators only. Please write an e-mail to FEMWIKI@ecdc.europa.eu if you need a page deleted.

Contacts

If you need technical support with your account, please address ECDC's ICT helpdesk by e-mail to

ICT.FrontOffice@ecdc.europa.eu

If you can't find the help you need in the documents above, please send an e-mail to FEMWIKI@ecdc.europa.eu.



Field Epidemiology Manual Wiki

Assessing the burden of disease and risk assessment

General Communication

Infection control and hospital hygiene

Introduction to Public Health and basic concepts

Statistical Concepts

Updated
Pages

User
manual
for
editors

Legal
notice

Help and
Support

About

User
manual
for
approvers

About

The FEM Wiki site is organised into: **Articles**, **Documents** and **Discussions**.

Articles

The FEM Wiki Manual **Articles** section centres around the Manual - a publicly available Wiki for Field Epidemiology that is organised by concept (see [Taxonomy](#)).

Documents

This section works like a folder structure where images, presentations or text documents can be uploaded to be referenced in the wiki articles and discussion topics.

Discussions

This section has:

- A place to [discuss Wiki Articles](#)
- a place to [ask the Expert](#), a public Question & Answer services
- a general interest forum on [disease prevention and control](#)
- a forum on [Epidemiologists in Europe – important persons](#)
- discussing [Manual content](#)
- [Feedback and Bugs](#) on the working of the site

There is also a section with **Help & Support** the **Legal notice** and '**About**' FEM Wiki (you are here now).

Help & Support

This section gathers together all information that might be useful for users interacting with the site. Where the wiki does not answer a query, the forum is provided to allow the community to assist. Answers in the forum can be used to grow the wiki.

Legal notice

Here you find information about the legal basis for FEM Wiki, the Privacy statement, and what you need to know about Licencing.

The History FEM Wiki

The Field Epidemiology Manual (FEM) was initially developed to support the European Programme for Intervention Epidemiology Training (EPIET). Trainers, supervisors, scientific coordinators, and facilitators created draft chapters using the lectures they delivered during the EPIET introductory course. The philosophy of sharing and building knowledge (in particular training materials) led to the idea and creation of a collaborative information space for the epidemiological training community - The FEM Wiki.

The aim of the FEM Wiki was to create a library of online resources for field epidemiology training programmes, as well as a market place for discussions on all theoretical and practical aspects of field epidemiology. Central to this effort will be the Field Epidemiology Manual, a core document for use when undertaking field investigations and research projects in intervention epidemiology.

As time went on, articles from other key disciplines for disease prevention & control such as public health microbiology, informatics and public health law were added. The aim of the FEM Wiki was to create a library of training materials for training programmes for all disciplines in disease prevention & control. In addition to [articles](#) (which are still available on this new platform), the FEM Wiki originally contained curricula for training programmes, core competencies, guidelines and work space for communities. These 5 pillars formed the FEM Wiki collaborative platform. [Curricula](#), [core competencies](#), [guidelines](#) and [communities of practice](#) are now hosted on other more adapted ECDC platforms.

FEM Wiki is now a resource for public health experts, hosted and funded by ECDC. The content of FEM Wiki was provided by users of the platform and does not necessarily represent opinion of ECDC.

FEM Wiki is not: an encyclopaedia, an indiscriminate collection of information, a publisher of original thoughts (though it may happen), or a newspaper.