II INTERNATIONAL SUMMER SCHOOL

Rare disease and orphan drug registries

Day 1
15.09.2014

Epidemiological study design, data sources, variables

Yllka Kodra MD PhD
National Centre for Rare Diseases
National Institute of Health, Rome, Italy

Organised by Istituto Superiore di Sanità
Rome (Italy), September 15-19, 2014


POPULATION
STUDY
DISEASE
EPIDEMIOLOGY
HEALTH
CONTROL
PUBLIC
BIAS
CASE
RISK
EFFECT
ERROR
MORE
ONE
CAUSE
RANDOM
BETWEEN
BASED
RESEARCH
CAUSAL
MEDICINE
STUDIES
SOURCE
TIME
COHORT
NON
SOCIAL
CLINICAL
EPIDEMIOLOGICAL
SYSTEMATIC
STATISTICAL
BIOLOGY
EXPOSURE
MANAGEMENT

Some epidemiological concepts

- Descriptive vs Analytic
- External Validity vs internal Validity
- Outcome vs Exposure
- Prospective vs Retrospective
- Prevalence vs Incidence

Descriptive Studies

- characterize the disease distribution in relation to: person, place, time (who, where, when);
- the first stage of epidemiological study and generate hypotheses;
- useful data for health planning and to evaluate the burden of diseases but not to understand the cause of a disease;
- simple statistical analysis (descriptive analysis);
- does not include a comparison group;
- generate hypotheses;
- non longitudinal data;
- provide rapid information.

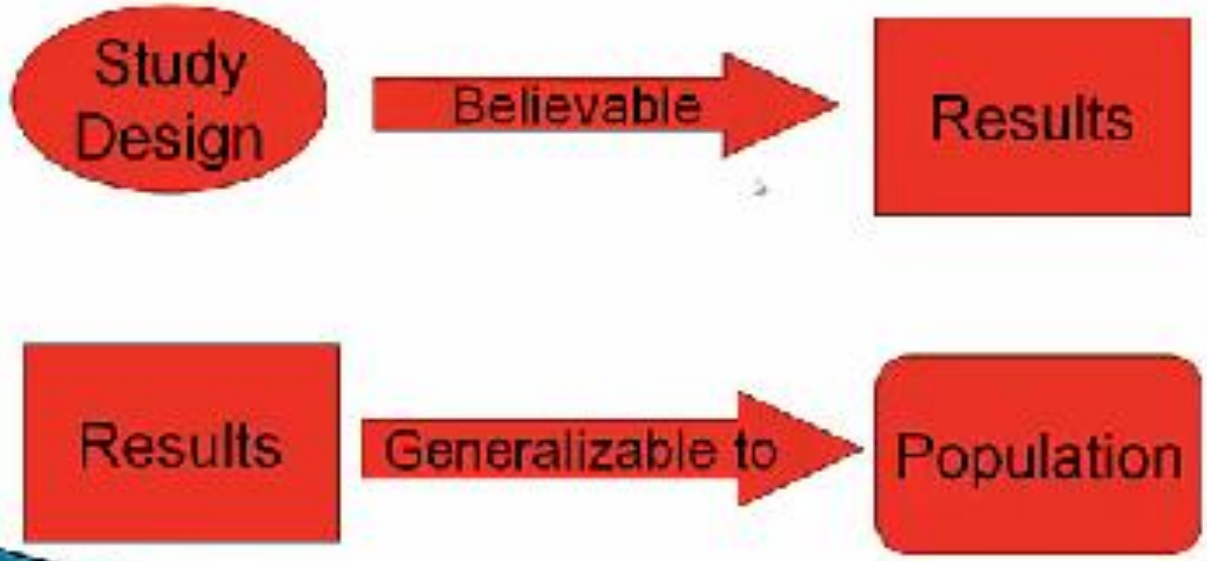
Analytic Studies

- understand the causes of a disease, risk factor/s, prognosis or treatment effect, analyse the strength of association between exposure and outcome (why and how);
- the second stage in which hypotheses generated in the descriptive phase are tested;
- useful data to understand the cause/s of a disease but not for health planning and burden of diseases;
- complex statistical analysis (inferential analysis);
- include a comparison group;
- follow-up is essential;
- don't provide rapid information.

Validity

- **Validity** refers to whether a study is able to answer accurately the questions it is intended to answer.
  - ***internal validity*** the extent to which a measurement accurately reflects what it is intended to reflect;
  - ***external validity*** the generalizability of the information to broader settings of patients, physicians, and health care settings.

Internal Vs. External Validity

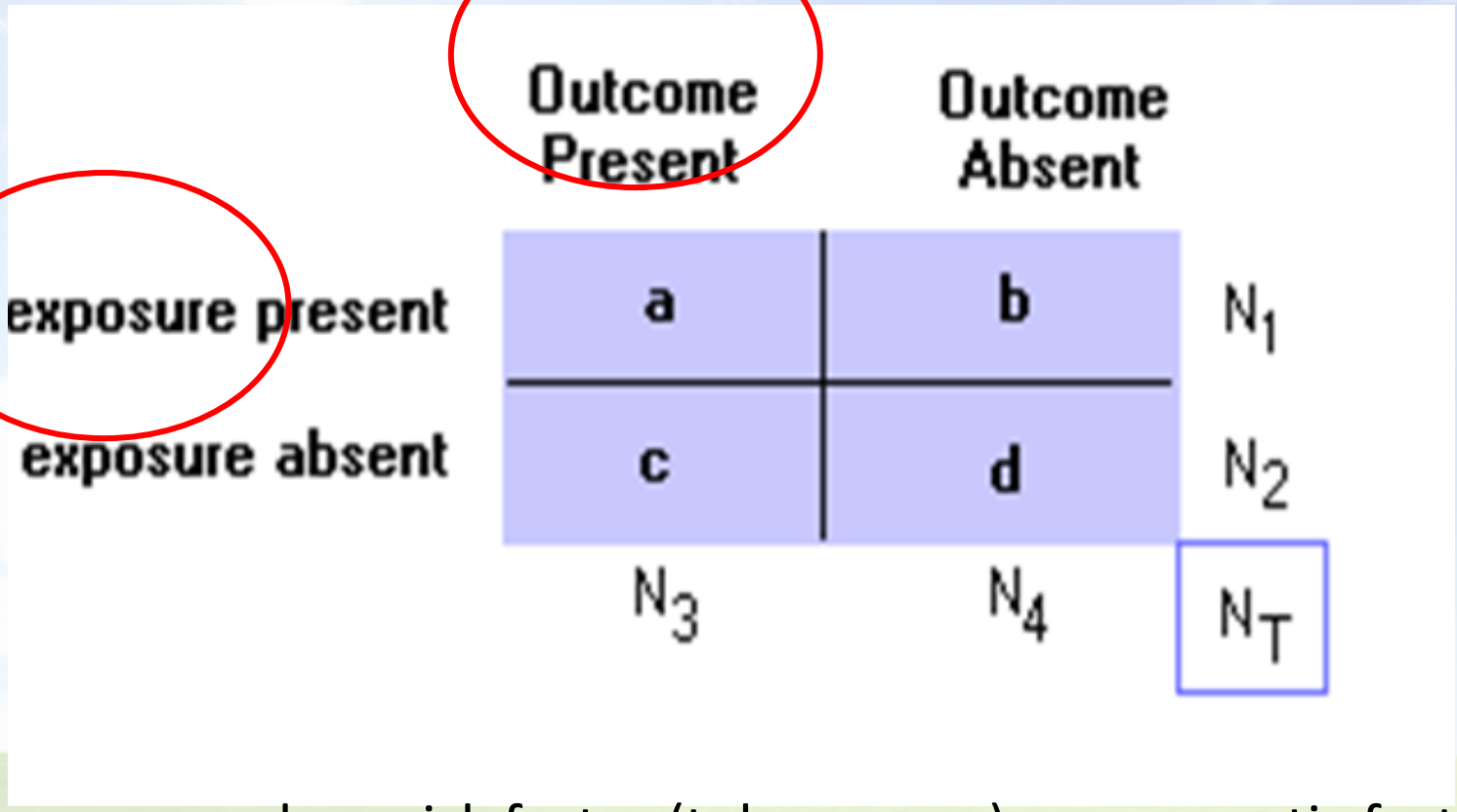


Descriptive Study

- characterize the disease distribution in relation to person, place, time (who, where when)
- the first stage of epidemiological study and generate hypothesis
- useful data for health planing and burden of diseases but not to undestand the cause of disease
- simple statistical analysis (descriptive analysis)
- not include a comparision group
- generate ipothesis
- non longitudinal data
- provide tempestive information

Analytic Study

- understand the cause/s of a disease, risk factor/s, prognosis or treatment effect, **analyse the strength of association between exposure and outcome** (why and how)
- the second stage in which hypotheses generated in the descriptive phase are tested.
- useful data to undestand the cause of disease but not for health planing and burden of diseases
- Complex statistical analysis (inferenzial analysis)
- include a comparision group
- follow-up is essential



- Exposure can be a risk factor (tobacco use), a prognostic factor, a diagnostic test, or a treatment.
- Outcome development of a disease or symptom (e.g. lung cancer) or curing a disease or symptom (e.g. reduction of pain).

Example

- *Does alcohol intake increase the risk of lung cancer?*

Alcohol → lung cancer

(exposure) (outcome)

Does treatment X improve the disease prognosis?

Treatment X → Death or remission

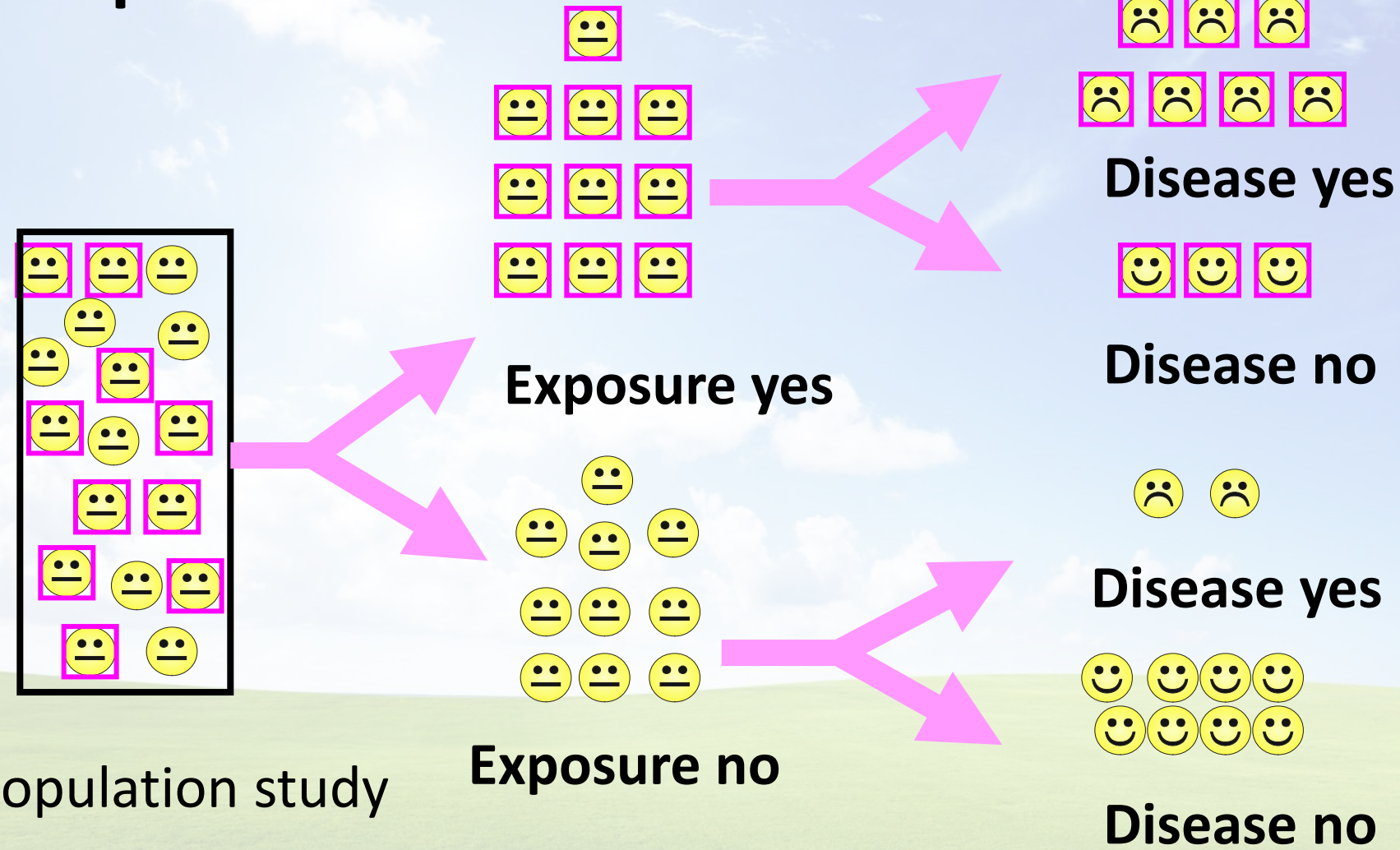
(exposure) (outcome)

Prospective vs **Retrospective**

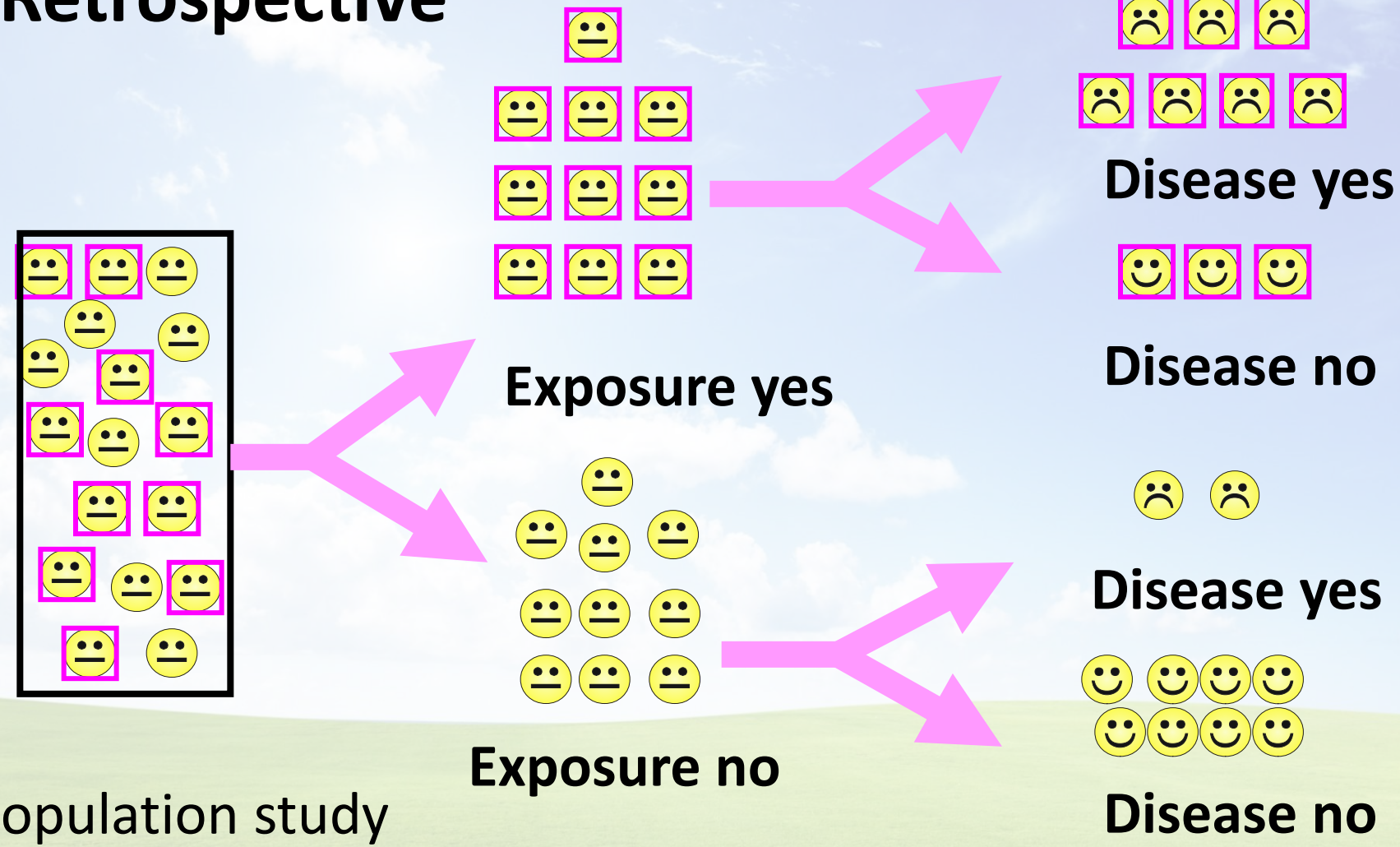
- **Prospective:** a group of people exposed are observed to see whether they develop the outcome.

- **Retrospective:** the same methodology but follow-up retrospectively. They use data already collected for other purpose. Offer advantage in term of cost and speed of execution.

Prospective



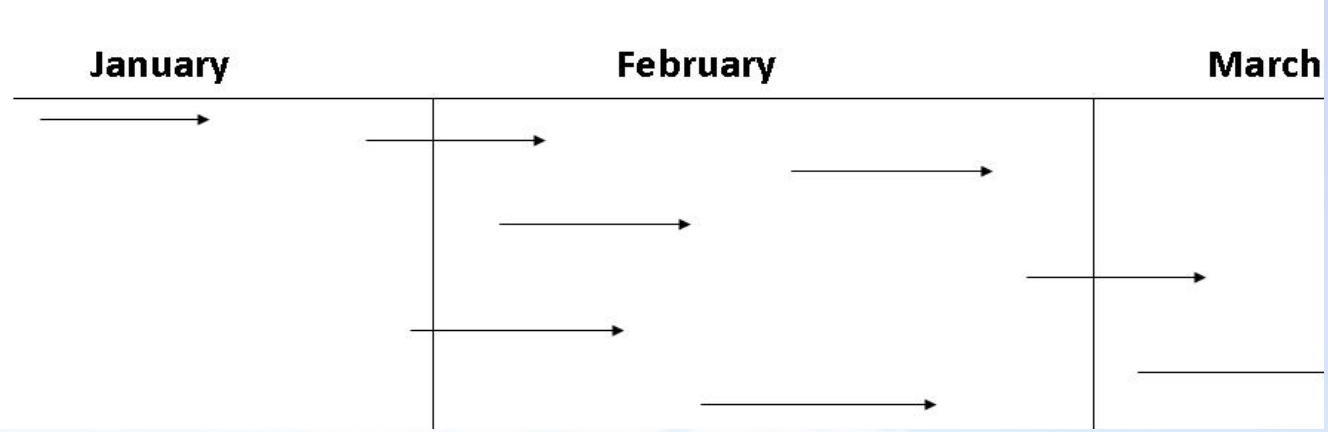
Retrospective



Prevalence vs incidence

- Prevalence: how many people (ALIVE) have this disease right now?
- Incidence: how new cases are diagnosed during the 2014?

Example



Prevalent cases at 28 febbrary =1

Incidence cases in febbrary are =4

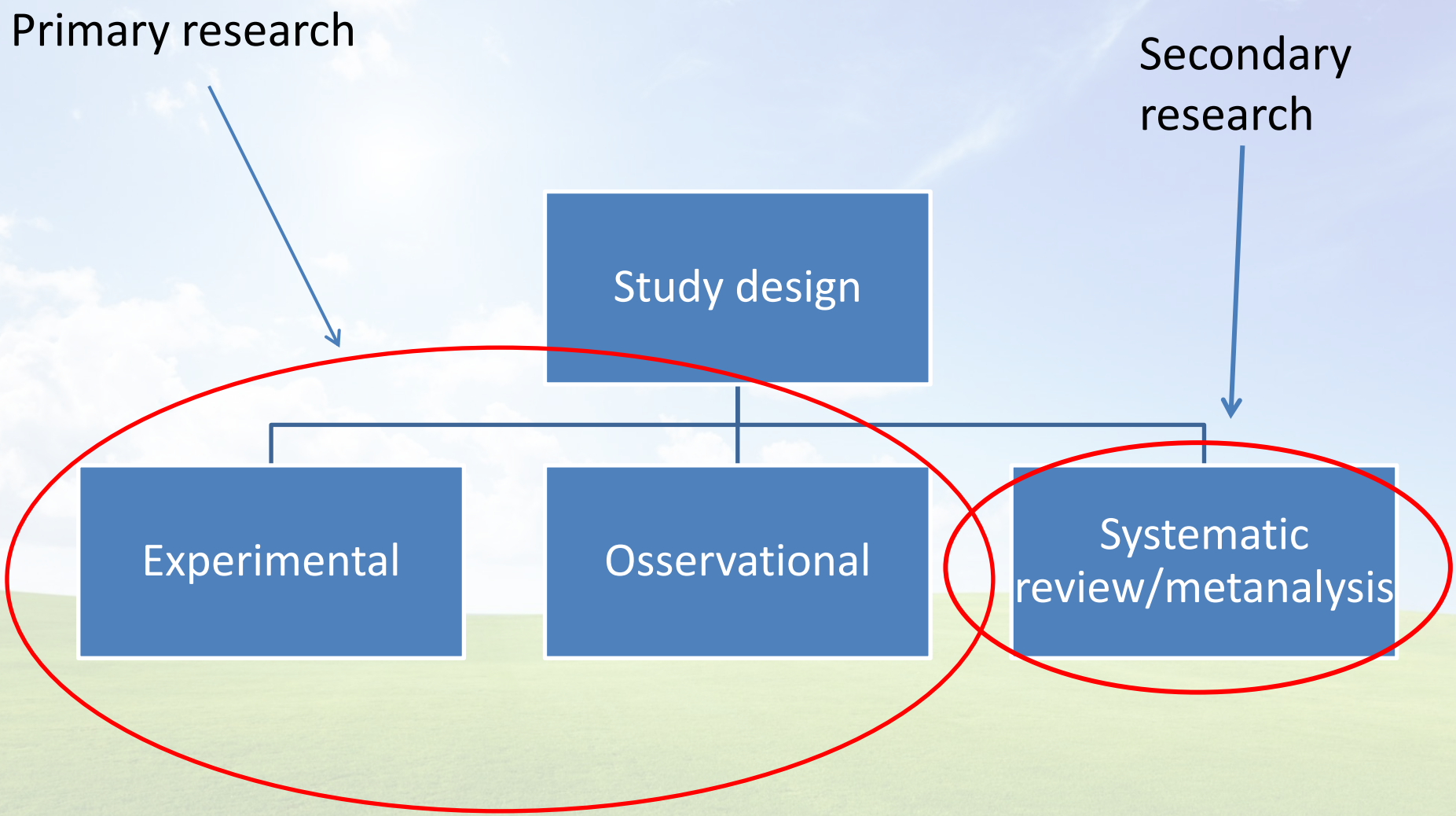
Estimation of prevalence need retrospective study and the up date of vital status is essential

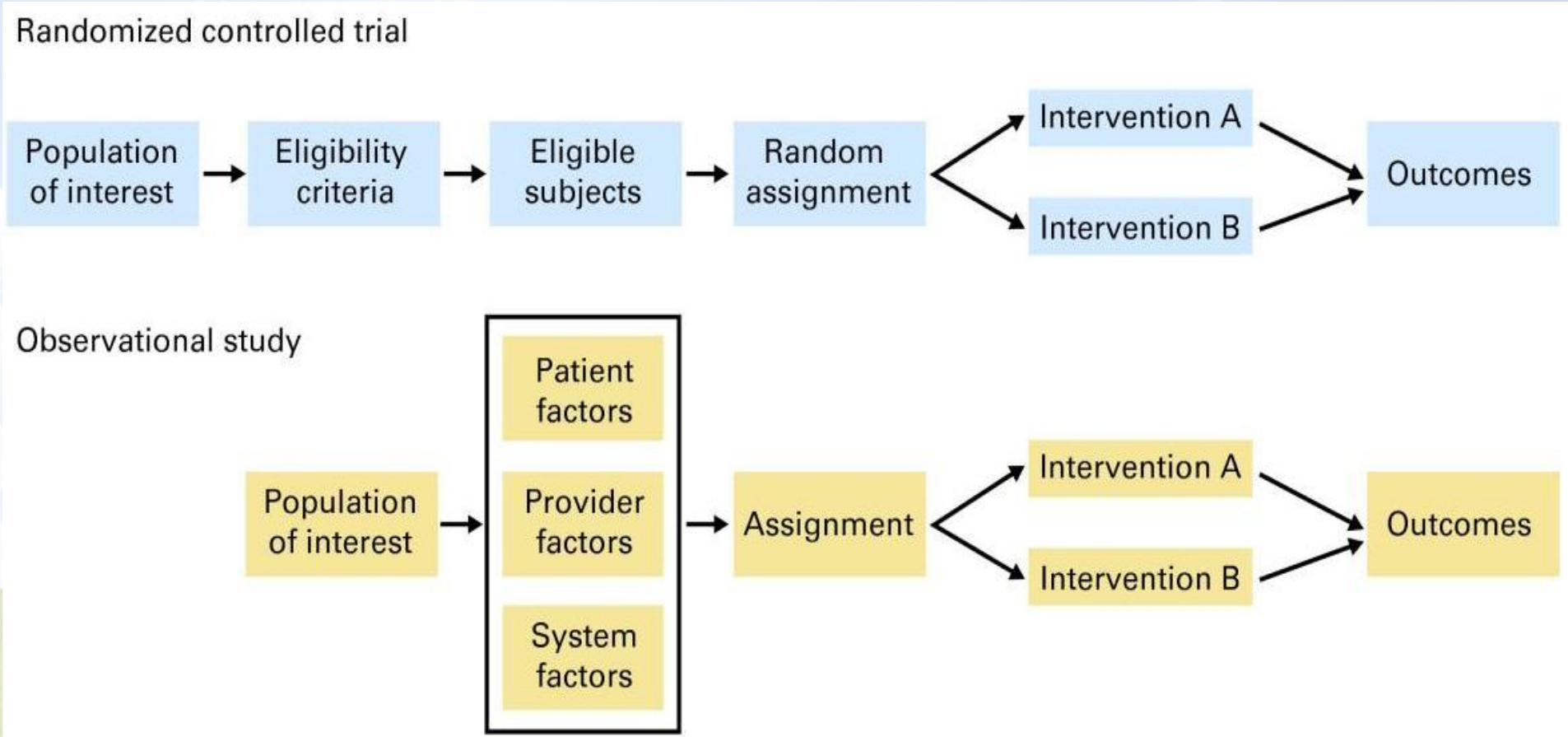
Estimation of incidence need propective study

Epidemiological study design

- which epidemiological study design will be the most appropriate to fulfill the purpose of the study;
- the choice of design for a particular issue will involve trade-offs between speed, cost, quality, and relevance to key policy questions.

The principle is: high quality information at a sufficiently low cost!


Primary research
Secondary research
Study design
Experimental
Osservational
Systematic review/metanalysis

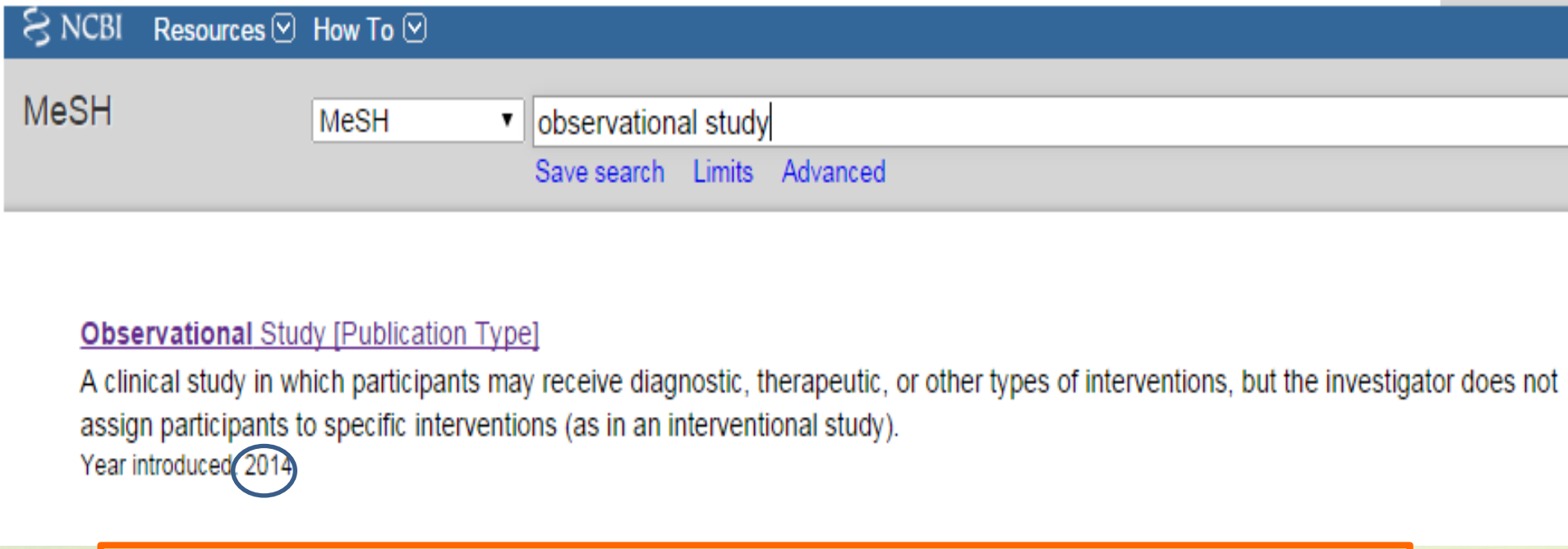

Randomized controlled trial
Population of interest
Eligibility criteria
Eligible subjects
Random assignment
Intervention A
Intervention B
Outcomes
Observational study
Population of interest
Patient factors
Provider factors
System factors
Assignment
Intervention A
Intervention B
Outcomes

A patient registry definition

An organized system using **observational study methods** for the collection, storage, retrieval, and use of a clearly defined set of data on identifiable individuals or a population defined by a particular disease, condition, or **exposure**, to evaluate specified **outcomes** for one or more predetermined scientific, clinical, or policy purpose(s).

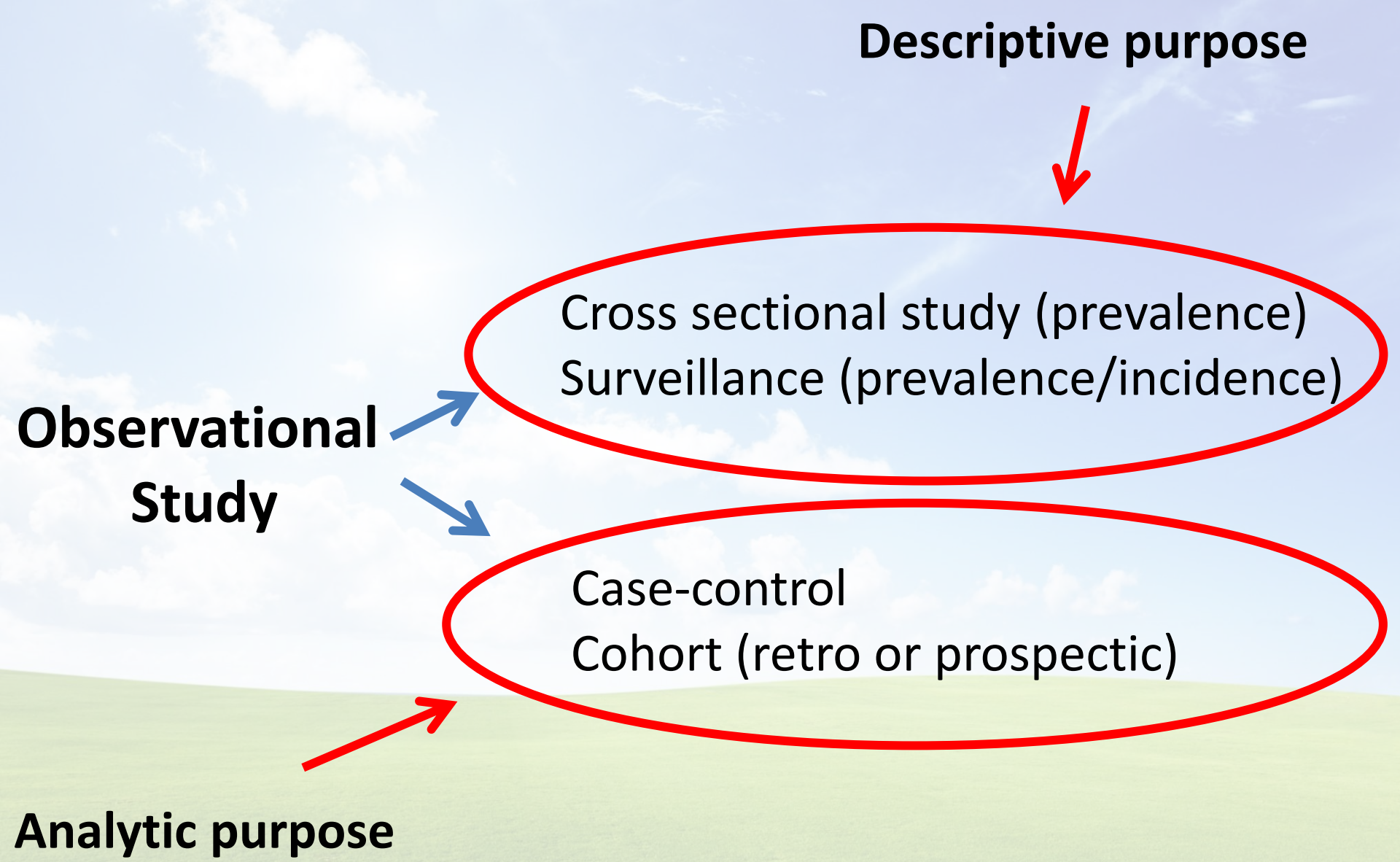
Source: Gliklich, RE and Dreyer, NA., editors. AHRQ. Registries for Evaluating Patient Guide. Agency for Healthcare Research and Quality; Rockville, MD: 2014

What are observational studies?



Observational studies are the only method where:

- a CT may be unethical
- or the condition is rare

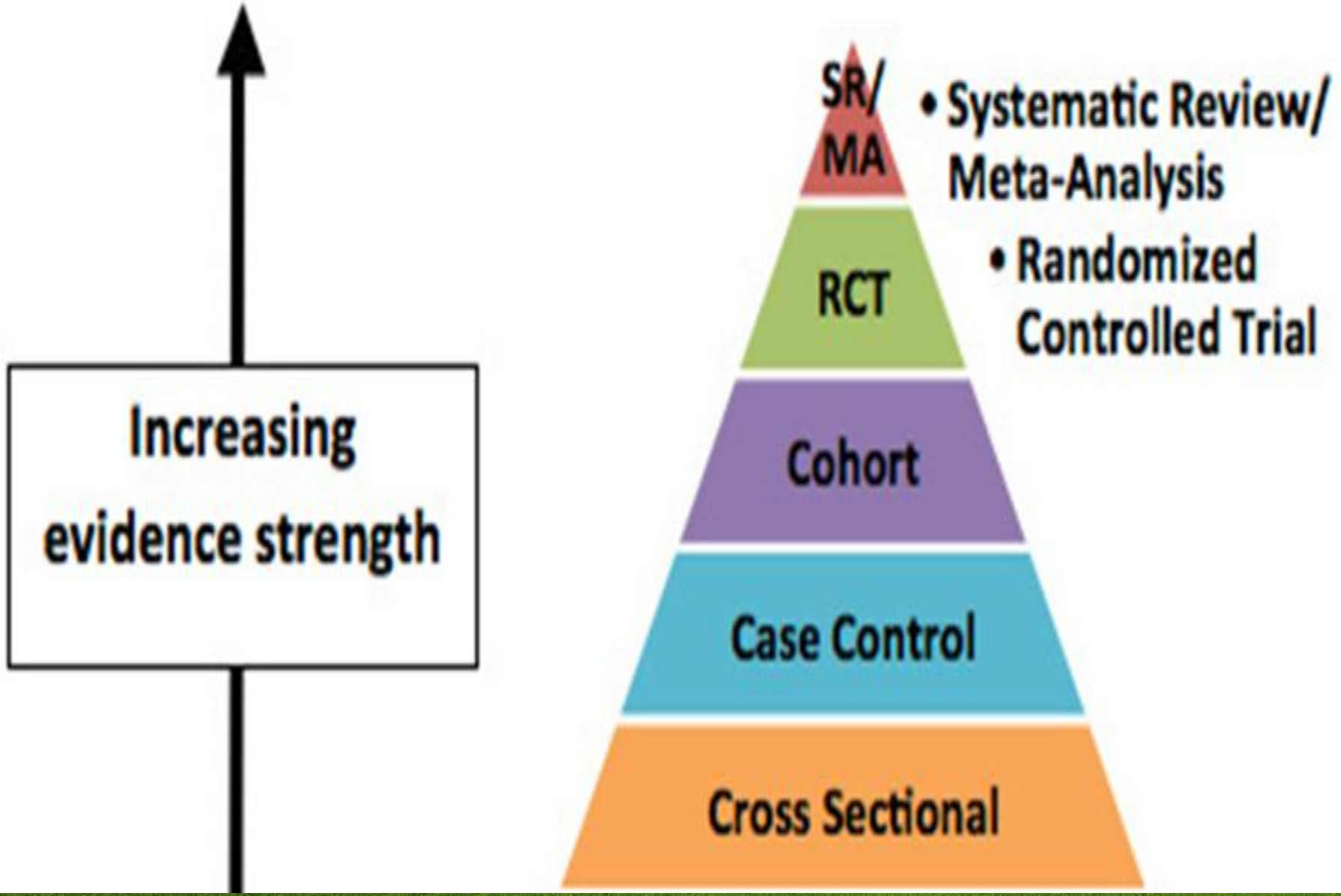

Descriptive purpose
Cross sectional study (prevalence)
Surveillance (prevalence/incidence)
Observational Study
Case-control
Cohort (retro or prospectic)
Analytic purpose

Key points on OS

- **Cross sectional studies** are the best way to determine prevalence
- Are relatively quick
- Can study multiple outcomes
- Do not themselves differentiate between cause and effect or the sequence of events

- **Case-control studies** are simple to organise
- Retrospectively compare two groups
- Aim to identify predictors of an outcome
- Permit assessment of the influence of predictors on outcome via calculation of an odds ratio
- Useful for hypotheses generation
- Can only look at one outcome
- Bias is an major problem

- **In Cohort** study subjects are selected before the outcome of interest is observed
- They establish the sequence of events
- Numerous outcomes can be studied
- They are the best way to establish the incidence of a disease
- They are a good way to determine the causes of diseases
- The principal summary statistic of cohort studies is the relative risk ratio
- If prospective, they are expensive and often take a long time for sufficient outcome events to occur to produce meaningful results


Increasing
evidence strength
SR/
MA
• Systematic Review/
Meta-Analysis
RCT
• Randomized
Controlled Trial
Cohort
Case Control
Cross Sectional

Registries

- Registries are the most difficult of the observational studies to organize as they require the most expensive resources to run effectively.

- A Registry is similar to a cohort study.

  The difference between the two is that, in the registry, a new subject will join the study after start, whereas in a cohort study interest is restricted to those in original group.

Public health surveillance/registry

- descriptive purposes (prevalence/incidence);
- the study population must be representative of the population (high external validity (population-based);
- need a strong collaboration between public with private;
- need the backing of health autorithies and this will happen when the authorities have something to gain from the registry.

Clinical registry

- Analytic purpose (cause, prognosis, treatment efficiency);
- The population coverage is not essential (non population based) and are based on selected bodies, clinical centers. The design is cohort. If the data are available a retrospective cohort design is the quickest method.

Descriptive Studies

- characterize the disease distribution in relation to: person, place, time (who, where, when);
- the first stage of epidemiological study and generate hypotheses;
- useful data for health planning and to evaluate the burden of diseases but not to understand the cause of a disease;
- simple statistical analysis (descriptive analysis);
- does not include a comparison group;
- generate hypotheses;
- non longitudinal data;
- provide rapid information.

Analytic Studies

- understand the causes of a disease, risk factor/s, prognosis or treatment effect, analyse the strength of association between exposure and outcome (why and how);
- the second stage in which hypotheses generated in the descriptive phase are tested;
- useful data to understand the cause/s of a disease but not for health planning and burden of diseases;
- complex statistical analysis (inferential analysis);
- include a comparison group;
- follow-up is essential;
- don't provide rapid information.

It is important to take familiarity with Osservational Study

STROBE

- **STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. 2007 Plos Med 4(10):e296.**

  Aim: provide guidance on how to report OS

ISPOR

- **A Checklist for Retrospective Database Studies—Report of the ISPOR Task Force on Retrospective Databases. Value in Health 6 (2003) 90-7.**

  Aim: evaluate the quality of published studies that use retrospective databases.

ISPOR

- **Prospective Observational Studies to assess Comparative Effectiveness: The ISPOR Good Research Practices Task Force Report. Value in Health 15 (2012) 217-230.**

  Aim: provide guidance on how to design and conduct OS

  ISPOR develop good research practices for Prospective Observational that focus on the effectiveness and/or comparative effectiveness of health care intervention so called **CER (Comparative Effectiveness Research)** study.

Comparative effectiveness research (CER)

- CER is defined as the:

  conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in the “real word setting”

- Interventions may include medications, procedures, devices, technologies, behavioral change strategies.

Comparative effectiveness research

- CER Comparative effectiveness research are defined as the:

  conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnosis, treat and monitor health conditions in "real word setting"

How efficient is a new treatment in practice

Intervention may include medications, procedures, devices, tecnologies, behavioral change strategies

CER and off-label treatments

- For off-label treatment, it exists the phenomenon of “clinical equipoise”;
- Clinical equipoise has been defined by Freedman B (*N Engl J Med 1987*) as the existence of “genuine uncertainty within the expert medical community about the preferred treatment”.
- When clinical equipoise exists, the most appropriate design is not a Clinical trial but a prospective observational study.

Data sources: Where can the necessary data be found?

- Primary data collected by the researcher for the purpose of the study.

- Secondary data (administrative data) are pre-existing databases originally collected for reasons other than this research. These data are used for retrospective study.

ADVANTAGES AND DISADVANTAGES OF USING PRIMARY AND SECONDARY DATA

Primary Data	Secondary Data
Advantages	*Advantages*
Investigator controls all aspects of the study including design, sampling techniques, data collection, and follow-up methods	Relatively fast and inexpensive
All variables of interest can be measured	Sample sizes tend to be large
	Sample may cover a large geographic area and thus provide ability to assess national trends
	Unobtrusive to subjects
Disadvantages	*Disadvantages*
Time consuming	Data may not include all variables of interest
Expensive	Maybe difficult to understand how and why data elements were collected

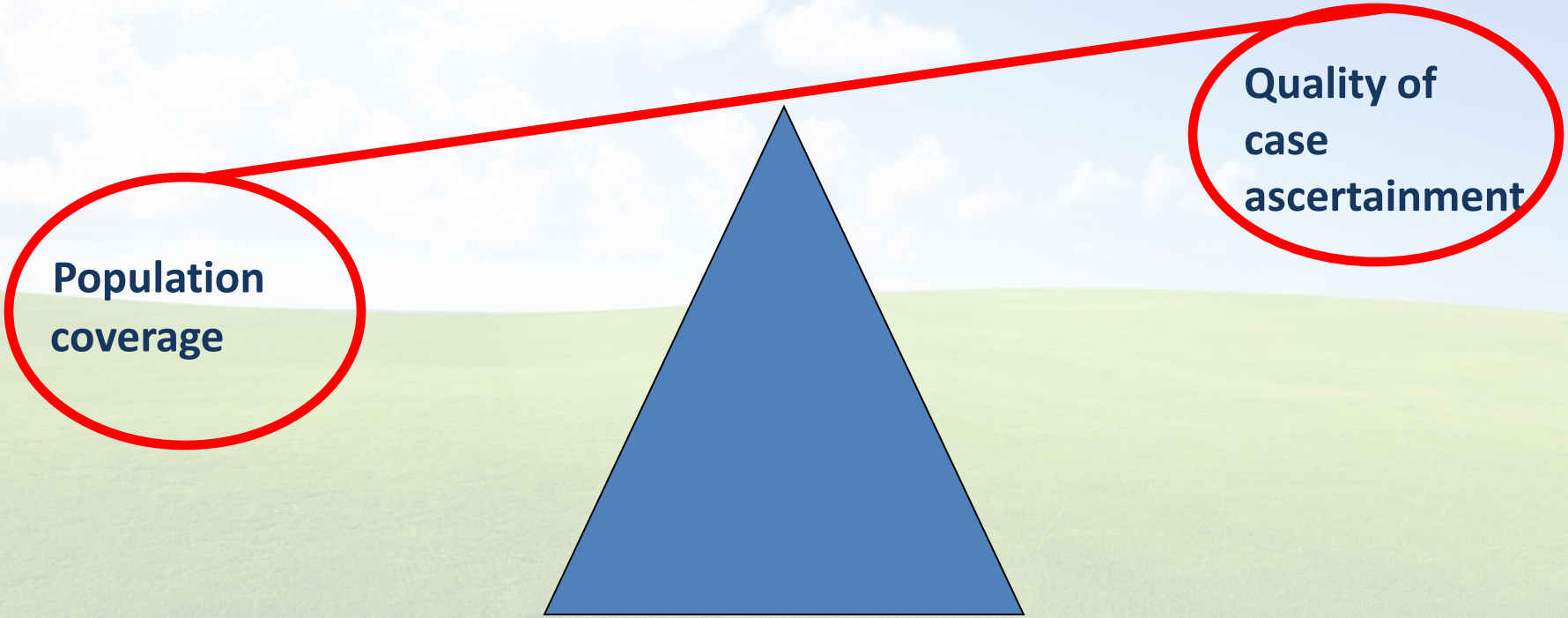
Melissa D.A. Carlson 2009

Example of secondary data source

Demographic and lifestyle data	Census, General Household Survey, social trends, annual reports
Morbidity data	GP morbidity statistics, communicable disease surveillance, hospital inpatient inquiry, hospital activity analysis, cancer registration
Mortality data	OPCS mortality statistics
Health services data	Immunisation data, screening levels, district and annual reports, confidential inquiry reports.
Specific databases	OPCS survey of disability, OPCS longitudinal study, abortion

Quality and quantity of data sources

- **a** rare condition could be identified from a large population;
- the best quality case ascertainment for rare events is possible when we focus on a limited population.



Variables (1)

- The gold rule is to keep the data sheet as short and as simple as possible.
- The content of the form (i.e., the data collected) is of course, driven by the goals:
  - public health surveillance needs relatively limited information;
  - clinical registry based on prospective cohort observational studies need more variables on risk factors and outcomes.
- Data collection is a tedious and time-consuming process, so it is important to limit metrics that are of secondary importance.
- List variables in exposure and outcome.
- Variable to uniquely identify patients (prevent duplicate records).

Patient-reported variables (2)

- Variables include patient-reported (data source patients)
- Example of patient-reported data (PROs Patient reported outcome) include health related quality of life; symptoms; complementary, and alternative medications; behavioral data (e.g., smoking and alcohol use); family history; and biological specimens.
- These data may rely on the subjective interpretation (recall bias)

Clinical variable (3)

- Variables include clinical data (data source: clinicians medical records).
- Examples of clinician data include clinical impressions, clinical diagnoses, clinical signs, differential diagnoses, laboratory results, and staging.
- The primary advantage of clinicians data is that clinicians are trained observers. Even so, the primary disadvantages are that clinicians are not necessarily accurate reporters of patient perceptions, and their responses may also be subject to recall bias. Also, the time that busy clinicians can devote to registry data collection is often limited.

Study protocol (1)

One of the most important components of research design is the creation of a study protocol, which is the researchers' blueprint to guide and govern all aspects of how a study will be conducted. A study protocol directs the execution of a study to help ensure the validity of the final study results. It also provides transparency as to how the research is conducted and improves the reproducibility and replicability of the research by others, thereby potentially increasing the credibility and validity of a study's findings.

Protocol outline (2)

1. **Purpose**: What is the key health policy question that the study is designed to answer?
2. **Background**: What is the current state of knowledge?
3. **Hypotheses**: What is the primary hypothesis? What are the secondary hypotheses (if any)?
4. **Study Design:**
   a. Study design and rationale
   b. Definition of population (patients, providers, sites) that will be studied (target of inference)
   c. Definition of treatment cohorts to be compared
   d. Definition of outcome measures to assess treatment effects
   e. Definition and justification of control outcome (if any)
5. **Data Analysis Plan**:
   a. Sample size justification
   b. Primary planned analyses, secondary planned analyses
   c. Definition of relative effectiveness measure or causal effect (average causal effect, local causal effect)
   d. Planned approaches to deal with bias, confounding, missing data, and multiple outcomes(if secondary outcomes)
   e. List confounders (whether collected or not)

Protocol outline (2)

6. **Study Governance and Implementation:**
   a. Governance and sponsorship
   b. Incentives for participation (if any)
   c. Safety reporting
   d. Informed consent and IRB approval (if required)
   e. Data processing and quality control
   f. Sample listing of data elements
   g. Plan for dissemination of results and publication planning
   h. If the study is designed to support a policy decision, explanation of decision to register study or not
   i. Anticipated timing of dissemination and publication of study results

Running successfully the registry will require expertise in Epidemiology and Biostatistics.