Research Methods

Kath O'Reilly
Associate Professor
Faculty of Epidemiology and Population Health

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

kathleen.oreilly@lshtm.ac.uk

Thank you to Lena Lorenz
for developing this session

Outline of this Session

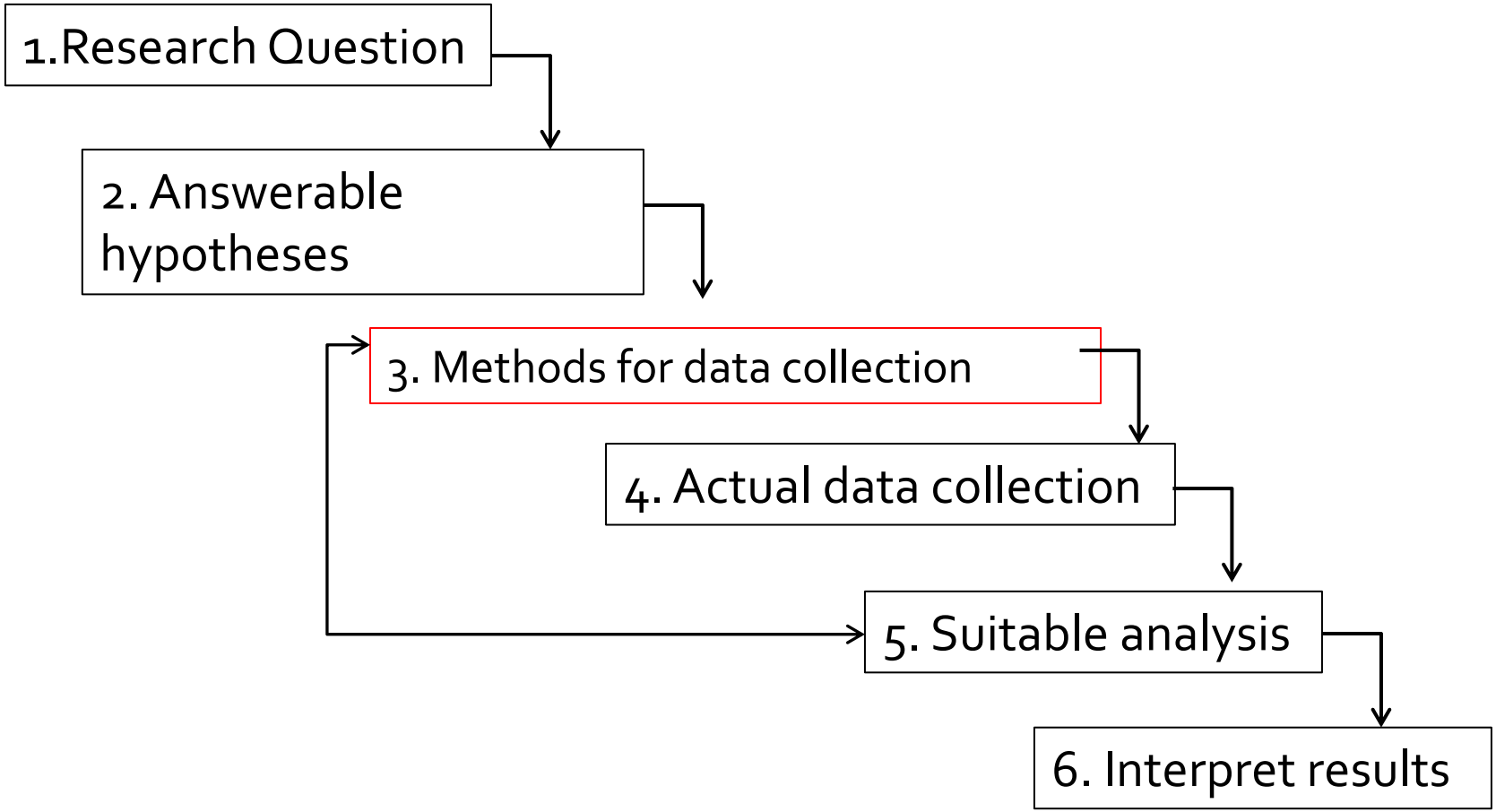
- **TODAY**
- **9:00-10:00** Essential 'Principles' of Research Methods
- **10:30-12:30** Applying principles in practice – Rapid Projects

- **14:00-15:30** Project development
- **15:30-16:30** Presentations and feedback

We have LG09 for this Session

Slides are here https://tinyurl.com/2p9easuz (GitHub page)

The "Big" Picture



(1/6) Research Questions

- Familiarity with the subject
- Establish a single primary research question at the beginning of a study
- Basis of the study plan
- Clear and concise
- Contain the topic being studied (purpose), the variable(s), and the population

Three main types of questions

- Descriptive – Describes an occurrence: "how much", "how often", or "what is the change".

e.g. How many tsetse flies land on cloths of different colours in rural Cameroon?

- Comparative – Examines the difference between two or more groups in relation to one or more variables.

e.g. What is the difference between the number of tsetse flies landing on white and black cloths in rural Cameroon?

- Relational – Compares two or more variables and determines if a relationship exists.

e.g. Does the cloth colour of traps affect the number of tsetse flies caught in rural Cameroon?

e.g. http://dissertation.laerd.com/how-to-structure-quantitative-research-questions.php

Translating Thoughts to Research Questions

Example: "Do mosquitoes bite more humans or cattle?"

- Revised: "Do female *Anopheles arabiensis* mosquitoes preferentially blood-feed on humans or cattle in Tanzania?"

Example: "How did the malaria burden in Nigeria change between 2010 and 2015?"

- Revised: "Did national malaria control interventions cause a change in human malaria incidence and prevalence in Nigeria between 2010 and 2015?"

Brainstorming

Good practices of formulating research questions

- Basic question to answer
- Achievable & answerable in frame of research
- Know your methods and analysis before formulating question
- Avoid repetition of info already out there
- Where, when and why?
- Focus on one issue only
- Clear, precise, specific; should not leave room for ambiguity
- Narrow scope (but not too narrow; it shouldn't be answerable with a single statistic)
- "So what?" – has impact

P&P: Take a pen and paper…

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

3 Word Project

P&P: Research Question

Start translating your "3 words" into a Research Question…

Consider:

- Basic question to answer
- Achievable & answerable in frame of research
- Know your methods and analysis before formulating question
- Avoid repetition of info already out there
- Where, when and why?
- Focus on one issue only
- Clear, precise, specific; should not leave room for ambiguity
- Narrow scope (but not too narrow; it shouldn't be answerable with a single statistic)
- "So what?" – has impact

2. Hypothesis Testing

Hypotheses

- Refinement of the research question into a testable prediction
- Hypotheses are not always appropriate – e.g. when studies are more descriptive or explorative with no clear expectations or defined variables
- Clear statement of what is intended to be investigated – formalise relationship between variables
- Must be specified before research is conducted

Cont.

Example: "Tsetse flies are attracted to a certain colour"

- Revised: "Tsetse flies are more likely to land and rest on black cloth than on cloth of other colours"

Example: "Rain affects malaria burden"

- Revised: "Higher rainfall increases the number of people testing positive for malaria with RDTs and microscopy"

Feedback & brainstorming

Good practices of formulating hypotheses

- It's a statement - it's not a question
- Define variables – how they are quantified
- Mirror the research question
- Neither too specific nor too general
- It's a prediction of consequences
- Considered valuable even if proven false

Hypothesis testing

Hypothesis (or significance) testing

- *The purpose of hypothesis testing is to make an inference about the population of interest on the basis of a random sample taken from that population.*

Defining the Null Hypothesis

Null hypothesis (H_o) = There is no effect in the population

Example: "There is no effect of cloth colour on tsetse fly landing."

Alternative hypothesis (H_A) = H_A formulated more specifically according to study objectives/theory

"Cloth colour has an effect on tsetse fly landing."

The final conclusion is always given in terms of the null hypothesis:

- e.g. 'reject H_o in favour of H_A' or 'do not reject H_o'

5 steps in hypothesis testing

1. Define the *null* and *alternative* hypotheses of the study;
2. Collect data from a sample of individuals;
3. Calculate the value of the *test statistic* specific to the null hypothesis and the data;
4. Compare the value of the test statistic to values from a known probability distribution;
5. Interpret the *P*-value and results.

'Errors' in Hypothesis Testing

Type I Error: We reject the null hypothesis when it is true. Denoted by α (significance level of the test). Reject H_0 if $P < \alpha$.

For example: *A type I error would occur if we concluded that the two drugs produced different effects when in fact there was no difference between them.*

Type II Error: We do not reject the null hypothesis when it is false. Denoted by β. (1- β) is the power of the test.

For example: *A type II error would occur if it were concluded that the two drugs produced the same effect when in fact they produced different ones.*

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Decision

Reality		**Reject H_0**	**Don't reject H_0**
	H_0 true	Type I Error	No error
	H_A true	No error	Type II Error

- Important to have studies that are adequately powered
- Power = Probability of detecting a real treatment effect given it exists
- Choice of α level and power based on a consideration of the relative costs of type I and type II errors.
- Choosing the correct **sample size** is important

P&P: Generate a Hypothesis Test

Consider how you would translate your Research Question into a testable hypothesis

3. Methods for Data Collection

Defining your variables

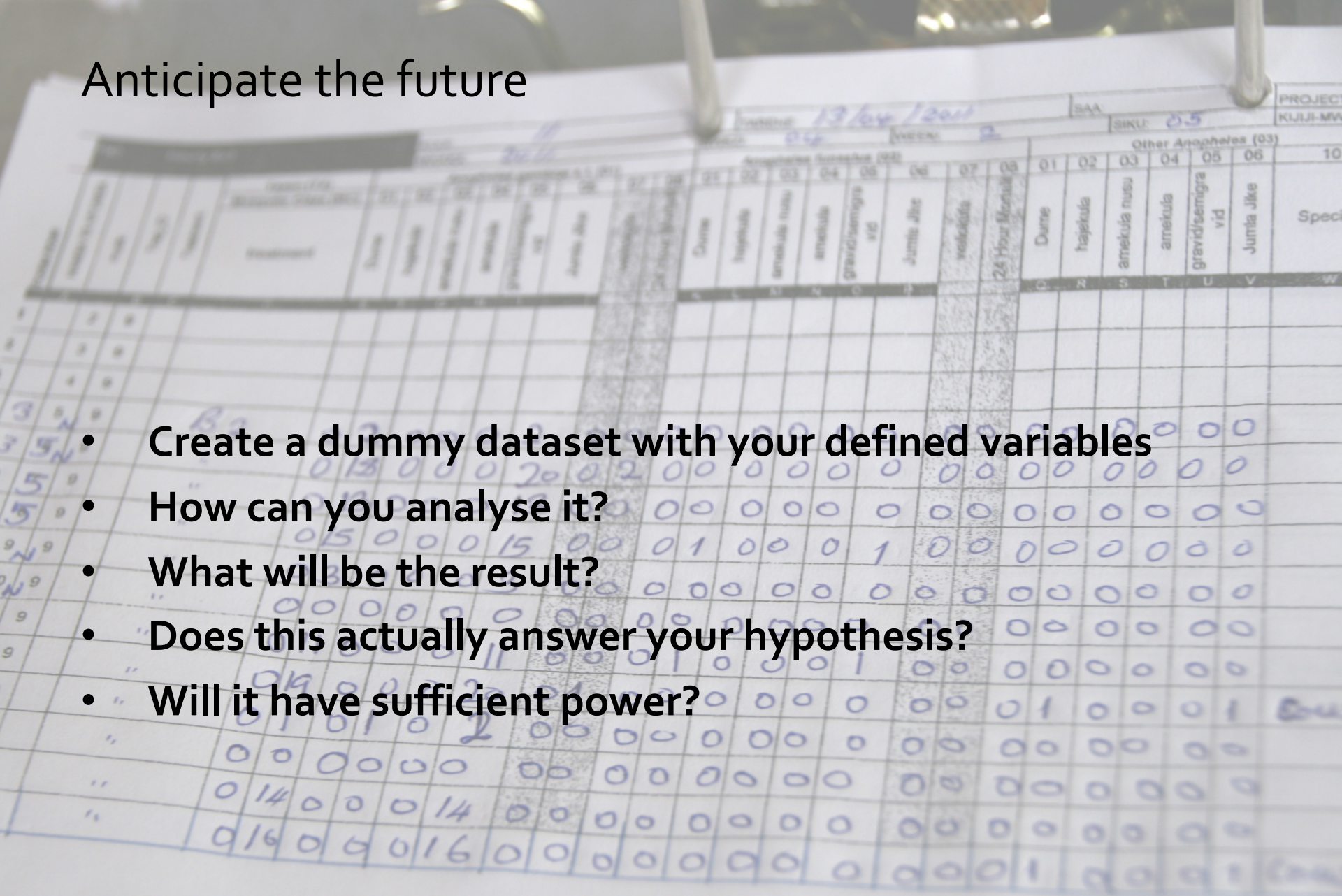
- What is your outcome (dependent) variable?
- What is/are your exposure (independent; explanatory) variable/s?
- What is your unit of observation? Determines scale of replication.
  – Mosquitoes, traps, houses, villages…
- What is your study population?
  – How generalisable will your data be?

Exposure variables

- Danger of including a large number of explanatory variables
- Common sense: Biological or clinical reasons to suspect relationship
- Univariable and multivariable analyses
- Interactions between explanatory variables (e.g. factorial experiments)
- Collinearity: 2 explanatory variables are highly correlated → difficult to evaluate their individual effects in multivariable models
- Confounding: 2 explanatory variables are both related to the outcome and to each other → failure to adjust may lead to biased estimates of model parameters

→ Think carefully which factors to measure and include in study design, analysis and how to interpret results

Anticipate the future



- **Create a dummy dataset with your defined variables**
- **How can you analyse it?**
- **What will be the result?**
- **Does this actually answer your hypothesis?**
- **Will it have sufficient power?**

Experimental design: Causality

We are often most interested in causality, but many experimental designs only illustrate association…

Does exposure to a factor cause the outcome?

- Ecological studies (more later) will illustrate association
- Laboratory experiments and RCTs are gold standard (more later)

Some rules for determining causality (Hill 1965):

- Cause must precede effect
- Association should be biologically sensible
- Consistent results from a number of studies
- Strong association between cause and effect
- Dose-response relationship with effect
- Removing factor of interest should reduce risk of disease

Beware of spurious correlations…

Experimental design: Controls

What would have happened in the absence of intervention?

Observational: compare units that happened to be "treated" with those that were not treated

Experiment: compare units assigned to "control" with "treatment" conditions

Negative control: No effect when there should be no effect

Positive control: A known outcome is expected (gold standard)

Experimental design: Replication

Everything varies!

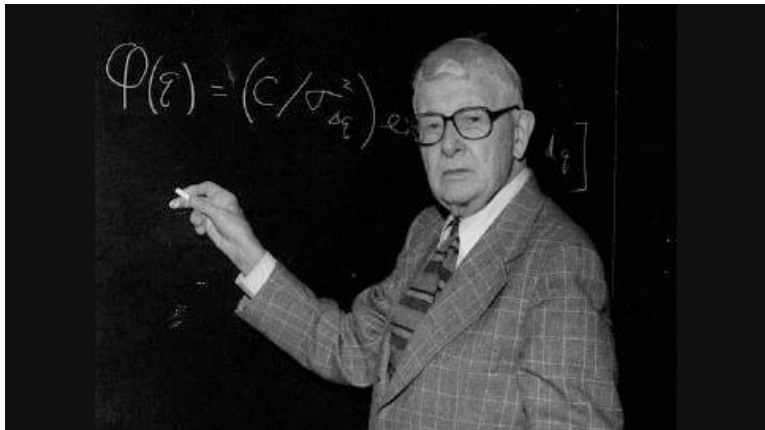
- Variation can affect your outcome variable, and design needs to account for natural variability
- Replication increases your power to detect important differences
- Not enough replication increases uncertainty around your estimates

Experimental design: Sample size

- We need to expand upon the 'how many?' question
- Balance between study precision/power and logistical feasibility
- Based on random variation in population (i.e. use of standard deviation)
- Specify H_A or *effect size* to be detected
- Usually focus on primary outcome

- Sample size calculator:

  EpiTools



Experimental Design: Bias

- Systematic difference between results from a study and reality
- Bias can crop up everywhere: study design, analysis, publication…
- Overestimate correlations/effect sizes; hide real associations
- Selection/sampling bias – study units not representative of study population
- Information bias – systematic errors in measurements of exposure or outcomes
  - E.g. recall bias, measurement bias, application bias

Randomisation

Random sampling: All subjects are equally likely to be selected. Avoids sampling bias.

Random allocation: Every subject has an equal opportunity to be allocated to the treatment/control groups.

Blinding

Study participants, researchers (double blind), statisticians…

do not know "control" vs "treatment" allocation

P&P: Tangible elements of the study

Consider the details you have written so far…

- What is the outcome you will measure?
- What other aspects will you need to measure?
- How long will the study need to happen for?
- How much data needs to be collected? When will 'enough' be collected?

3. Methods for Data Collection – study types

Cross-sectional studies

- Describe frequency and characteristics of an outcome at single point in time
- Measure disease prevalence but not incidence – cannot observe trends
- Relationship between exposure and outcome non-directional
- Cheap and quick; easy to standardise
- Useful as preliminary studies to formulate hypotheses for intervention studies/trials
- Survey or census, e.g. Malaria Indicator Surveys

Case-Control studies

- Retrospective: Compare a group of individuals with outcome of interest (=cases) with a group that do not have it (=controls)
- Compare exposures to (risk) factors in the past to explain outcome of interest, measured by odds of outcome (calculate odds ratios; ORs)
- Selection bias – have precise and unambiguous eligibility criteria
- Matching of cases and controls
- Recall bias
- Cheap, quick, easy

Ecological/correlation studies

- Associations between outcome and exposures at population level
  - Examine exposure-disease relationship between groups at one time point
  - Examine exposure-disease relationships over time in one group (time trends)
- Useful to explore hypotheses; quick and cheap
- Overestimate degree of association between variables
- Does not prove cause-effect relationships

Cohort studies

- Prospective: Follow group of individuals over time to investigate how exposures affect outcome of interest
- Outcome = estimate of incidence of disease or change in variable over time measured by risk of disease, which allows calculation of relative risk (RR)
- Cohort should be representative of relevant population
- Expensive; time consuming (long-term); not suitable for rare outcomes but good for rare exposures
- Risk of loss to follow up (take into account when calculating sample sizes)
- Low risk of recall or selection bias

Laboratory experiments

- Animal studies; molecular studies; model systems
- Manipulate variables of interest
- Controlled conditions, exclude as much unexplained variation as possible
  - Environmental conditions
  - Physiological status of subjects
  - Genetic factors of subjects
- Look out effect of exposures in isolation – reductionist?
- Simplify the natural world – generalisable?

Factorial experiments

Simultaneous analysis of any number of factors of interest

2x2 factorial experiment: 2 factors each at 2 levels

	Untreated Net	ITN
No spraying		
IRS		

Test whether 2 factors are interactive, i.e. whether the effect of net treatment is different whether insecticide has been sprayed or not

Randomised Control Trials

- Exposure randomly allocated to subjects
- Measurements of outcomes: cross-sectional or longitudinal
- Avoid allocation bias
- Provide strong evidence of cause-effect relationships
- Expensive and difficult to design; ethical issues
- External validity – are results valid outside of experimental setting?

Smoking and Cancer

In the 1940s the UK had the highest lung cancer rates, and incidence exceeded tuberculosis for the first time…

Motor cars? Smoking?

1948

- Richard Doll and Austin Bradford Hill began a case-control study of cancer in London hospitals and interviewed patients about daily habits

1949

- Richard Doll stopped smoking

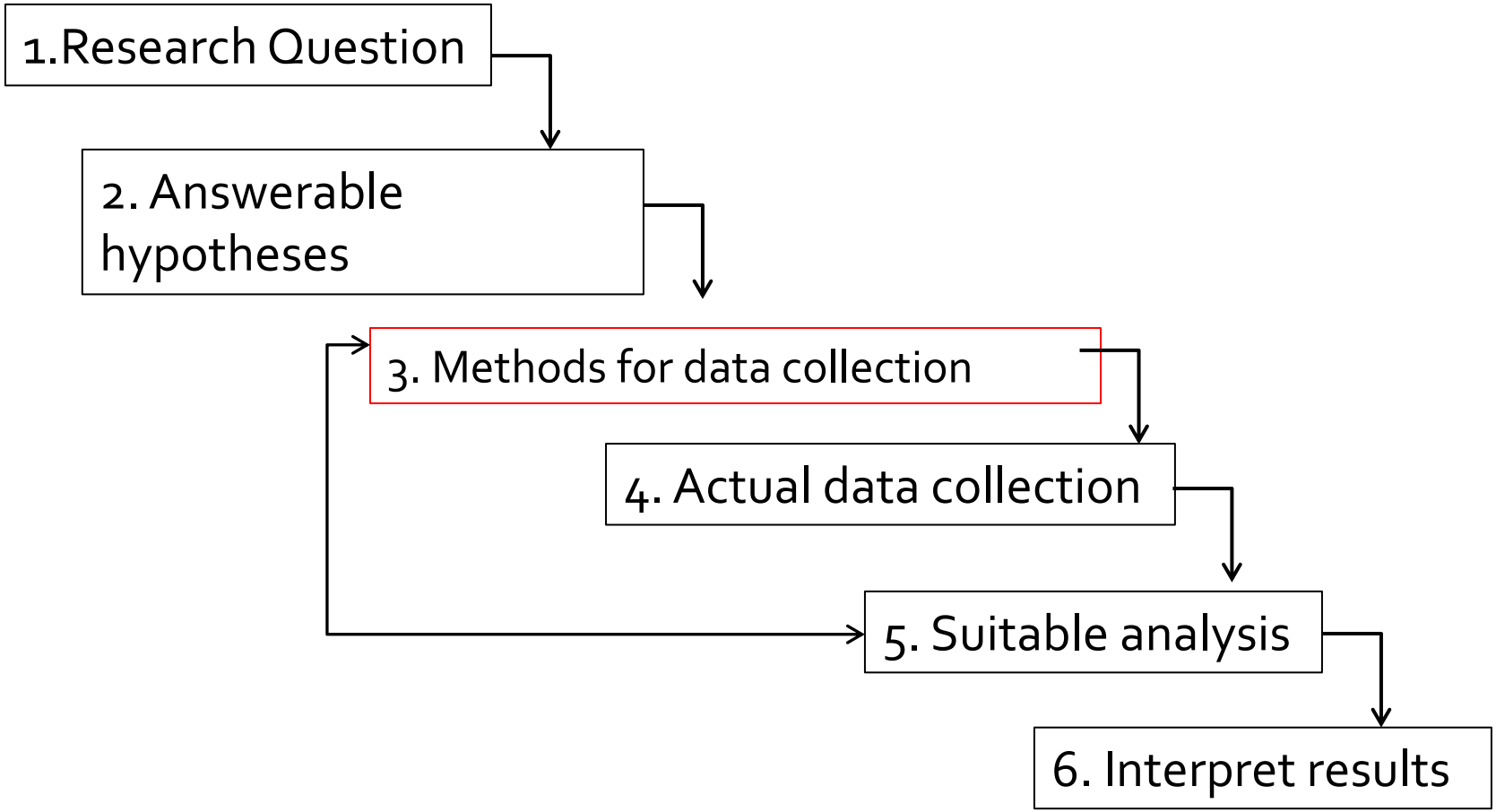
1950

- "Smoking and Carcinoma of the Lung" BMJ – smoking assoc. x16 risk

1954

- Doll and Hill: "The Mortality of Doctors in Relation to Their Smoking Habits" BMJ
- A cohort study of doctors where they were followed until death
- A dose response relationship with lung cancer and increased risk of heart disease

The "Big" Picture



P&P: Study design

Does your study fit a specific study design?

Rapid Project

- Reconvene at 10:30
- Come back with:
  - A research Q you could do over the summer *(…think feasible!)*
  - The hypothesis
  - The (rough) methods
- Online discussion of ideas…we will then coalesce towards 4 ideas and form groups to develop further
- Plan your Research
- Present your Research and give feedback to other groups

Recommended reading

A Short Introduction to Epidemiology by Neil Pearce (LSHTM), online eg here

Medical statistics 2nd edition by Betty R Kirkwood & Jonathan A.C. Sterne (library and online)

Medical Statistics at a Glance 3rd edition by Aviva Petrie & Caroline Sabin (library)

Great for (simple) sample size calculations:
EpiTools (http://epitools.ausvet.com.au)