Digital health and computational epidemiology

Lesson 4

Michele Tizzoni
Dipartimento di Sociologia e Ricerca Sociale
Via Verdi 26, Trento
Ufficio 6, 3 piano

UNIVERSITÀ DI TRENTO

FONDAZIONE BRUNO KESSLER

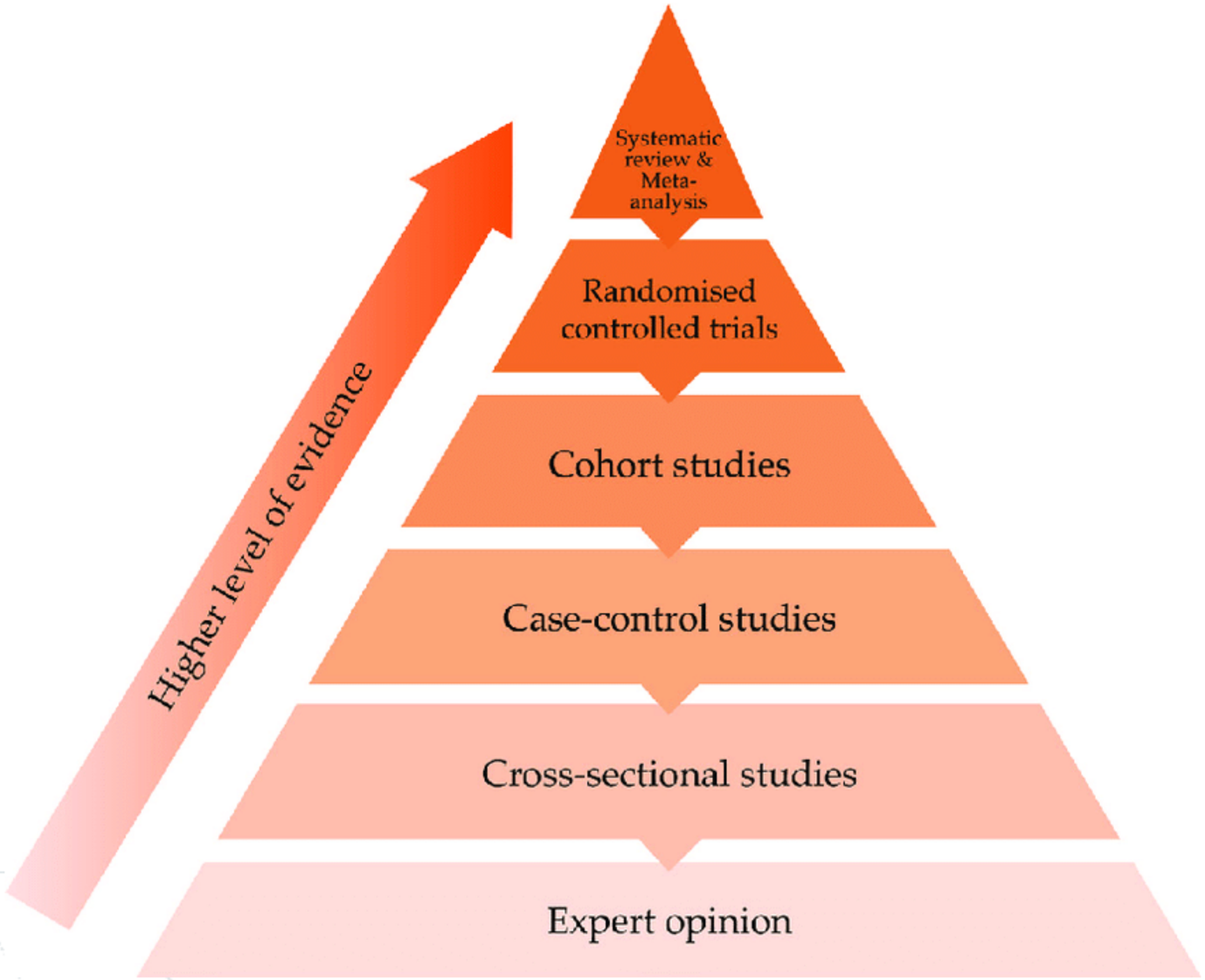
C2S2 Center for Computational Social Science and Human Dynamics

Epidemiological studies

Epidemiological studies

- How do we know what we know in epidemiology? **What constitutes sufficiently strong evidence?**
- Medical practice has mostly followed standard-of-care practices or the beliefs of medical experts.
- The term “evidence-based medicine” is barely 30 years old. It was first used in 1998 in a published manuscript.
- In the past decades, a generic **hierarchy of evidence** has emerged (first published in 1995).

Evidence ranking

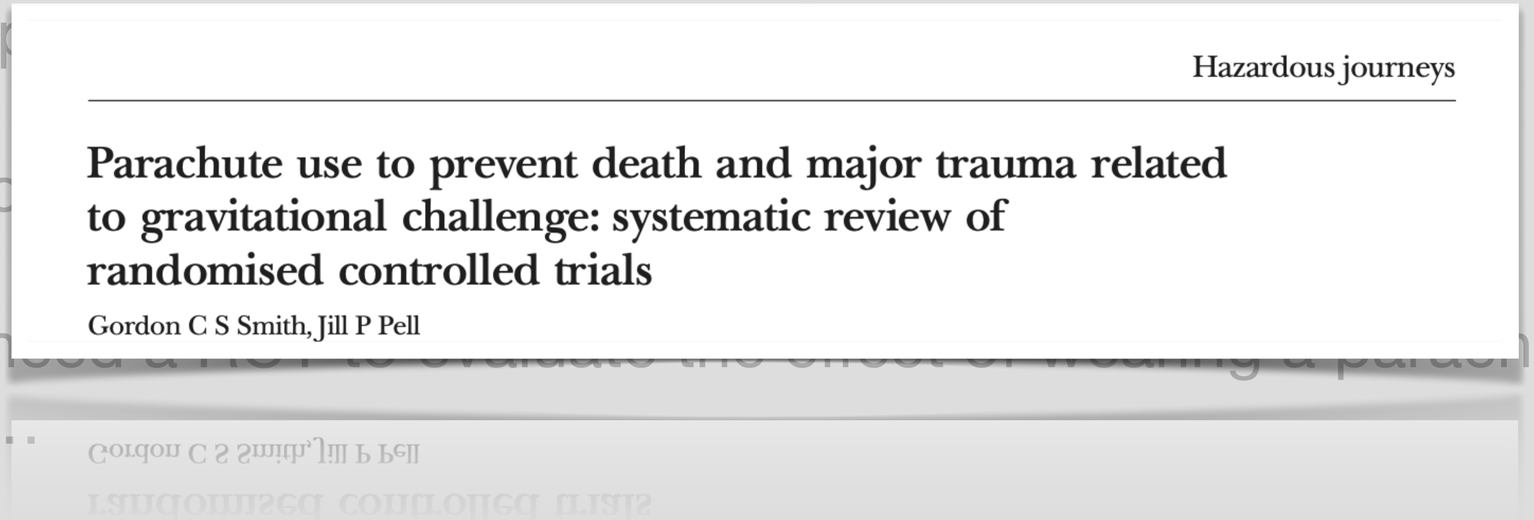


Evidence ranking

- Be aware: the pyramid can be misleading! It does not mean that RCTs are the best option available to gain evidence *in all circumstances*.
- RCTs are often not possible for several reasons.
- We don’t need a RCT to evaluate the effect of wearing a parachute when skydiving…
- Expert opinion could be valuable in many cases, especially at the beginning of an outbreak of an unknown pathogen.

Evidence ranking

- Be aware: the pyramid can be misleading! It does not mean that RCTs are the best op
- RCTs are c
- We don't n ute when skydiving…
- Expert opinion could be valuable in many cases, especially at the beginning of an outbreak of an unknown pathogen.



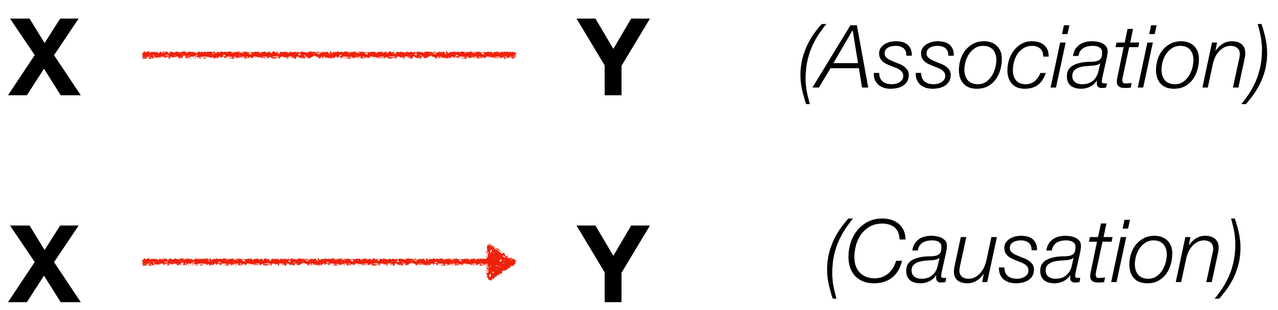
Hazardous journeys

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell

Exposure and outcome

- There is a certain factor or condition that people are exposed to (**the exposure**) which may or may not lead to a certain **outcome**.
- Assessing the link between exposure and outcome is what epidemiological studies are all about.



Exposure and outcome

- It would easy to think that exposure is something external, like a virus or an environmental factor, like heat, while outcome means health or disease.
- **This is not always the case.**
- **Exposure** can really mean anything, such as age, location, disease, behavior, etc. - any factor that may be associated with a given outcome of interest.
- The **outcome** can be behavior, such as the decision to vaccinate, or to adopt preventive measures.
- Studies will be called differently depending on whether people are selected based on exposure or on outcome.

Observational studies

Observational studies are studies where we observe a group, and in particular the effect of a factor (exposure) on an outcome of interest, of that group.
The key aspect is that we do not control this factor.

Case reports

- The first level of evidence is that of **case reports** and **case series**
- A case report is a report on a particular patient, with a certain outcome such as disease
- Case series is a group of similar case reports.
- Case reports are not designed to test a hypothesis but they can be a trigger to formulate one.
- Case reports are often instrumental in the early description of newly emerging diseases.

Case reports

Letter | Published: 23 April 2020

Clinical and virologic characteristics of the first 12 patients with coronavirus disease 2019 (COVID-19) in the United States

The COVID-19 Investigation Team

Nature Medicine **26**, 861–868 (2020) | Cite this article

CDC Home | Search | Health Topics A-Z

MMWR Weekly

June 5, 1981 / 30(21);1-3

Epidemiologic Notes and Reports

Pneumocystis Pneumonia --- Los Angeles

In the period October 1980-May 1981, 5 young men, all active homosexuals, were treated for biopsy-confirmed *Pneumocystis carinii* pneumonia at 3 different hospitals in Los Angeles, California. Two of the patients died. All 5 patients had laboratory-confirmed previous or current cytomegalovirus (CMV) infection and candidal mucosal infection. Case reports of these patients follow.

Ecological studies

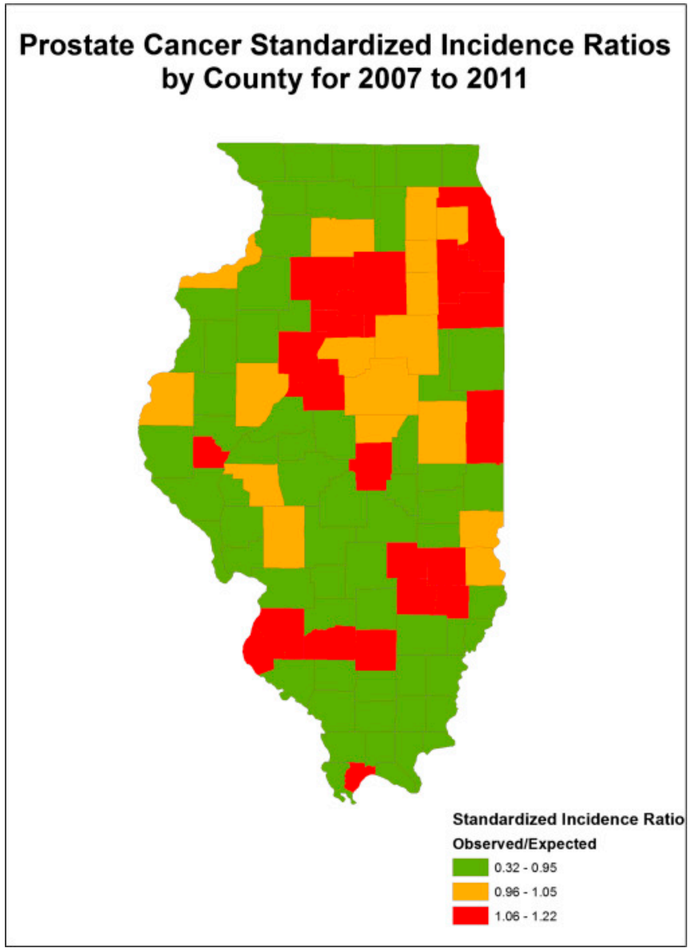
- An **ecological** - or aggregate - study compares quantities of groups, rather than quantities of individuals.
- A case series may have indicated that most people with a given disease of interest had a certain type of exposure.
- We could then design an ecological study, comparing different groups - for example, different countries, or different towns - by looking at their incidence of the disease, and their aggregate level of exposure.

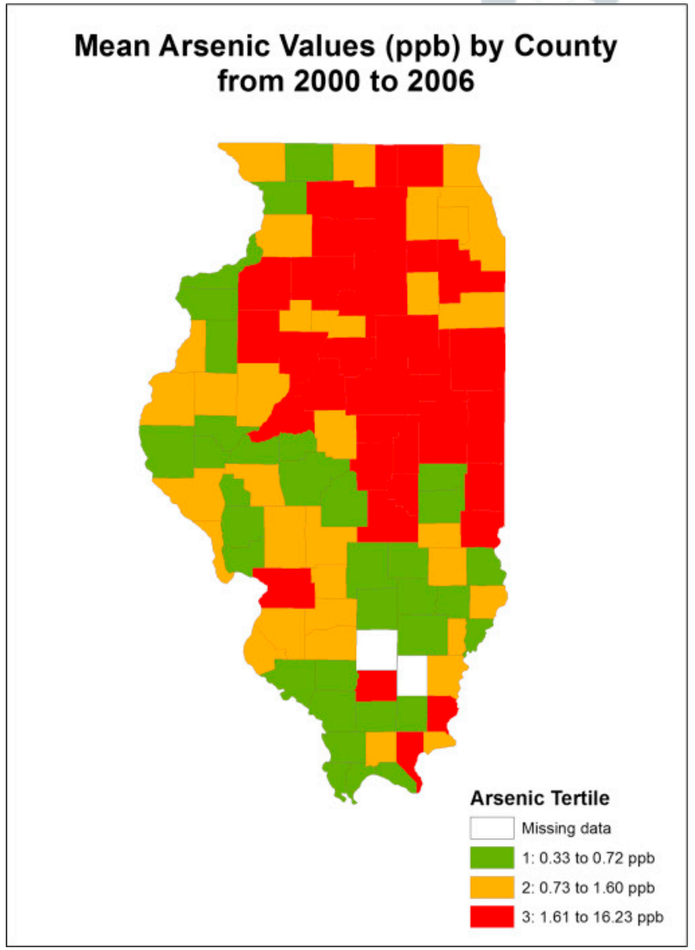
Ecological studies

Environmental Research
Volume 148, July 2016, Pages 450-456

Arsenic in drinking water and prostate cancer in Illinois counties: An ecologic study

Catherine M. Bulka [a], Rachael M. Jones [b], Mary E. Turyk [a], Leslie T. Stayner [a], Maria Argos [a]

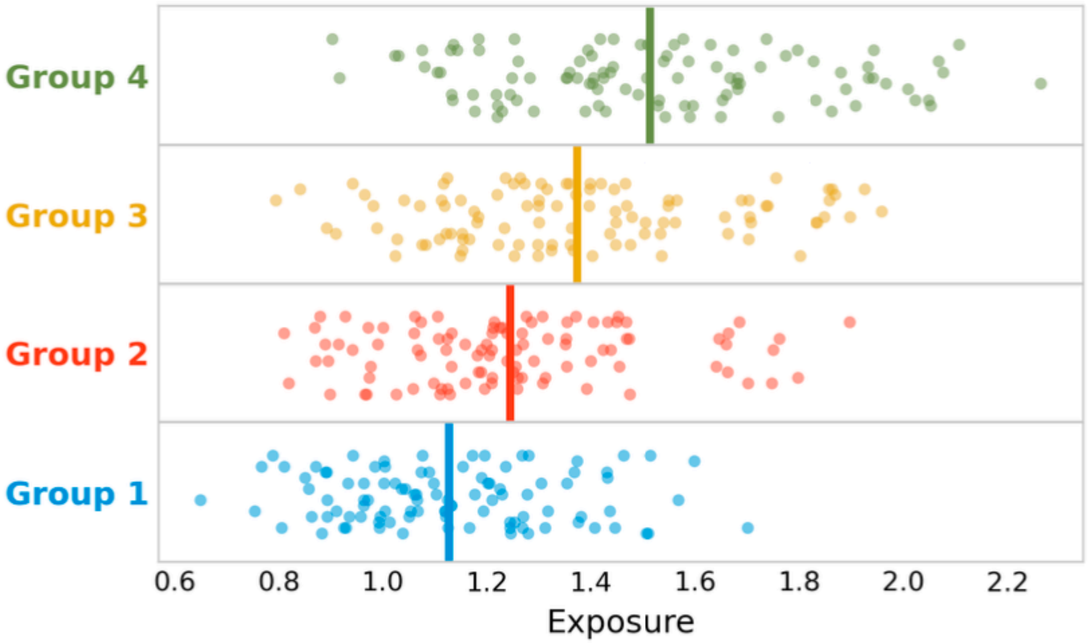




Ecological studies

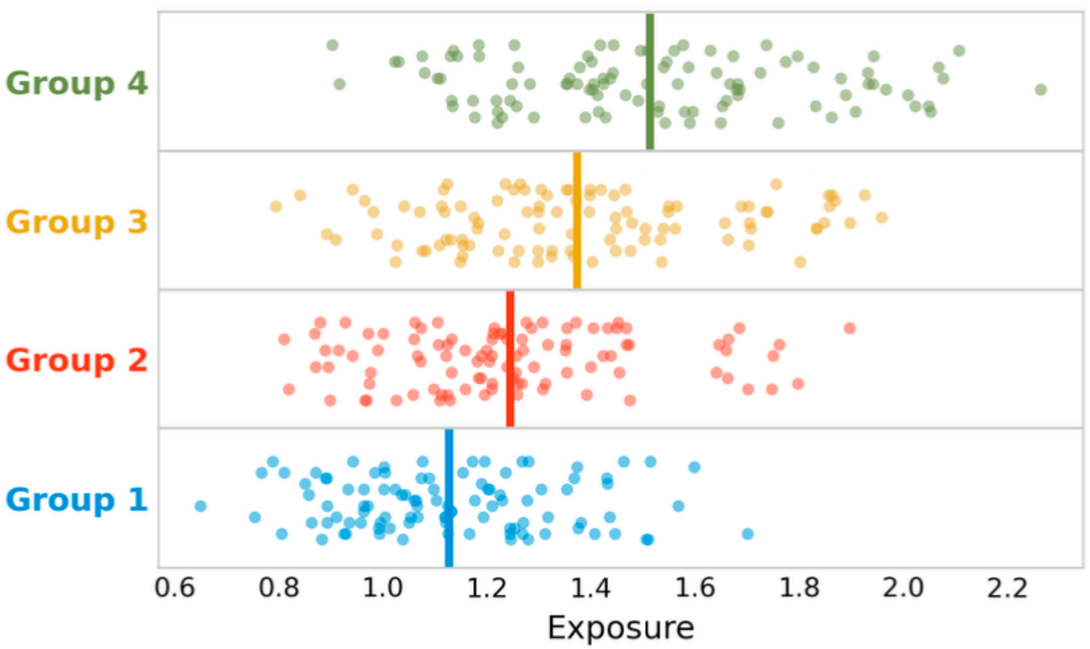
- This type of studies doesn’t tell us anything about the individual-level risks.
- Misinterpretation of the results based on the group means is called **ecological fallacy.**
- **The ecological fallacy happens when conclusions about an individual are drawn from based solely on the general tendency observed in a group.**

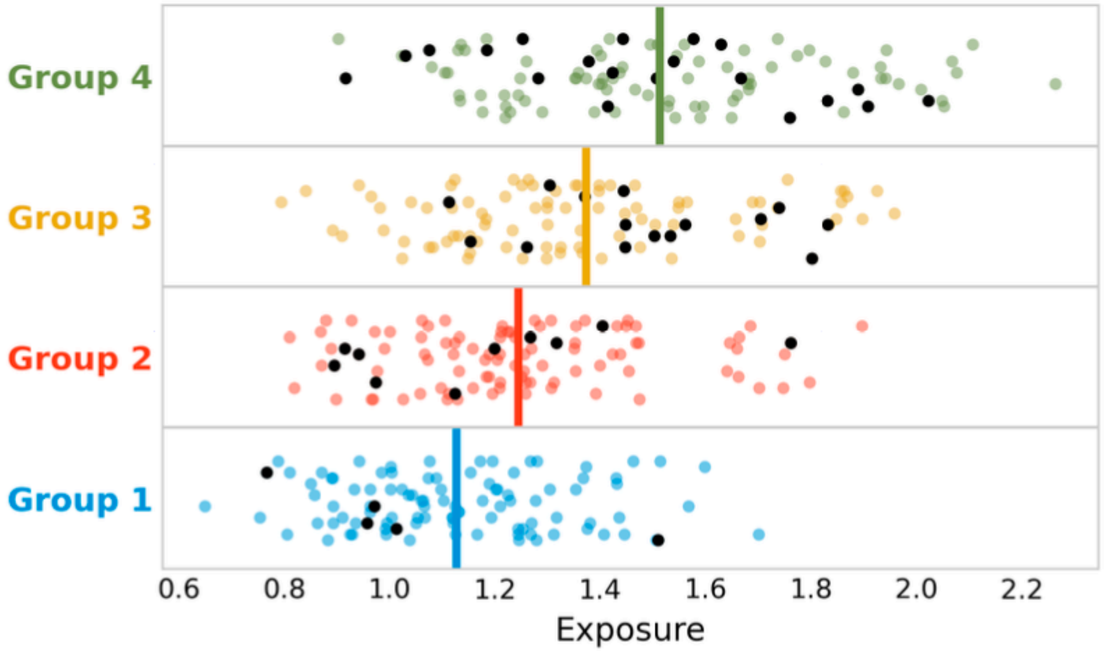
Ecological fallacy



- Let’s imagine we compare the average exposure with disease incidence in 4 groups.
- We find that the average exposure is positively correlated with incidence at the level of groups.

Ecological fallacy





Black dots indicate the diseased individuals.
They are randomly distributed
with respect to the level of exposure.

No correlation at the individual level!

Ecological fallacy

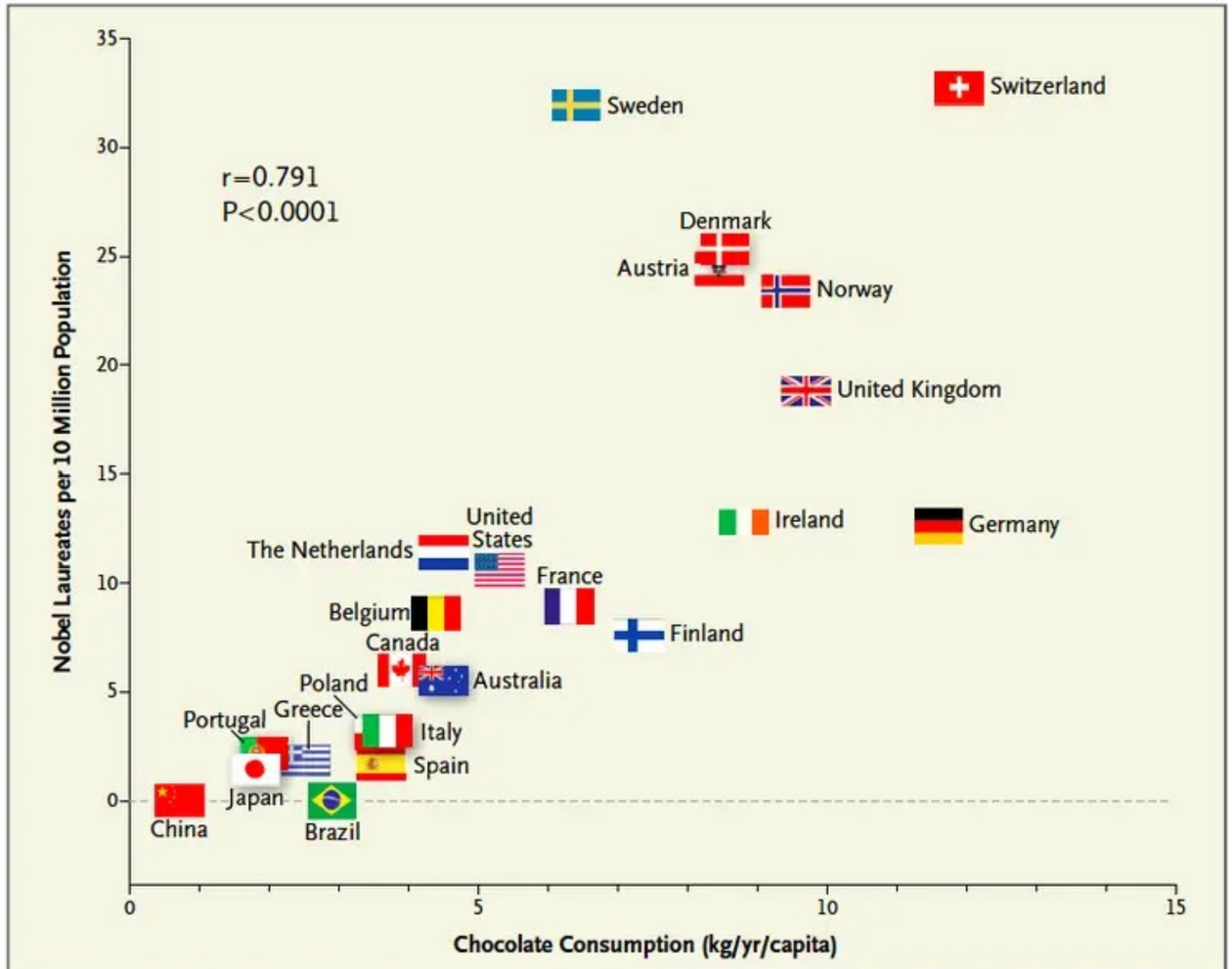


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

OCCASIONAL NOTES FREE PREVIEW

Chocolate Consumption, Cognitive Function, and Nobel Laureates

Franz H. Messerli, M.D.

Cross-sectional studies

- **Cross-sectional studies** are studies where data is collected at a **given point in time**.
- Because cross-sectional studies can mostly assess prevalence, they are often called **prevalence studies**.
- The opposite of a cross-sectional study is a **longitudinal study**, where data are collected at **multiple time points**.
- Longitudinal studies are adequate to assess the incidence of a disease.

Cross-sectional studies

The internet and children's psychological wellbeing

Emily McDool[a], Philip Powell[a,b], Jennifer Roberts[a], Karl Taylor[a,c,*]

[a] Department of Economics, University of Sheffield, UK
[b] School of Health and Related Research, University of Sheffield, UK
[c] IZA Bonn, UK

ARTICLE INFO

Article history:
Received 12 December 2018
Received in revised form 31 July 2019
Accepted 7 December 2019
Available online 13 December 2019

JEL classification:
D60
I31
J13

Keywords:
Digital society
Social media
Wellbeing
Children
Happiness

ABSTRACT

Late childhood and adolescence is a critical time for social and emotional development. Over the past two decades, this life stage has been hugely affected by the almost universal adoption of the internet as a source of information, communication, and entertainment. We use a large representative sample of over 6300 children in England over the period 2012–2017, to estimate the effect of neighbourhood broadband speed, as a proxy for internet use, on a number of wellbeing outcomes, which reflect how these children feel about different aspects of their life. We find that internet use is negatively associated with wellbeing across a number of domains. The strongest effect is for how children feel about their appearance, and the effects are worse for girls than boys. We test a number of potential causal mechanisms, and find support both for the 'crowding out' hypothesis, whereby internet use reduces the time spent on other beneficial activities, and for the adverse effect of social media use. Our evidence adds weight to the already strident calls for interventions that can reduce the adverse effects of internet use on children's emotional health.

Case-control studies

- **Case-control studies** are studies that look at groups that differ in the outcome.
- Oftentimes, case-control studies compare a group with a disease to a group that does not have the disease (but the outcome can be anything).
- **A good case-control study** design compares two groups that are identical except for the outcome of interest.
- It then looks at the levels of exposure in the case group, and in the control group.
- In order to quantify the association between exposure and outcome, we can calculate an **odds ratio**.

Odds ratios

	Exposed	Not exposed
Disease	80	15
No disease	244	1018

$$\textbf{Odds ratio} = \frac{\textbf{Odds event in the exposed}}{\textbf{Odds event in the non-exposed}}$$

$$\textbf{Odds disease in the exposed} = \frac{80}{244} = 0.328$$

$$\textbf{Odds disease in the non-exposed} = \frac{15}{1018} = 0.0147$$

$$\textbf{Odds ratio} = 22.25$$

Case-control studies

- **Case-control studies cannot establish causality, not matter how high is the OR**
- Study participants are always drawn from a sample, thus we need to report the odds ratios with a **confidence interval (CI)**.
- **If the CI includes 1, then the association between exposure and outcome is not statistically significant**.
- In 2007, a case-control study of Human Papilloma Virus (HPV) and oropharyngeal cancer (a cancer in the back of the throat) reported an odds ratio of 32.2 (95% CI, 14.6 to 71.3) for oropharyngeal cancer and seropositivity for the HPV-16 [1].
- In 2016, a case-control study of Guillain-Barré Syndrome, an autoimmune disorder leading to muscle weakness, and the Zika virus infection reported an odds ratio of 59.7 (95% CI, 10.4 to ∞) for Guillain-Barré syndrome and Zika virus positivity [2].

Case-control studies

Advantages:

- We can start with the outcome and look back.
- We can study rare outcomes, by simply starting with all known cases.

Biases:

- **Selection bias.** Cases and controls are selected based on a factor related to exposure. Cases are not representative of all cases, controls are not representative of the larger population.
- **Recall bias**. Cases are more likely to recall and report exposure than the control individuals.

Cohort studies

- **Cohort studies** are studies that start with **different levels of exposure**, and try to see if the outcome of interest is associated with the levels of exposure.
- The goal is the same as in a case-control study, the difference is that the cohort starts by comparing different levels of exposure, while the case-control study starts by comparing different levels of outcome.
- A case-control study, which starts with different groups of outcome, is **necessarily a retrospective study**, i.e. going back in time.
- Cohorts start with different groups of exposure. This could be now, and we can plan the cohort to observe the outcomes in the future. This is called a **prospective cohort**.
- However, we can also find different groups of exposure in the past, and see how the outcome developed. This would be called a **retrospective cohort.**

Relative risk

	Exposed	Not exposed
Disease	80	15
No disease	244	1018

$$\textbf{Relative risk} = \frac{\textbf{Risk when exposed}}{\textbf{Baseline risk}} = \frac{P(D \mid E)}{P(D \mid \neg E)}$$

$$P(D \mid E) = \frac{80}{80 + 244} = 0.247 \qquad P(D \mid \neg E) = \frac{15}{15 + 1018} = 0.0145$$

RR $\simeq 17$

Relative risk

	Exposed	Not exposed
Disease	80	15
No disease	244	1018

$$\textbf{Relative risk} = \frac{\textbf{Risk when exposed}}{\textbf{Baseline risk}} = \frac{P(D \mid E)}{P(D \mid \neg E)}$$

$$P(D \mid E) = \frac{80}{80 + 244} = 0.247 \qquad P(D \mid \neg E) = \frac{15}{15 + 1018} = 0.0145$$

RR $\simeq 17$

$$\textbf{Odds ratio} = \frac{P(D \mid E)}{1 - P(D \mid E)} \Big/ \frac{P(D \mid \neg E)}{1 - P(D \mid \neg E)} = 22.25$$

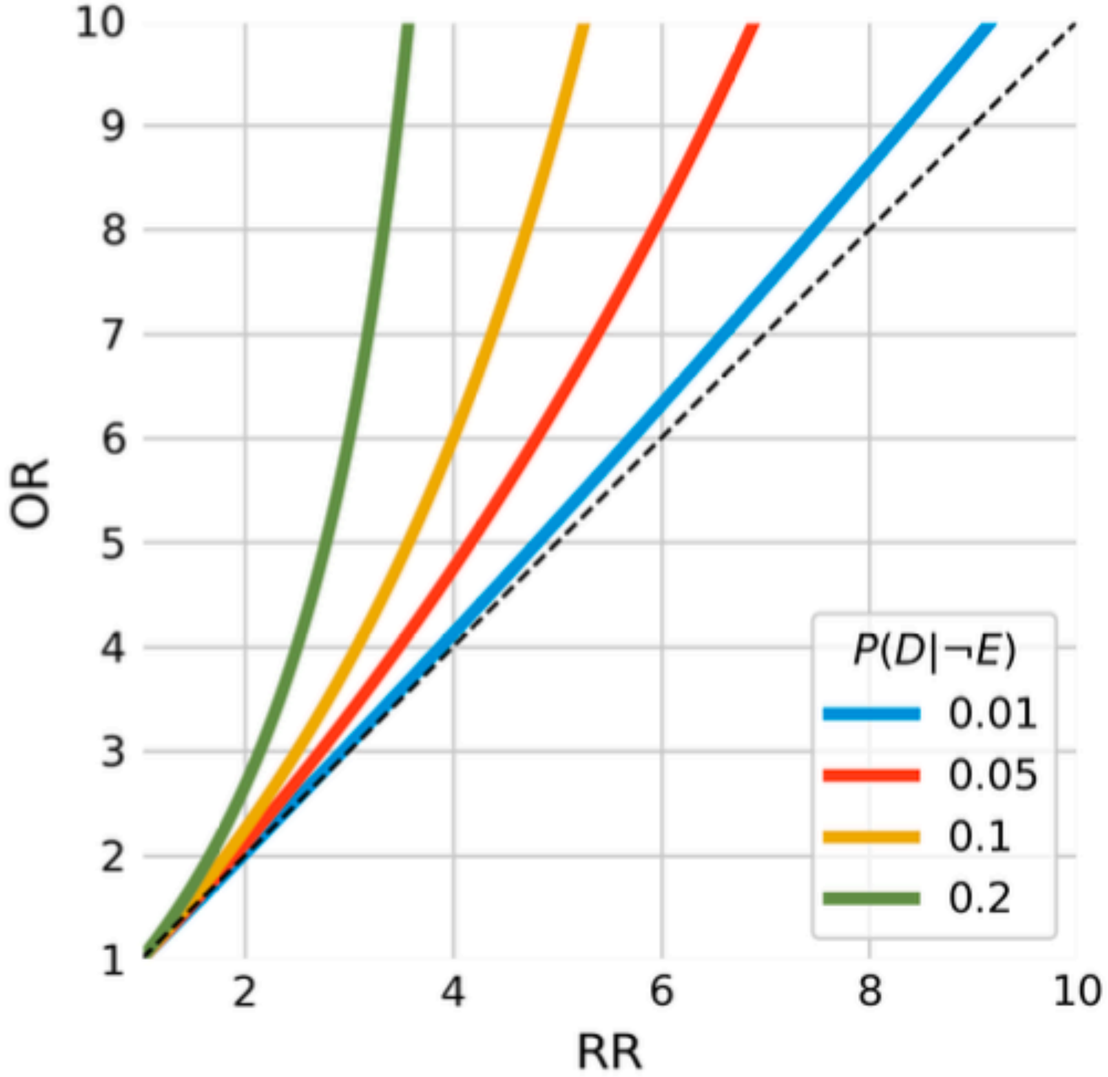
Relative risk

- Given a disease vs exposure contingency table, we can calculate both odds ratio and relative risk.
- For cohorts, that is fine. But **for case-control studies, calculating a risk ratio makes no sense.**
- **When we start from the exposure**, as we do in **cohorts**, it makes sense to talk about risk, or probability, of the outcome.
- **When we start from the outcome**, as we do in **case-control studies**, it makes no sense to talk about the risk of outcome: indeed, we started at the outcome. That's why we use odds ratios.

Relative risk

$$\textbf{Odds ratio} = \frac{P(D\,|\,E)}{P(D\,|\,\neg E)} \times \frac{1 - P(D\,|\,\neg E)}{P(D\,|\,\neg E)}$$

$$\textbf{Odds ratio} = \textbf{Relative Risk} \times \frac{1 - P(D\,|\,\neg E)}{P(D\,|\,\neg E)}$$



Cohort studies

- The biggest advantage of cohorts is that cohort study designers can specifically determine what data should be gathered, and how.
- The downside is that cohorts are difficult for diseases that take a long time to develop, or are rare (or both).
- Well designed, long-term cohorts are expensive.
- The **Framingham heart study** is a cohort that started in Framingham, Massachusetts (US) in 1948 in order to better understand the development and risk factors of cardiovascular diseases.

The spread of obesity

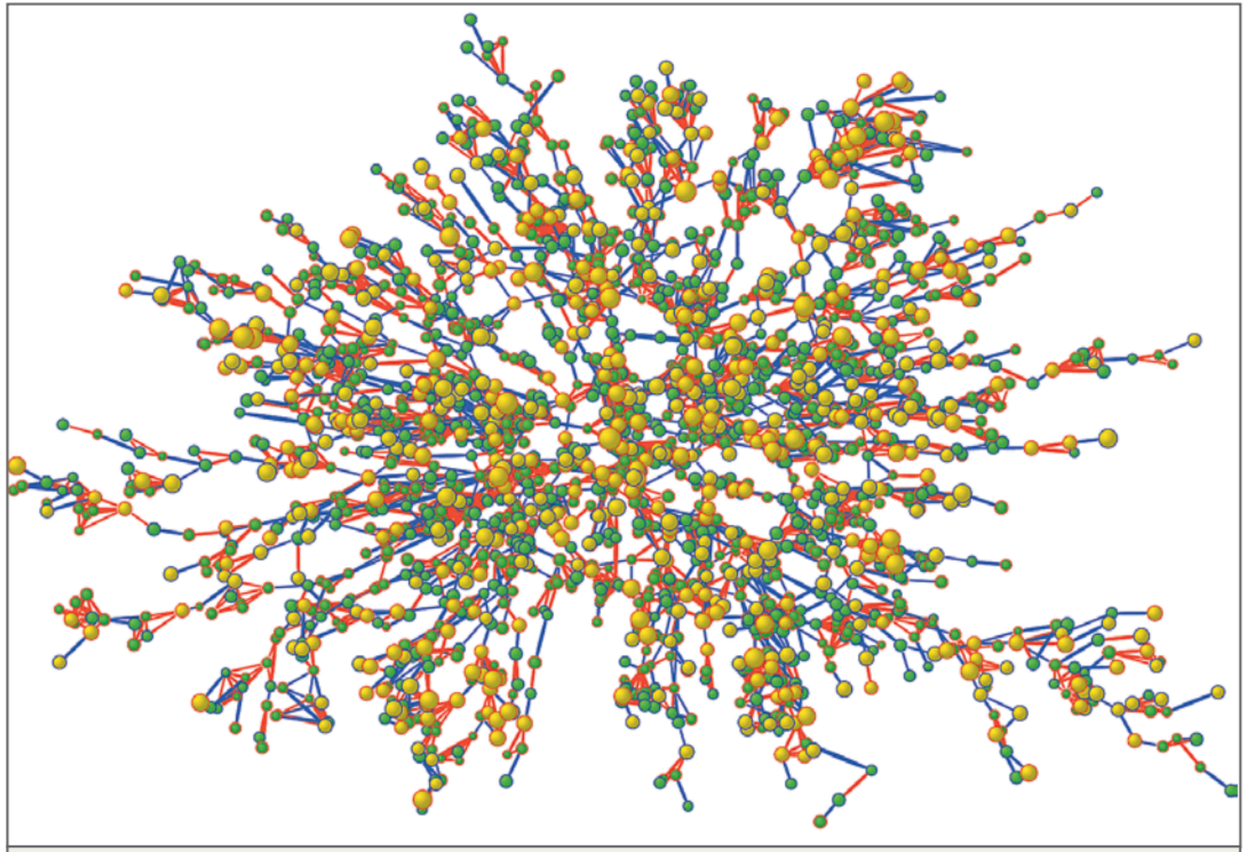


Figure 1. Largest Connected Subcomponent of the Social Network in the Framingham Heart Study in the Year 2000.
Each circle (node) represents one person in the data set. There are 2200 persons in this subcomponent of the social network. Circles with red borders denote women, and circles with blue borders denote men. The size of each circle is proportional to the person's body-mass index. The interior color of the circles indicates the person's obesity status: yellow denotes an obese person (body-mass index, ≥30) and green denotes a nonobese person. The colors of the ties between the nodes indicate the relationship between them: purple denotes a friendship or marital tie and orange denotes a familial tie.

SPECIAL ARTICLE

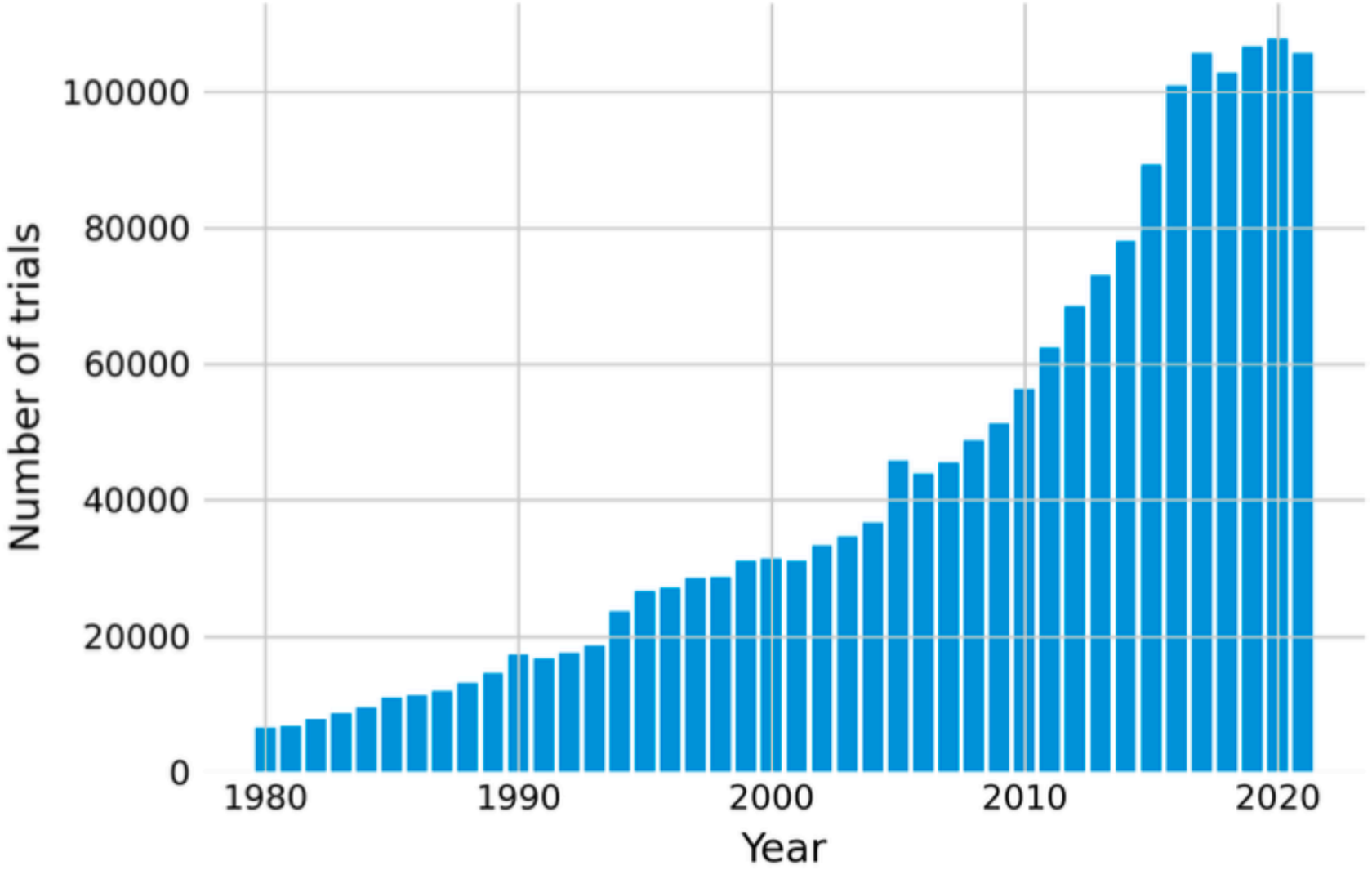
The Spread of Obesity in a Large Social Network over 32 Years

Nicholas A. Christakis, M.D., Ph.D., M.P.H., and James H. Fowler, Ph.D.

Experimental studies

Experimental studies are studies where we don't just observe the exposure; we control it with interventions.
Studies that experimentally test interventions are called ***trials****.*

Experimental studies



Randomized controlled trials

- Experimentally controlling interventions is obviously only possible if the deliberate application of the intervention is ethically justifiable.
- In many trials, the goal is to assess a type of intervention where one can assume that it is relatively safe.
- In the **RCT study design**, participants are assigned randomly to one of two (or more) groups. In the classical scenario of two groups, one group receives the intervention, and the other doesn’t.
- **Blinding**: ideally, nobody knows in which group they are.

Randomized controlled trials

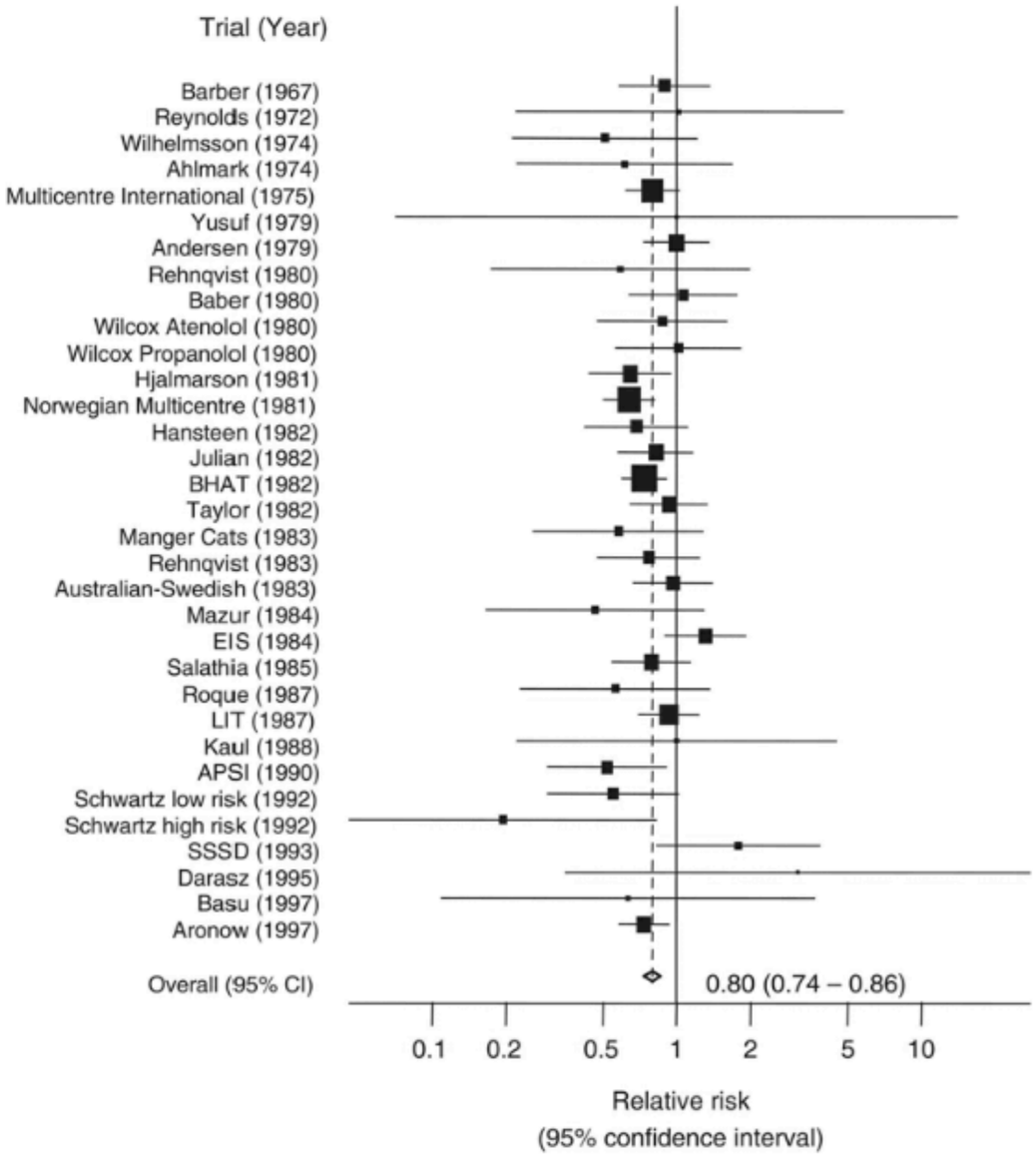
- **Single-blind trial**: only the participants are unaware of their group assignment.
- **Double-blind trial**: both the participants and the researchers don’t know the group assignment during the trial.
- **Open label trial**: there is no blinding.
- Overall, cohorts and trials, which can provide some of the strongest epidemiological evidence, are resource-intensive. **Digitization can help address this issue.**

Systematic reviews and meta-analyses

- One RCT provides a single data point. Over time, multiple studies may appear with different results.
- The evidence from such studies can be collected, analyzed, and synthesized in **systematic reviews**.
- Systematic reviews systematically select and assess the studies for a given topic, minimizing biases and errors.
- Systematic reviews can be done with different types of studies, not just randomized controlled trials.
- If the studies are similar enough, and the corresponding data available, **a statistical pooling of the results - a so-called meta-analysis - may be done** to determine the size of the effect, reported over multiple studies.

Forest plots

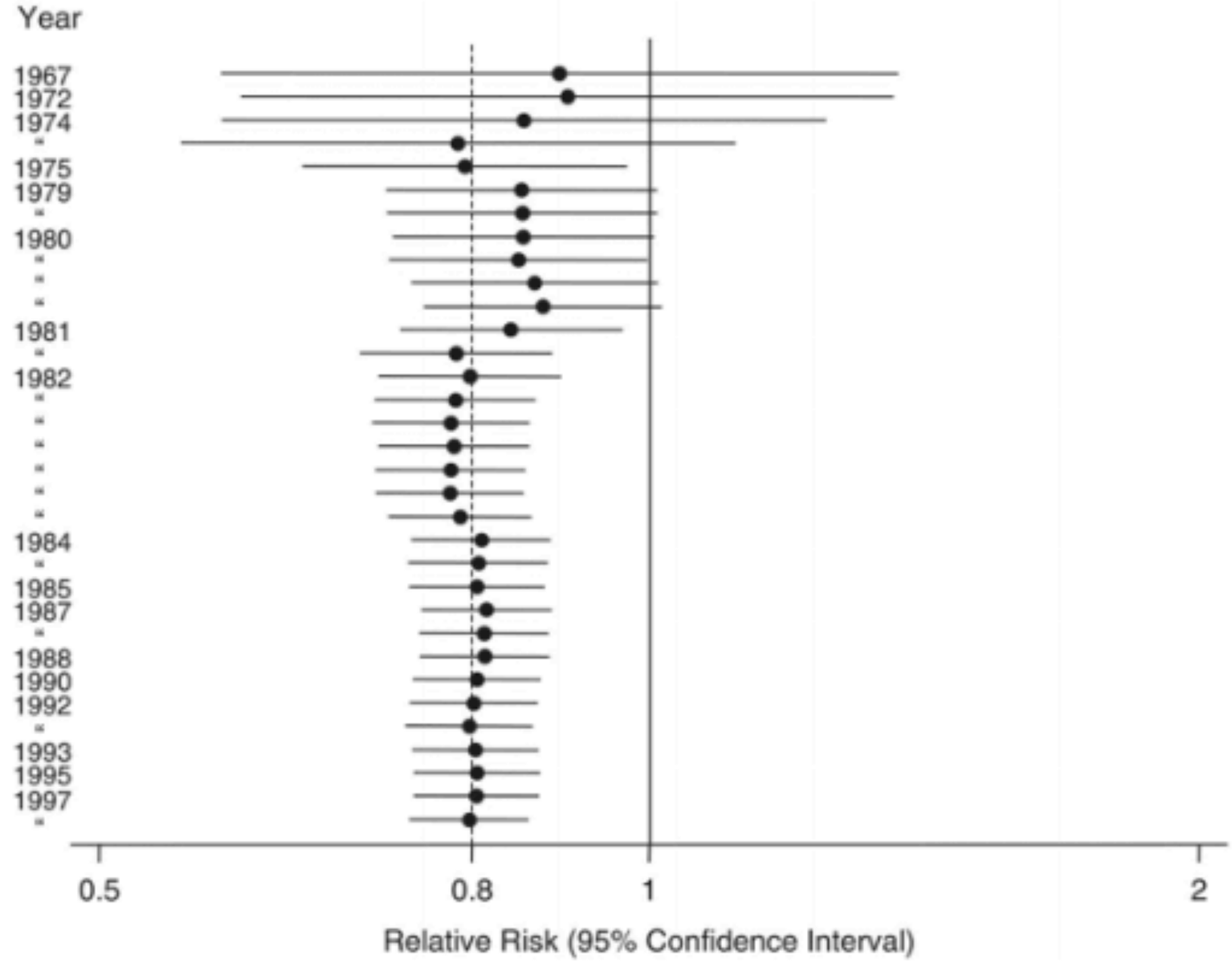
- Meta-analyses weigh studies according to their size
- The plot shows the mortality-preventing effect of beta-blockers after myocardial infarction.



Egger, Matthias, Julian PT Higgins, and George Davey Smith, eds. Systematic reviews in health research: Meta-analysis in context. John Wiley & Sons, 2022.

Cumulative meta-analysis

- What would the mean risk have been after every new study?
- A significant risk reduction became evident in 1981, which also raises **ethical questions** about the need of further studies.



Digital methods

- Digital methods can assist with the (semi-)automated assessment of new published studies.
- As an example, deep learning language models can support the creation and curation of **living systematic reviews** for a specific disease.
- **Living systematic reviews** are reviews which are continually updated, incorporating relevant new evidence as it becomes available.

Next… infectious disease epidemiology