Causal inference: challenges for health data analysts

Dr Jeremy Wyatt DM FRCP, Professor of Digital Healthcare, University of Southampton; Clinical Advisor on New Technologies, Royal College of Physicians

IMAGE: KACPER PEMPEL

Asthmopolis

Some advantages of big health data & “real world evidence”

- Datasets 100-1000 times larger than for RCTs, so can examine patient subgroups
- Data captured from routine care, so more representative / pragmatic
- Wider variety of data items, so can answer more questions eg. on side effects, effect modifiers
- Uses existing data, so quicker to start up and cheaper to answer questions (but EPIC in Cambridge cost £200M + 1-2 years of lower Care Quality Commission ratings)

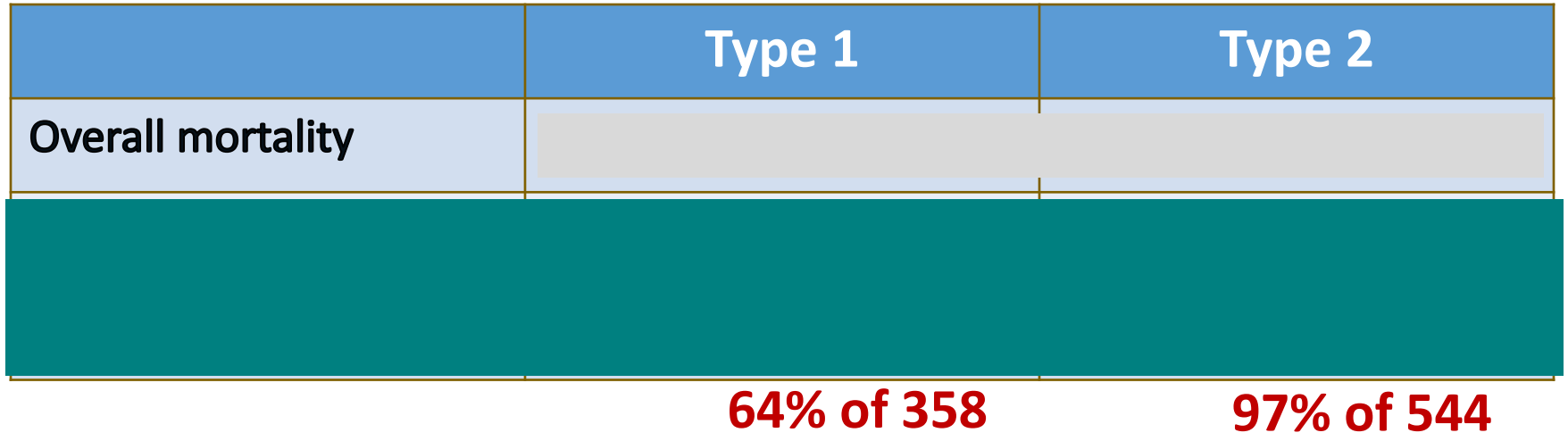
Sherman et al – FDA view on RWE - NEJMed 2016

Lars Hemkens, Ioannidis et al – Routinely collected data, promises & limitations. CMAJ 2016

Concerns about making inferences from routine data

https://utmost.org/going-through-spiritual-confusion/

Simpson's Paradox: mortality in diabetes



Data from Poole Diabetes cohort, cited by Julious et al BMJ 1994

Association vs. causation: Rochester library study

Study question: is hospital length of stay (LOS) shorter in patients whose doctors used the Rochester NY library ?

Method: compared LOS in patients of library-using Drs vs. patients of Drs who do not (case-control)

Result: LOS 1 day less in library-using Drs; savings would easily pay for the library !

Possible interpretations:

a) Library use is the *cause* of reduced LOS
b) Library use is a *marker* of doctors who keep their patients in hospital for less time
c) Library use *result*s from doctors keeping patients in hospital less !

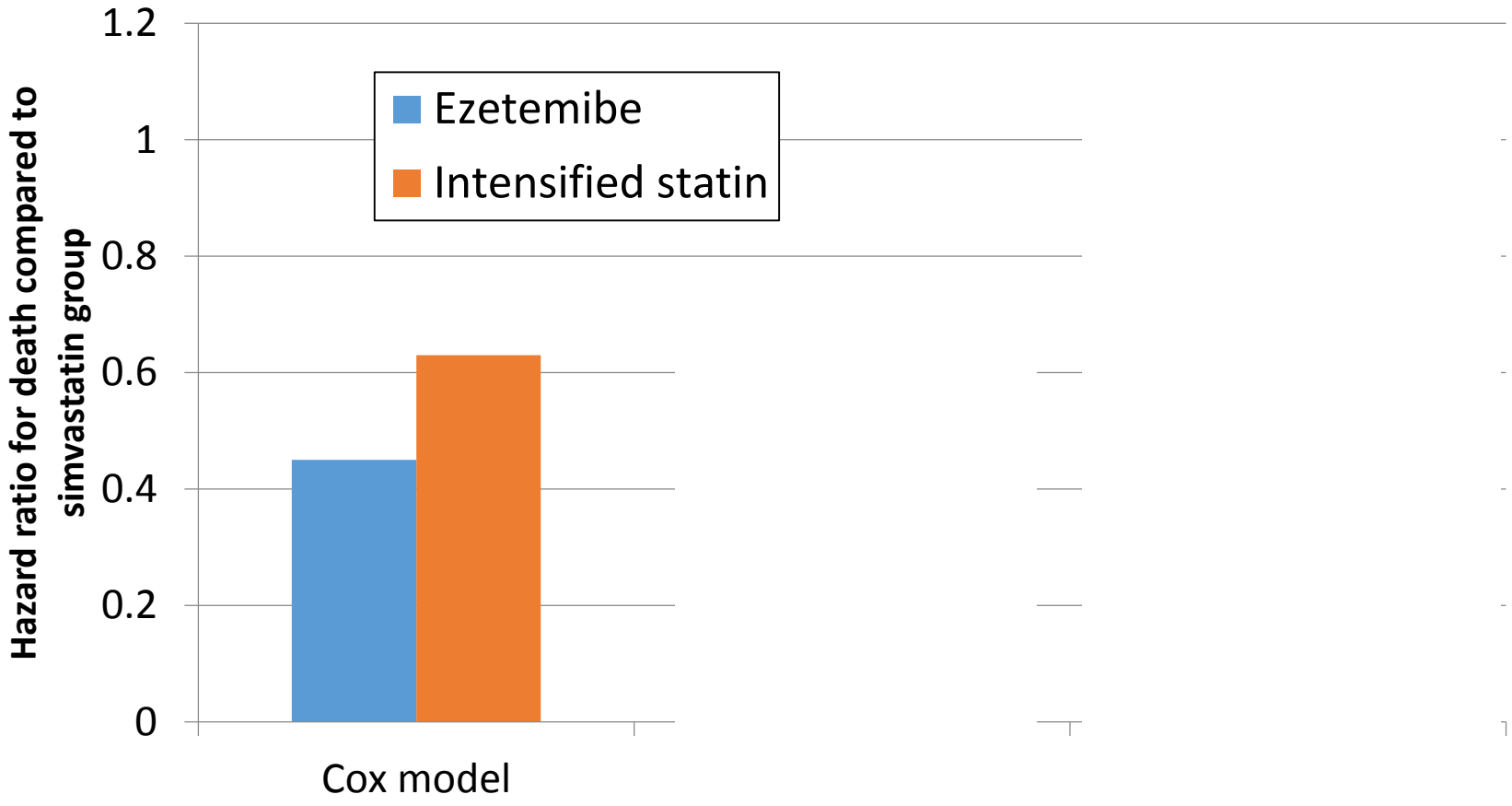
A better question:
What is the impact on LOS of providing a sample of doctors with access to the library ?

Confounding by indication

- 40% of cancer patients treated with new drug survive 5 years versus 30% of patients treated with old drug
- Difference persist despite taking account of differences in age, baseline cancer severity, genetic markers…
- Conclusion: the new drug reduces mortality by 10%
- But ***maybe*** allocation to the new drug depends on the doctor's intuition on who will survive (subtle predictive feature not recorded in any database)
- So, receipt of the new drug is a ***marker*** of better outcome - not the cause

Propensity scoring as a potential solution to this

The impact of bias on estimating mortality for ezetimibe in 2233 post-MI deaths (all cause mortality)



Eg. First incident MI; missing cholesterol levels; medication covariates

Source: Pauriah et al. Ezetimibe Use and Mortality in Survivors of an Acute Myocardial Infarction: A Population-based Study. **Heart** 2014

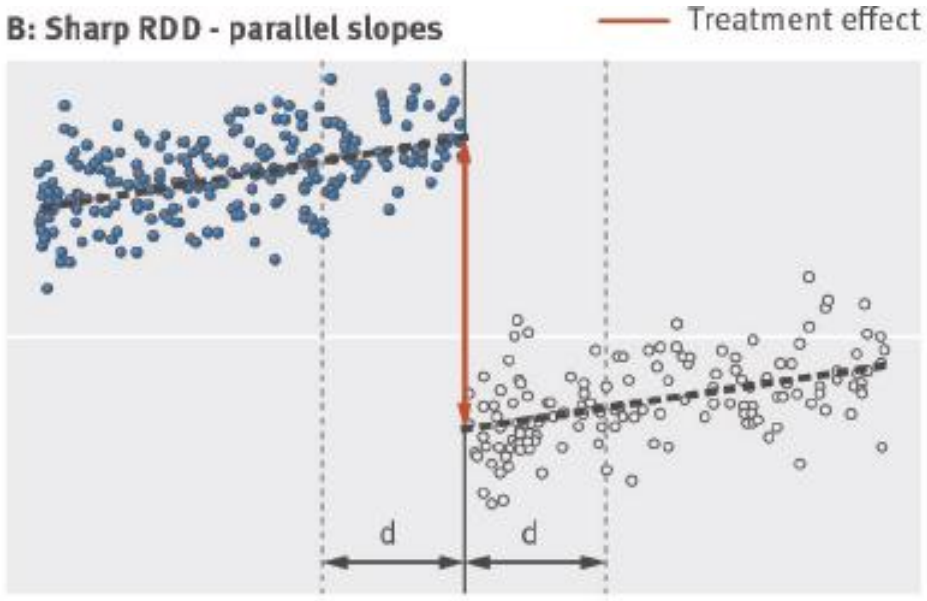
Estimating causality from big health data: some possible solutions

Understand & quantify the biases & apply expertise in relevant analytical methods:

- life course epidemiology
- multi-level modelling
- functional data analysis for intermittent monitoring data
- case-crossover design (Farrington)
- mediation and Rubin causal modelling
- instrumental variable analysis eg. regression discontinuity

Regression discontinuity design

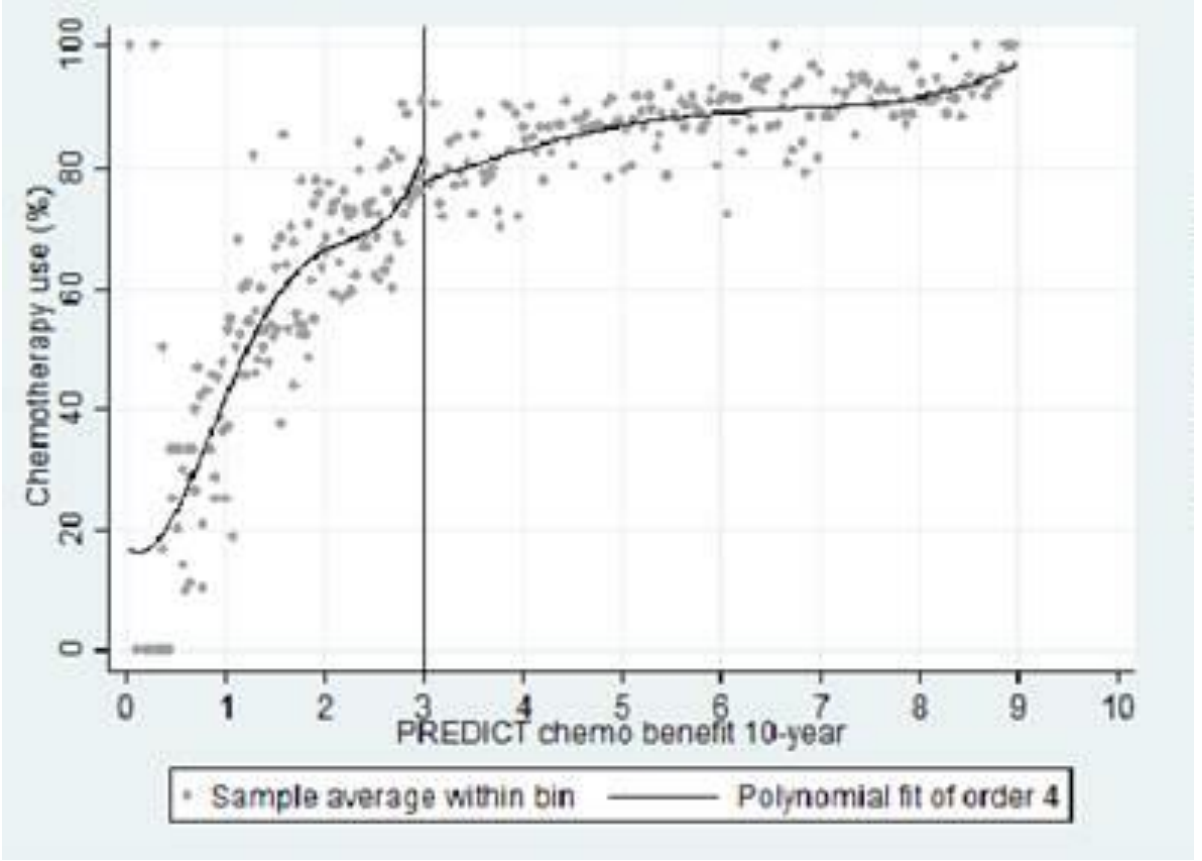
- Some drugs / procedures are used according to the threshold in a continuous variable eg. test result or predicted risk
- But due to measurement error, people ***just*** above & ***just*** below an allocation threshold are **very** similar
- So, if you have enough people to compare, you can *estimate* the impact of the intervention, just like an RCT…



Thistlethwaite & Campbell, 1960

Our *attempted* RDD study in 45,000 Scottish women with breast cancer

- NHS Predict score is an accurate, well calibrated algorithm for predicting p(Response|Chemotherapy)
- NICE: doctors should usually offer women chemotherapy when p(R|C) >5%, be reluctant to give it if <3% and discuss it with woman if 3-5%
- However, this is what happens in Scotland:



Gray, Hall, Marti, Brewster, Wyatt, to be submitted. Funded by CSO Scotland

Beware: non-randomised study designs are associated with replication failure !

Intervention studied	Original study design	Claim from original study	Findings from later studies / SRs
Post menopausal HRT	Non randomised	Prevents CAD & stroke	Ineffective
Vitamin E	**RCT**	1º CAD prevention	Ineffective
Vitamin E	Non randomised	2º CAD prevention	Ineffective
Inhaled nitric oxide	Non randomised	Treats ARDS	Ineffective
Endotoxin antibodies	Non randomised	Treats gram neg sepsis	Ineffective
Flavonoids	Non randomised	Prevents CAD	Effect smaller
Carotid endartectomy	Non randomised	Treats high grade stenosis	Effect smaller
Coronary stent vs. PTCA	Non randomised	Treats CAD	Effect smaller
Zidoudine	Non randomised	Treats HIV infection	Effect smaller

Ionnidis et al. Contradicted and initially stronger effects in ***highly cited*** clinical research. **JAMA** 2005 [original articles with 1000+ citations,1990-2003]

Conclusions

- We must use routine health data to improve patient safety, target interventions, evaluate process innovations and create the “Learning Health System”
- But it’s often hard to know if our data is biased or lacks key unmeasured variables
- Propensity scoring can help some times - but not other times
- More research is needed to understand when we can trust the results of PS, RDD and other inferential methods