AusHSI
AUSTRALIAN CENTRE FOR HEALTH SERVICES INNOVATION

Bringing health innovation to life

www.aushsi.org.au

Common pitfalls in data analysis

How infection prevention and control professionals should be using the infection control literature

Nicole White

ACIPC 2018

@nicolem_white

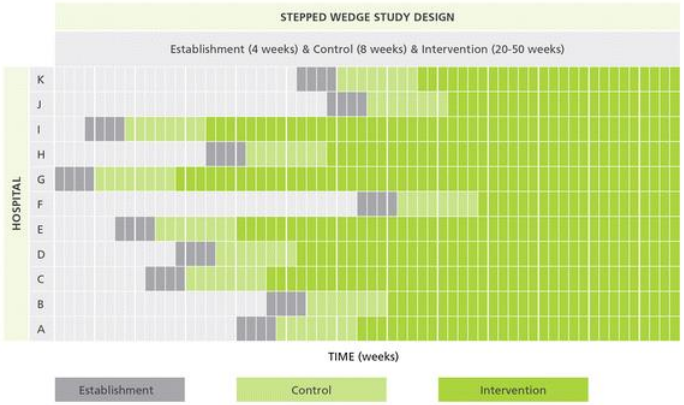
Disclosure statement: no conflicts of interest to disclose

Why pitfalls occur: Study designs, data sources and data analysis techniques used

Study designs for evaluating interventions:

Experimental/randomized

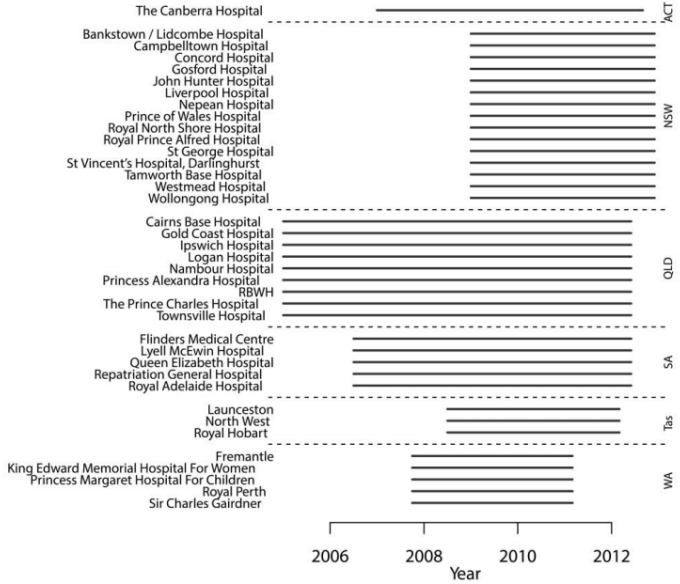
- Cluster parallel, cross-over, stepped wedge
- Prospective



Sources: Hall et al (2016), *Implementation Science*; Barnett et al (2014), *ICHE*

Observational/'Quasi-experimental'

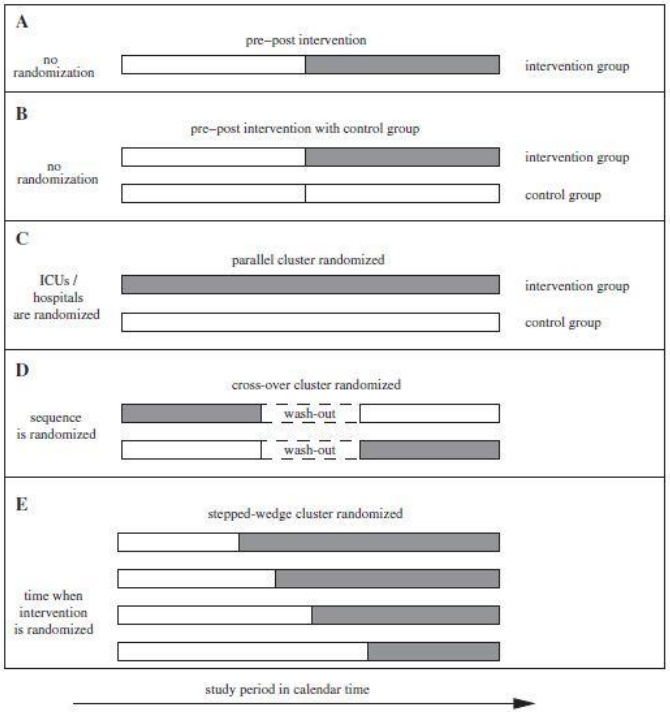
- Pre-post cohort, Matched case-control
- Retrospective



Why pitfalls occur: Study designs, data sources and data analysis techniques used

Variations of the same study design require different statistical considerations

- Single vs. multiple study sites
- Comparator
- Outcome type (binary, count, continuous)
- Confounders
- Nature of intervention effect



Source: Wolkewitz et al (2014); *Journal of Hospital Infection*

Transparency around data analysis techniques

Guidelines to promote best practice in reporting methods and findings (https://www.equator-network.org/)

- Systematic reviews: PRISMA statement
- Randomized trials: CONSORT statement
- Observational studies: STROBE statement

- At times, limited reporting of statistical methods used

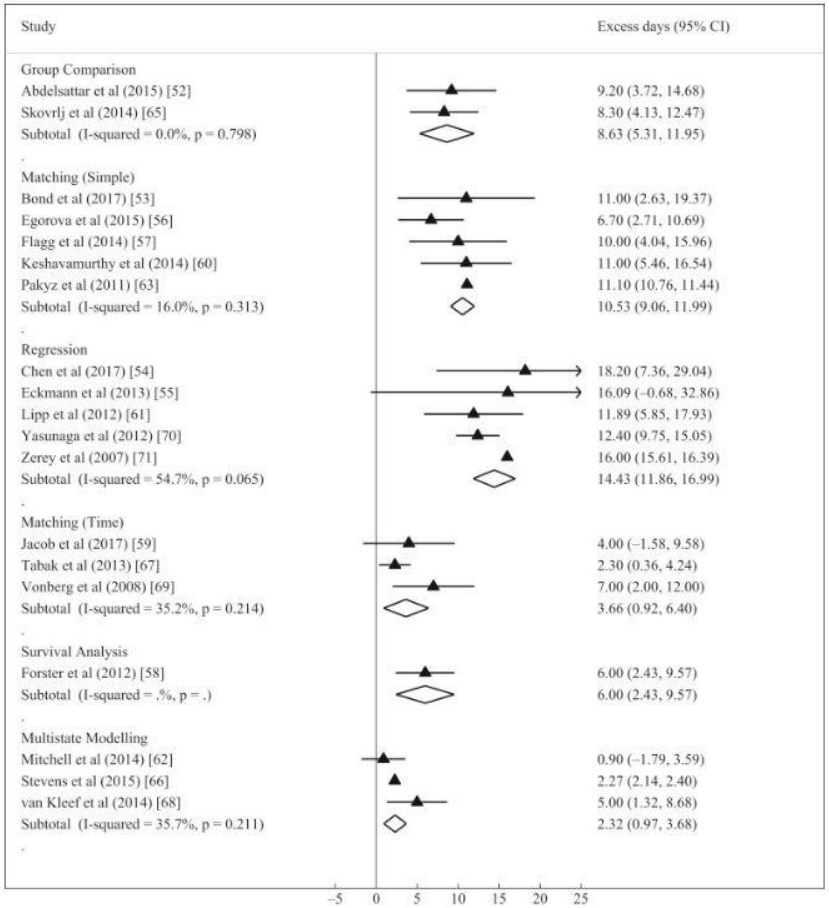
No clear consensus of which method(s) are "best"

"All models are wrong, but some are useful" (George Box)

Impact of data analysis method on HAI extra length of stay

Meta-analysis of methods for estimating CDI extra length of stay

- Group comparison
- Matching (simple, time)
- Regression
- Survival analysis
- Multistate modelling



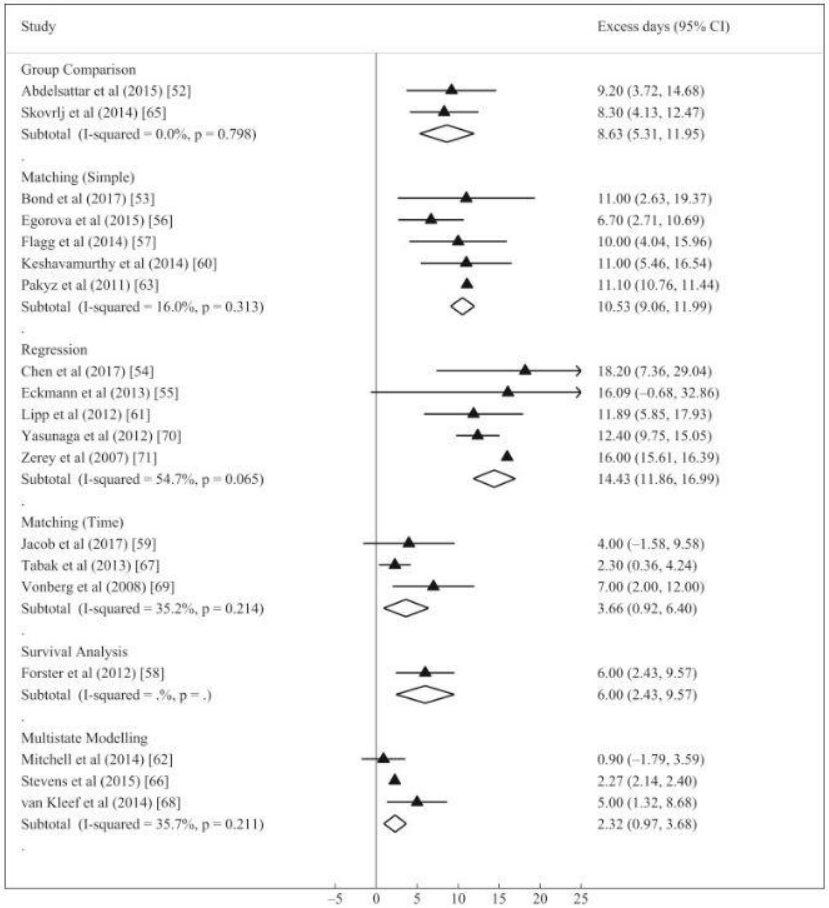
Source: Manoukian et al (2018), *Journal of Hospital Infection*

Impact of data analysis method on HAI extra length of stay

Meta-analysis of methods for estimating CDI extra length of stay

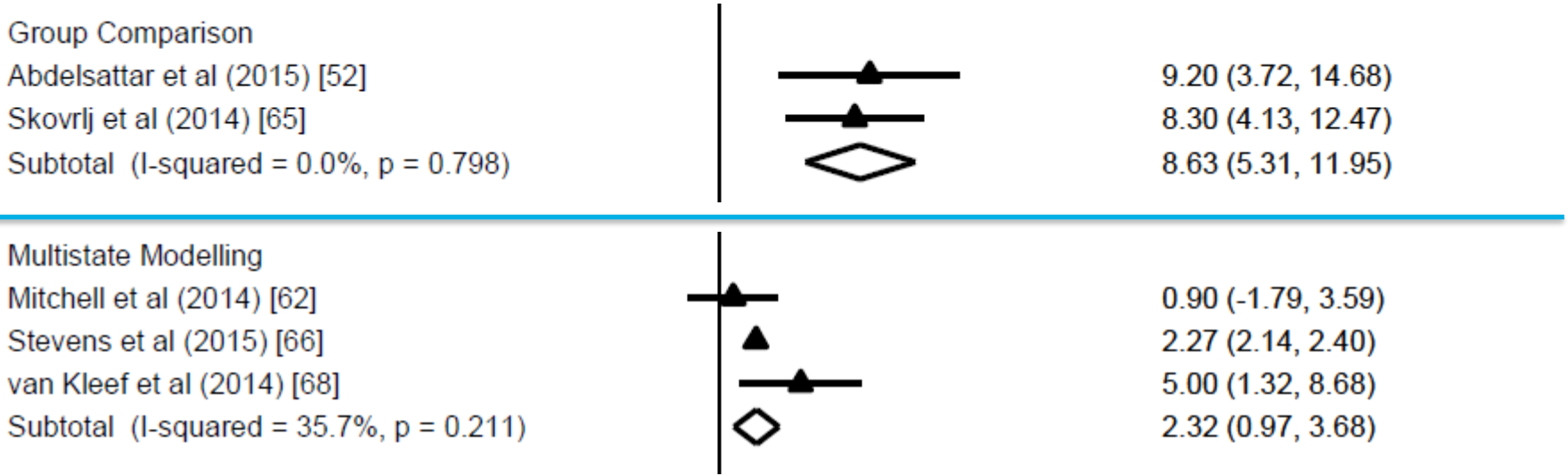
- Group comparison
- Matching (simple, time)
- Regression
- Survival analysis
- Multistate modelling

Methods ignoring time-dependence overestimated extra length of stay



Source: Manoukian et al (2018), *Journal of Hospital Infection*

Impact of data analysis method on HAI extra length of stay

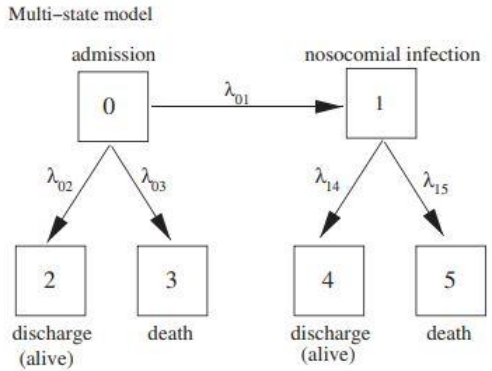


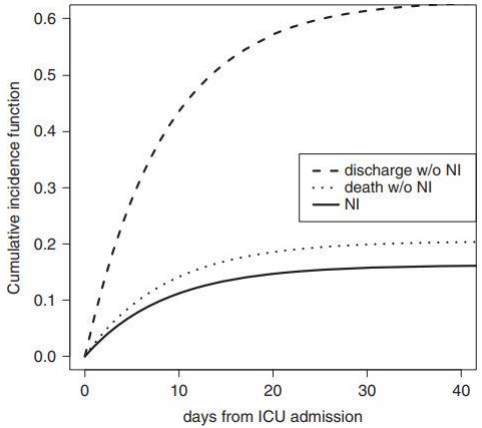
Difference in cost savings per infection prevented = $1,363
@ WTP of $216/bed day released (Page et al, 2017, *BMC health services research)*

Same outcome, many methods: Extra length of stay associated with HAIs

Methods for estimating extra length of stay should account for the dynamics of infection (Schumacher et al, 2013, *IJE)*

- Risk is time-dependent
- Competing risks (death, discharge)



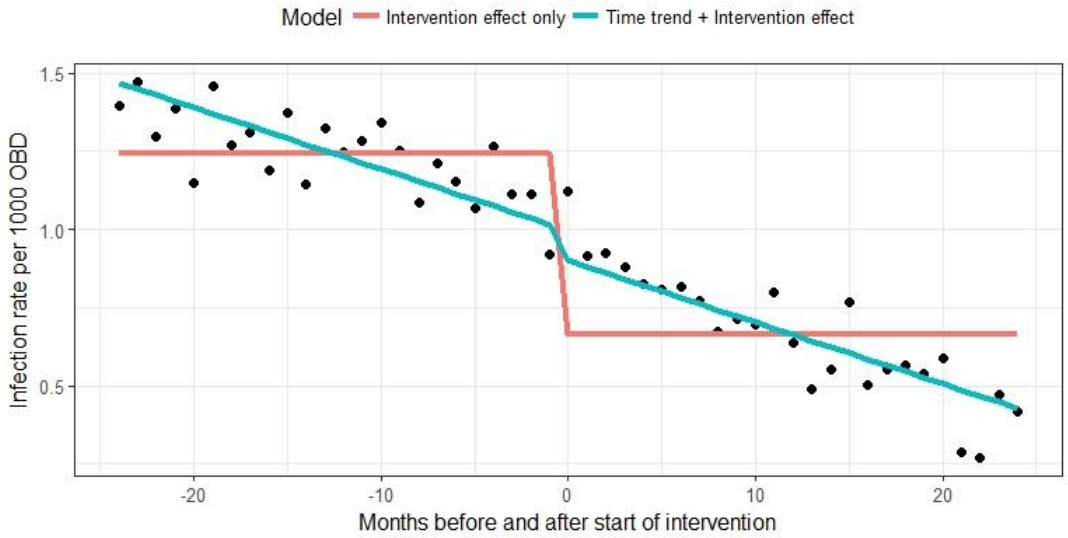


Source: Wolkewitz et. al (2017); *ICHE*

Confounding in intervention studies

Analysis of pre/post changes should account for pre-existing trends in outcome

- Impact on size of intervention effect

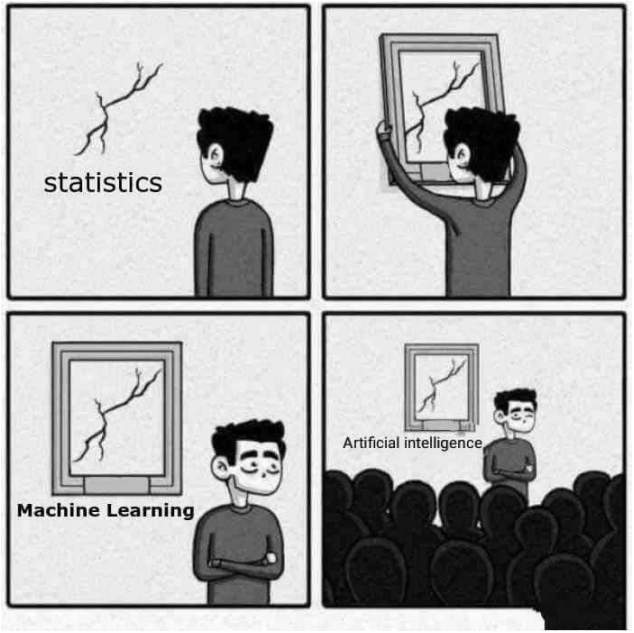


Intervention effect only: -0.58 (95% CI: -0.68 to -0.48)

Time trend + Intervention effect: -0.10 (95% CI: -0.21 to 0.02)

The promise of big data in infection control

- Use statistical theory combined with computational algorithms to harness volume and variety of new data sources
  - Electronic health records
  - Social media
  - Internet search queries
- Big data analysis as the new form of observational design
  - Infectious disease surveillance
  - Risk prediction
- Many are applications of traditional statistical methods, so the same rules apply!
- Potential pitfalls:
  - Unsupervised – focus on prediction, not meaningful associations
  - “Big Data hubris”



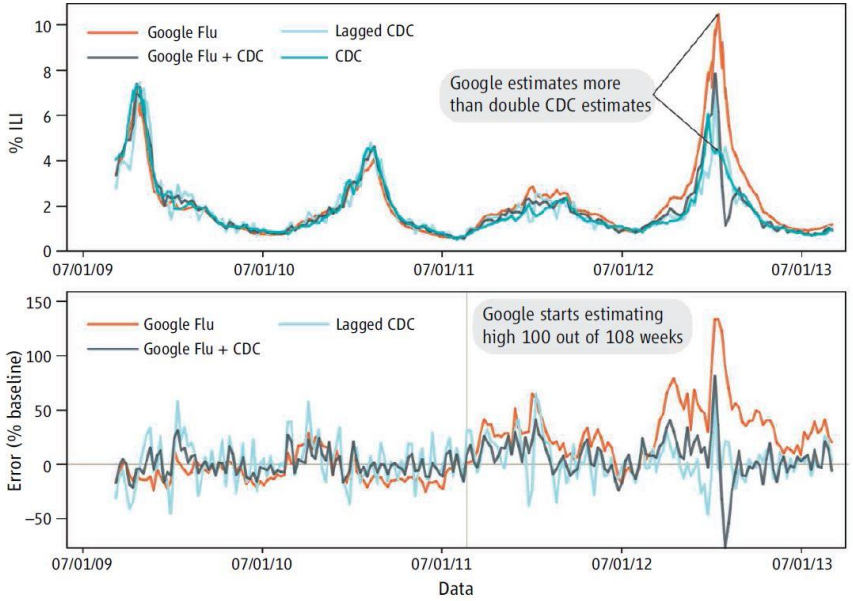
Source: sandserifcomics

Pitfalls in big data analysis: Lessons from Google Flu Trends

- Model used Google search terms to predict % Influenza like illness(ILI)

Case of GFT:

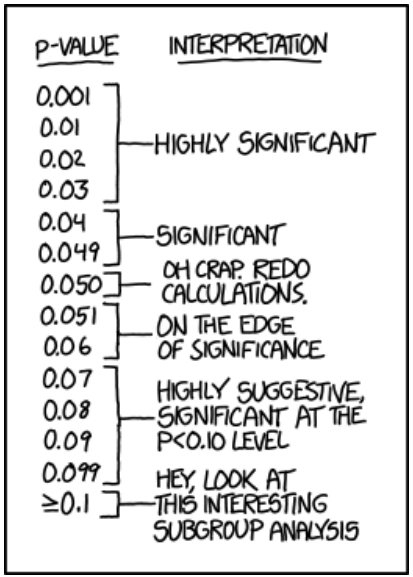
- In 2013, two-fold overestimation of %ILI compared with traditional surveillance (CDC)
- Spurious correlations between search terms and influenza rates
  - 'High school basketball' (Ginsberg et al (2008); *Nature)*
- Changes in search behaviour over time
- Outperformed by predictions based on historical CDC data



Source: Lazer et. al (2014); *Science*

Concluding thoughts: how to identify pitfalls better

- Reporting of data analysis techniques based on CONSORT/STROBE guidelines
  - How have studies justified their choice of method?
  - Technical supplements
  - Interrogation of statistically significant results
- Big data analytics – beware big data hubris
  - Validation and Evaluation



Source: https://xkcd.com/1478/ (Not a guide!)

AusHSI

AUSTRALIAN CENTRE FOR HEALTH SERVICES INNOVATION

Bringing health innovation to life

www.aushsi.org.au

PARTNERS

Queensland Government

QUT ihbi Institute of Health and Biomedical Innovation

The University Of Queensland AUSTRALIA