# Evaluation of healthcare apps: a UK view on challenges and alternatives

Jeremy Wyatt DM FRCP ACMI Fellow

Professor of Digital Healthcare & Director,
Wessex Institute of Health Research,
University of Southampton, UK

Digital health adviser to Royal College of Physicians, NICE, Medicines & Healthcare Regulatory Authority, Care Quality Commission

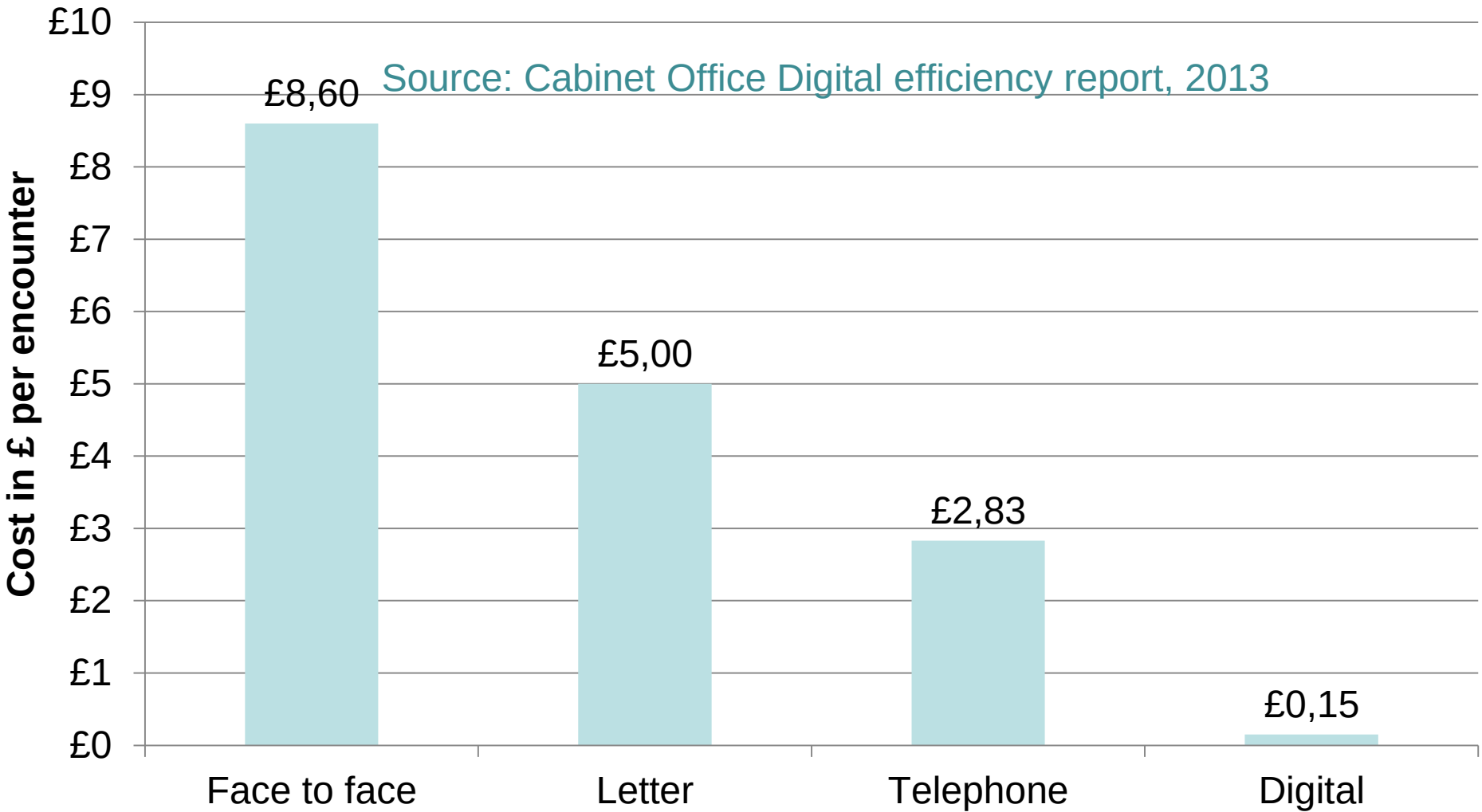
# Overview

UNIVERSITY OF Southampton

- Why are apps important ?
- Why is app evaluation hard ?
- *[What to test and how ? - 11am panel session, Sal AB]*
- A few options for app regulation and quality improvement
- How these work out in practice
- What NICE and the NHS are doing
- Conclusion: we need various approaches to add new survival pressures to the app ecosystem

# Why use digital channels ?

## Mean UK public sector cost per completed encounter across 120 UK councils



# Why should apps work ?

UNIVERSITY OF Southampton

1. Face-to-face contacts with health professionals do not scale, but software does

2. Smart phones are used by 75%+ of UK adults:
   - Cheap, convenient, fashionable, trusted by users
   - Inbuilt sensors +/- wearables allow easy measurements
   - Multiple communication channels: SMS, voice, video, apps, VR…

3. mHealth apps enable:
   - Unobtrusive alerts to take actions, record data eg. PROMs
   - Delivery of Susan Michie's 94 behaviour change techniques
   - Tailoring, which makes behaviour change more effective (d=0.16, Lustria, J H Comm 2013)

# *The risk of "We know it works"*

UNIVERSITY OF Southampton

Motorbike paramedics *must* be effective:

- Get to accident faster than ambulance
- Paramedic is trained to resuscitate
- Carry relevant equipment


What could go wrong ?

*"Full advanced life-support did not decrease mortality or morbidity... **mortality was greater** among patients with Glasgow Coma Scale scores above 9"* Stiell IG. CMAJ. 2008

# Some plausible eHealth technologies that failed so far

UNIVERSITY OF Southampton

Diagnostic decision support (Wyatt RCT, MedInfo ‘89)

Integrated medicines management for a children’s hospital (Koppel, JAMA 2005)

MSN Messenger patient triage (Eminovic, JTT 2006)

Smart home applications for fall detection etc.:

- “*The effects of [these] technologies is not known. Better quality research is needed*.” (Martin, Cochrane Review 2008)
- “*The technology readiness level for smart homes & home health monitoring technology is still low. There is no evidence that [these] technologies address disability prediction, health-related quality of life or fall prevention.”* (Liu L et al. Int J Med Inform. 2016)

# Bad quality apps

UNIVERSITY OF Southampton

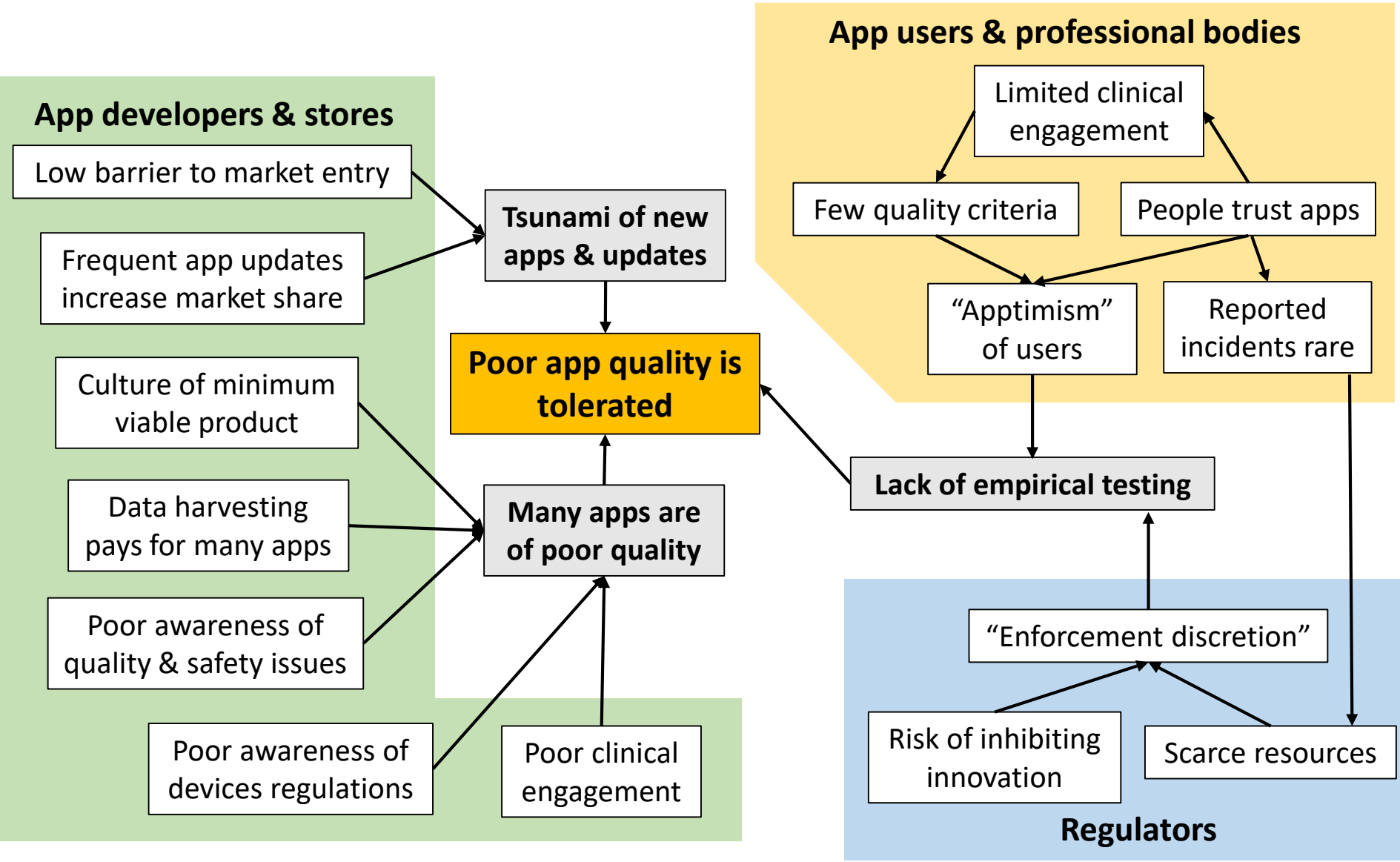
Melanoma de
Opiate drug c
Cardiac risk c
https://www.y
Smoking cess
Asthma, diab
Fake iPhone
Acne treatme
Harvesting of
Heart rate ca
Drink thin app


# Why ineffective health apps matter

UNIVERSITY OF Southampton

- Can cause harm directly: unsafe apps, inaccurate online diagnostic triage…
- Or indirectly ("opportunity cost"): ineffective health promotion app delays person from using effective app, going to GP or dietician
- Can waste health system resources: money, professional time, facilities
- Risk making users, professionals and policy makers cynical about digital health: a "great revulsion" (Muir Gray)

# Why poor app quality is tolerated



# Some challenges of app evaluation

UNIVERSITY OF Southampton

Tsunami of new apps: 1000 new apps on iPhone platform *per da*y – c. 5-7% health related (https://www.statista.com)

Rapid update cycle – often weekly (partly to retain place in app store top 50)

About 1/3 of asthma apps disappear each year (Huckvale 2014)

Zero barriers to market entry (eg. MIT app inventor toolkit) so huge variation in app quality

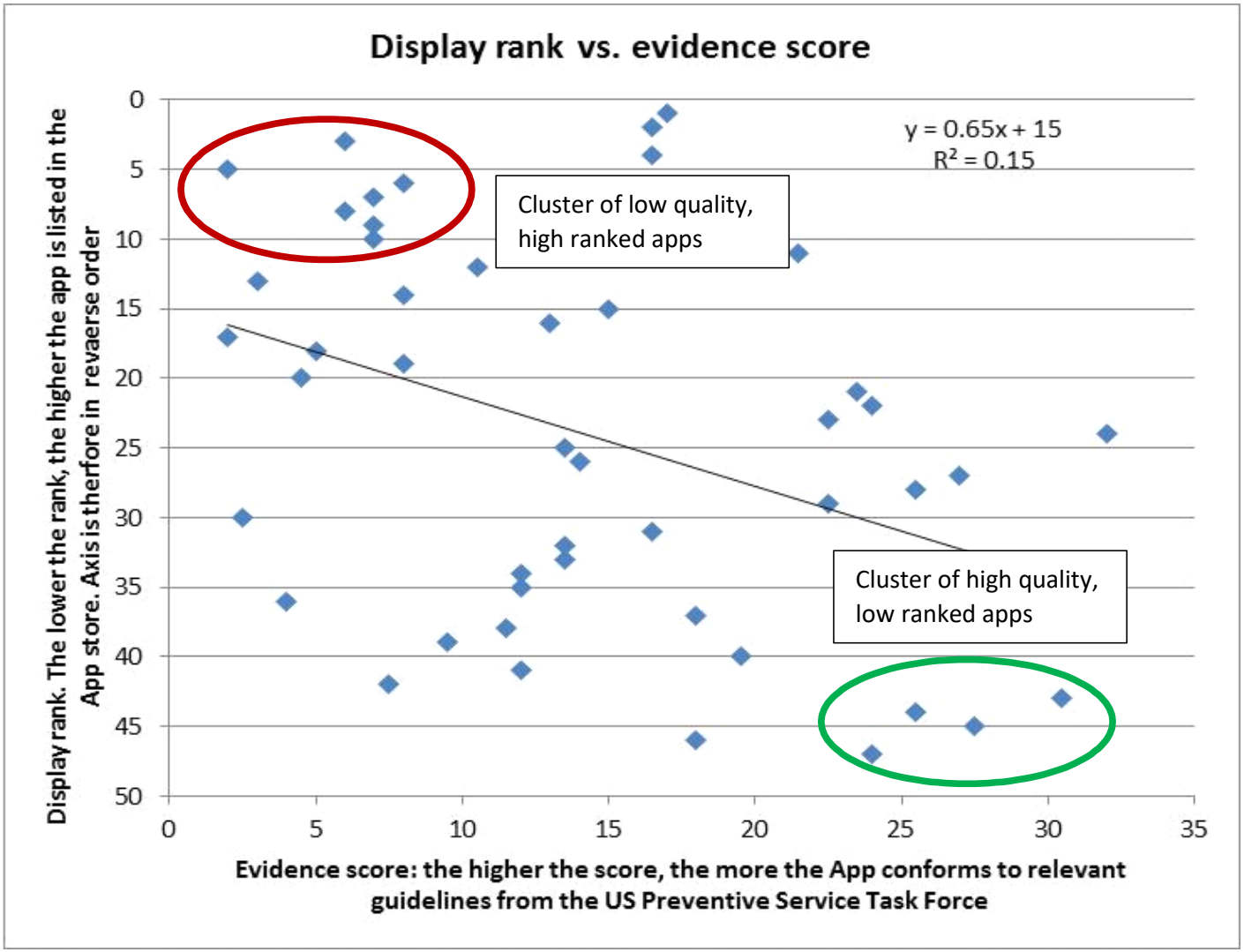
Huge variety of users (public, patients, professionals) & use cases - from lifestyle improvement to controlling a surgical tele-manipulator or insulin pump…

**So, we need a quality approval process !**

# Possible quality approval processes for apps

| Methods | Advantages | Disadvantages | Examples |
|---|---|---|---|
| **Wisdom of the crowd** | Simple user ranking | Hard for users to assess quality; click factory bias | **App stores** MyHealthApps |
| **Users apply quality criteria** | Explicit | Requires dissemination; will all users apply criteria ? | **RCP checklis**t |
| Classic peer reviewed article | Rigorous (?) | Slow, resource intensive, doesn't fit App model | 470 PubMed articles |
| Physician peer review | Timely, Dynamic | Not as rigorous Scalable ? | iMedicalApps, MedicalAppJournal |
| Developer self-certification & labelling | Dynamic | Requires developers to understand & comply; checklist must fit apps | HON Code **RCP checklist** **NHS App store** |
| Developer support | Resource light | Technical knowledge needed Multitude of developers | BSI PAS 277 |
| **CE marking**, external regulation | Credible | Slow, expensive, apps don't fit national model | FDA, **MHRA** |
| **Curated store** | Credible | Resource intensive | **NHS App Store** |

# User ratings: app display rank versus app adherence to evidence




Re-analysis of data on 47 smoking cessation apps from Abroms 2013 (Wyatt, BMC Med 2018 – in press)

# RCP quality criteria for physician apps, based on Donabedian 1966

UNIVERSITY OF Southampton

**Structure** = the app development team, the evidence
change r

**Processe**

**Outcome**
efficacy, u

## What makes a good clinical app? Introducing the RCP Health Informatics Unit checklist

**Authors:** Jeremy C Wyatt,[A] Harold Thimbleby,[B] Paul Rastall,[C] Jan Hoogewerf,[D] Darren Wooldridge[E] and John Williams[F]

Doctors increasingly rely on medical apps running on smart phones or tablet computers to support their work. However, these apps vary hugely in the quality of their data input screens, internal data processing, the methods used to handle sensitive patient data and how they communicate their output to the user. Inspired by Donabedian's approach to assessing quality and the principles of good user interface design, the Royal College of Physicians' Health Informatics Unit has developed and piloted an 18-item checklist to help clinicians assess the structure, functions and impact of medical apps. Use of this checklist should help clinicians to feel more confident about using medical apps themselves, about recommending them to their staff or prescribing them for patients.

Wyatt JC, Thimbleby H, Rastall P, Hoogewerf J, Wooldridge D, Williams J. Clin Med (Lond). 2015 (15):519-21.

# First NHS Apps Library: ignored data protection

UNIVERSITY OF Southampton

Huckvale et al 2015 “man in the middle” study of 79 accredited lifestyle apps from the NHS Apps library:

- Only 53 (67%) had a privacy policy: policies vague, did not explain types of data being shared
- No app encrypted data held on device
- 70 (89%) of apps leaked confidential data over network
- 35 included identifiers, 23 sent IDs without encryption
- 4 (5%) apps sent **both** IDs and health information without encryption

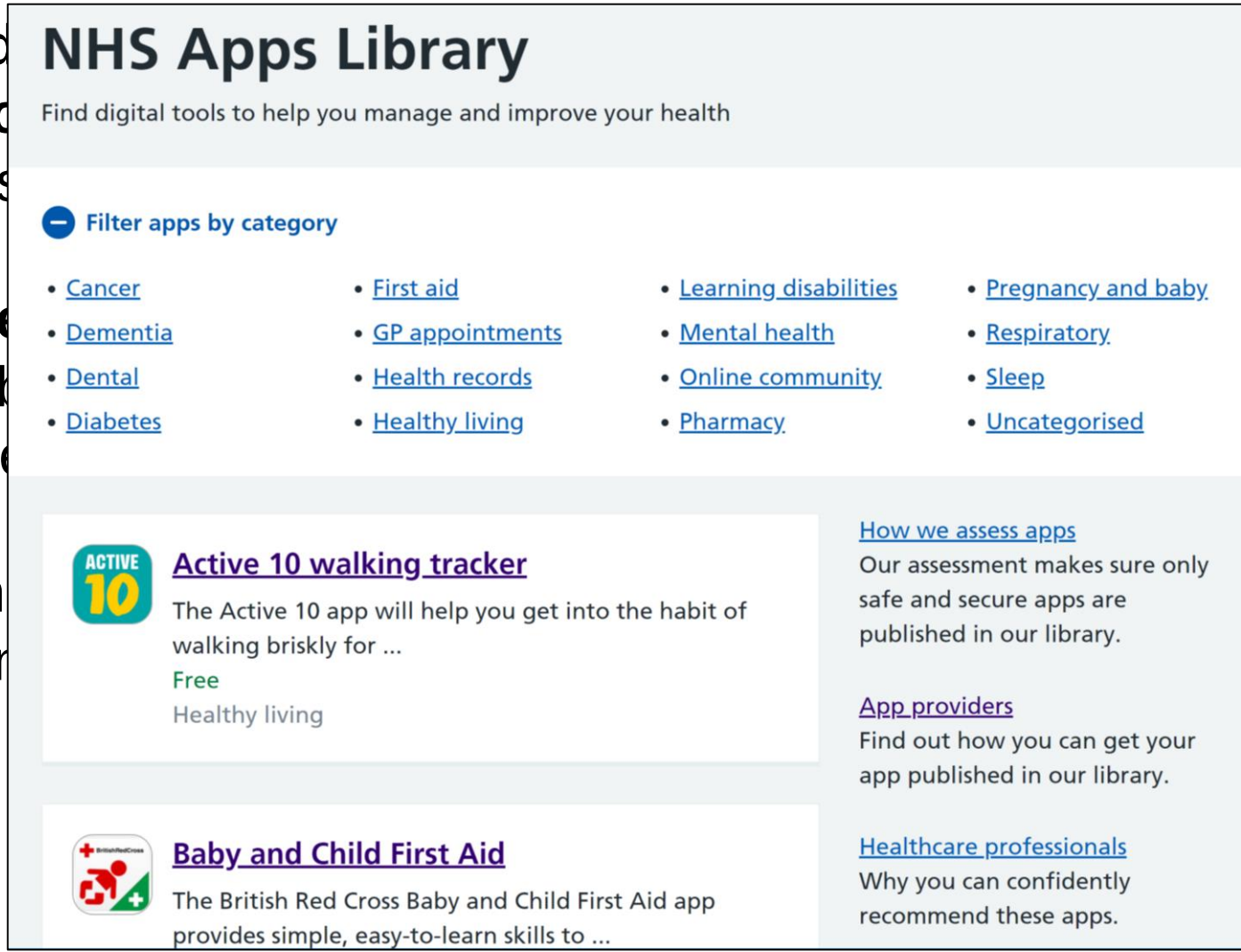
# New NHS Apps Library, 2017 on https://apps.beta.nhs.uk/

UNIVERSITY OF Southampton

**48 million visits** a month; about 120 apps so far

**Three labelled**

1. **NHS Appro**
   effectivenes
   "supportive
2. **Being teste**
   safety, usab
   for evidence
3. **No badge**:
   usability an
   NHS for clir



# NHS Apps Library 3 stage approval process (35 page form)

1. Submit: if app aligns with NHS priorities ie. maternity, social care, long term conditions, cancer, mental health (2 pages of questions)
2. Pre-assess: fits clinical expectations of NHS apps, has CE mark if medical device (4 pages)
3. Assessment: if it adheres to core obligations: effectiveness (3), clinical safety (standard DCB0129, - 1 page), **data protection (20),** cyber security (OWASP standard - 2), usability & accessibility (2), interoperability (1), technical stability (1)

https://developer.nhs.uk/digital-tools/daq/

# NICE app assessment process

Use Medtech Innovation Briefing approach, https://www.nice.org.uk/about/what-we-do/our-programmes/nice-advice/medtech-innovation-briefings

Not guidance for NHS – review of evidence, including:

- Evidence on effectiveness
- Costs and resource use
- Usage and user experience
- Specialist commentator comments
- Patient organisation comments

Only 5 apps reviewed 2015-17: Sleepio, GDm-health, ChatHealth, AliveCor, Mersey Burns – no resources for more

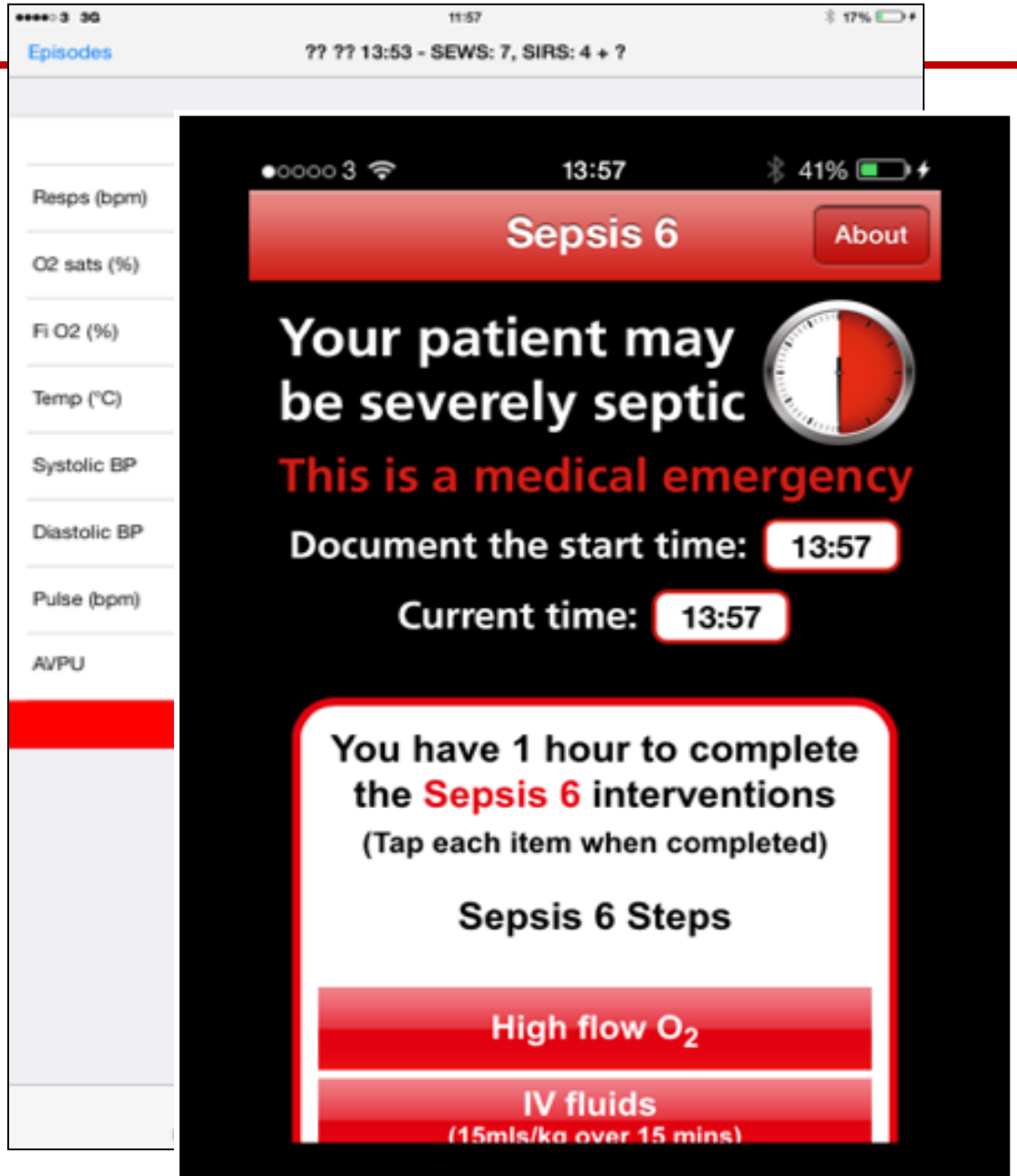
# New NICE / NHS eHealth “Evidence for Effectiveness" programme

UNIVERSITY OF Southampton

- Developing guidance & standards to streamline support for NHS digital adoption
- Led by NHS England with NICE, Public Health England, MedCity, DigitalHealth.London,
- Builds on previous NICE Health App Briefings to develop functional taxonomy of apps; requires higher level evidence for higher risk apps
- In HTA tradition, but aims to be rapid and embrace tools like real world evidence
- Probably *cost consequence* economic analysis for apps with significant NHS impact

http://www.medcityhq.com/evidence-for-effectiveness/

# Does CE marking identify high quality apps ?




Sepsis 6 app: screenshot shows data from ***two*** patients

**A different Sepsis 6 app:** screenshot shows ***clipped list of key actions*** to complete – cannot scroll down

Source: Harold Thimbleby, Swansea

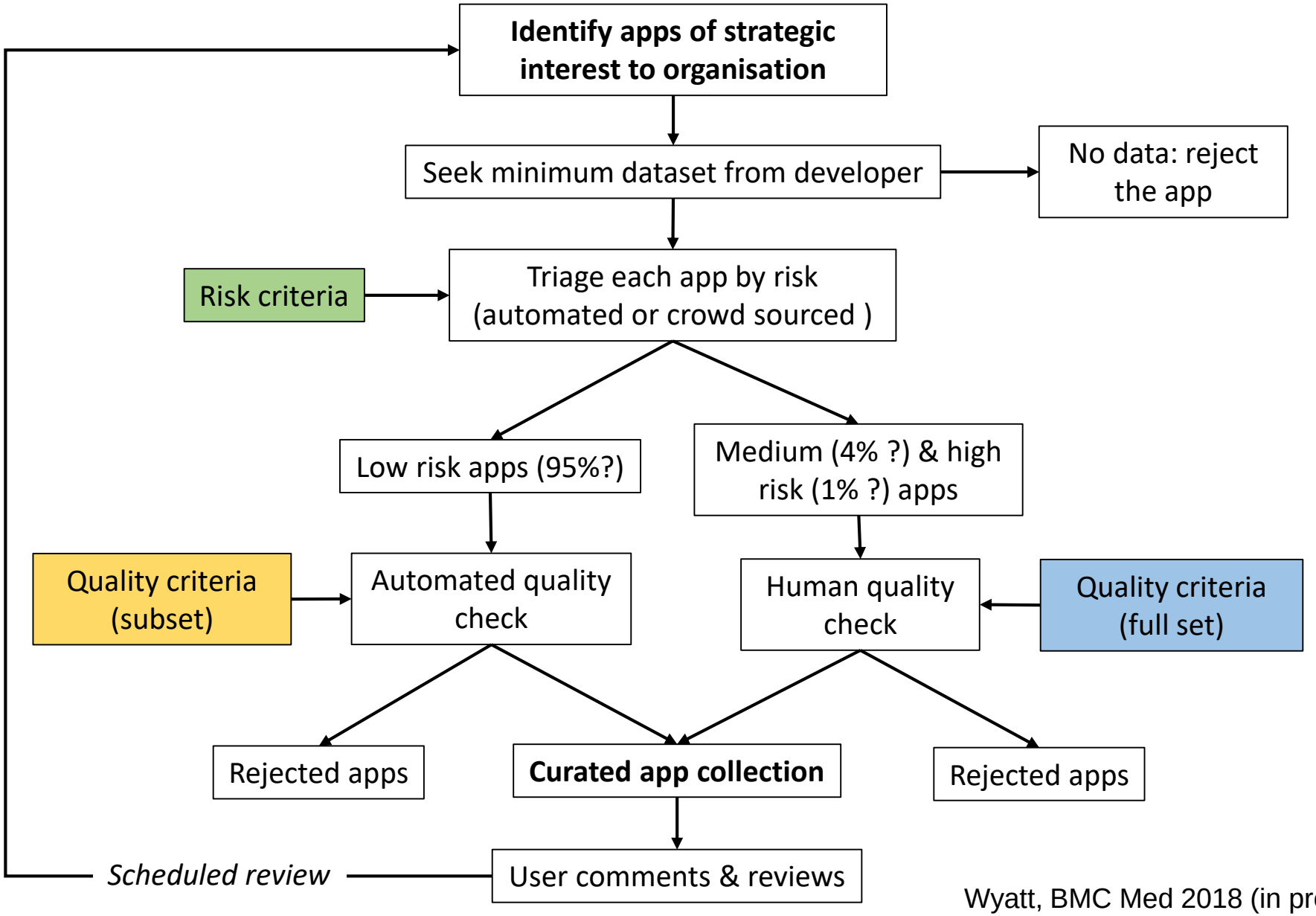
# Some criteria for an app quality approval process

UNIVERSITY OF Southampton

1. Empower patient & professional choice
2. Promote survival of the fittest and a proper market, not just innovation for its own sake
3. Use criteria that make sense to patients, professionals, health systems & industry
4. Scalable to thousands of apps
5. Proportionate to clinical risk
6. Resistant to manipulation, and auditable

*Source: JW submission to NICE / PHE, Feb 2016*

# A process for organisations to develop a risk and quality based curated app store




Wyatt, BMC Med 2018 (in press)

# Some actions physicians themselves can take to improve app quality

UNIVERSITY OF Southampton

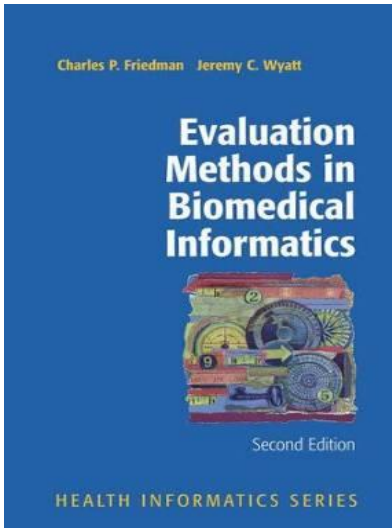
1. Report unsafe apps or apps which harvest data to professional / regulatory authorities
2. Use checklist to guide informal study of app before you recommend it to patients or staff
3. Discuss app quality, “apptimism”, methods to report poor quality apps with peer / patient groups
4. Help app developers identify good evidence or algorithms
5. Carry out well-designed evaluations of app accuracy, impact or effectiveness
6. Support professional societies, patient groups, regulators, the media etc. promoting better quality apps

# Conclusions

UNIVERSITY OF Southampton

1. Apps ***can*** bring great benefits to patients and professionals
2. However, their quality varies too much, there are huge numbers and they change all the time
3. They therefore pose a real challenge to evaluators, regulators and health systems
4. Some useful innovations may include:
   a) Open, agreed, risk-based criteria
   b) Self-declared label with intended user, purpose, test results + random checks of these
   c) Research to identify quality predictors (eg. developer)
   d) Specific curated app stores built using a moderated crowd-sourcing process (patients or professionals)

j.c.wyatt@soton.ac.uk



# Maybe we need to think differently

UNIVERSITY OF Southampton

| Old think | New Think |
|---|---|
| Paternalism: we know & determine what is best for users | Self determination: users decide what is best for them |
| Regulation will eliminate harmful Apps after release | Prevent bad Apps - help App developers understand safety & quality |
| The NHS must control Apps, apply rules and safety checks | Self regulation by developer community Consumer choice informed by labelling |
| App developers are in control | Aristotle's *civil society** is in control |
| Quality is best achieved by laws and regulations | Quality is best achieved by consensus and culture change |
| Apps symbolise innovation (and many harvest data for resale) | App innovation must balance benefits and risks |
| An Apps market driven by viral campaigns, unfounded claims of benefit | An Apps market driven by *fitness for purpose* (ISO) & evidence of benefit |

**The elements that make up a democratic society, such as freedom of speech, an independent judiciary, collaborating for common wellbeing*

# RCP app checklist part 1

UNIVERSITY OF Southampton

App name and version: _____ For iPhone / Android / other:_____

Date of filling out this checklist: _____

1. **Who developed the app, and what's inside it?**
   a) Is it clear who this app is for and how it should be used? Yes / No / Don't know
   b) Is it clear which problem the app is designed to alleviate or what outcome it helps to promote? Yes / No / Don't know
   c) Do the app developer and sponsor seem well informed about this problem or outcome, and likely to be unbiased in their approach to it? Yes / No / Don't know
   d) Have they located sound, relevant, up-to-date evidence, images, video etc to use in their app? Yes / No / Don't know
   e) Do the app screens look well designed, is text clear? Not applicable / Yes / No / Don't know
   f) Is it clear what data the app needs from the user with units defined, out of range detection and a 'clear last patient' button? Not applicable / Yes / No
   g) Does the app collect any identifiable patient information? Yes / No / Unclear
   h) Does it seem to keep user and patient data secure and private? Yes / No / Don't know
   i) If the app is designed to support any medical task,* is it CE marked? Not applicable / Yes / No / Unclear

# RCP app checklist part 2

UNIVERSITY OF Southampton

**2. How well does the app work?**

a) Is the app fast and easy to use in clinical settings? Yes / No / Don't know

b) Does the app give the user usable answers or advice quickly? Not applicable / Yes / No / Unclear

c) Do the answers, advice or calculated risks appear to be correct? Yes / No / Unclear

d) Is there a way to feed back user comments to the app developer? Yes / No / Don't know

**3. Is there any evidence that the app does actually alleviate the problem?**

a) Have any studies been carried out to measure the impact of using the app on clinical or patient knowledge, actions or (preferably) patient outcomes? Yes / No / Don't know

b) Were these studies independently conducted, well designed, large enough, and applicable to the user? Not applicable / Yes / No / Don't know

c) Did any study also examine health resource use, potential harms caused by the app, or quantify cost effectiveness? Not applicable / Yes / No / Don't know

d) Overall, do the benefits of using this app seem likely to outweigh inconvenience and costs to the user? Yes / No / Don't know

e) Is there any specific clinical scenario or patient subgroup in which using the app seems particularly likely to be useful? Yes - Which? _____ / No / Unclear

# What is an “effective” digital health product?

UNIVERSITY OF Southampton

One that:

- Is *designed* to be effective ?
- The developers *believe* / state is effective ?
- That users *like & feel* is effective ?
- That users *state* in a survey is effective ?
- That users *continue to use* ?
- That evidence from studies *demonstrates* is effective ?

# Dimensions of effectiveness

UNIVERSITY OF Southampton

**Better patient experience** / quality of life

**Better clinical outcomes** eg. fewer complications, slower disease progression

**Lower usage of healthcare resources** with same clinical outcomes

**Incremental cost effectiveness** £20k per QALY or less

# How to measure if a digital health product is effective ?

UNIVERSITY OF Southampton

Within-person pre-post study: OK if disease stable, outcomes are easily measured & change fast (eg. asthma, diabetes) and no therapy is altered

Compare outcomes in two “similar” groups (control and intervention) - but how to **ensure** similarity:

- users vs. non-users ?
- patients last month vs. pts. this month ?
- alternate patients ?
- randomly allocated patients (Liu & Wyatt, JAMIA 2011)

# Factors likely to promote the uptake of digital health

UNIVERSITY OF Southampton

1. High quality products – functionality, flexibility, resilience, interoperability
2. Political will and leadership - funding
3. Incentives for professionals – direct benefits (EM Rogers), reimbursement…
4. Transparent market – certification, labelling
5. NICE or other national guidance based on evidence of effectiveness from studies about which patients & organisations benefit, and when

# Big challenge: study validity

- Sufficiently rigorous for the findings to be correct
- Include typical patients, outcomes, version of the product - so results are relevant to others
- Minimise role of the manufacturer / sponsor, to ensure others value & trust study results

Murray E et al. Design & evaluation of digital interventions. Am J Prev Med Nov 2016

# Some specific challenges

UNIVERSITY OF Southampton

Who to study:

- Volunteer effect

Measurement problems:

- Social response bias
- Digital health system collecting only outcome data

**Inference:**

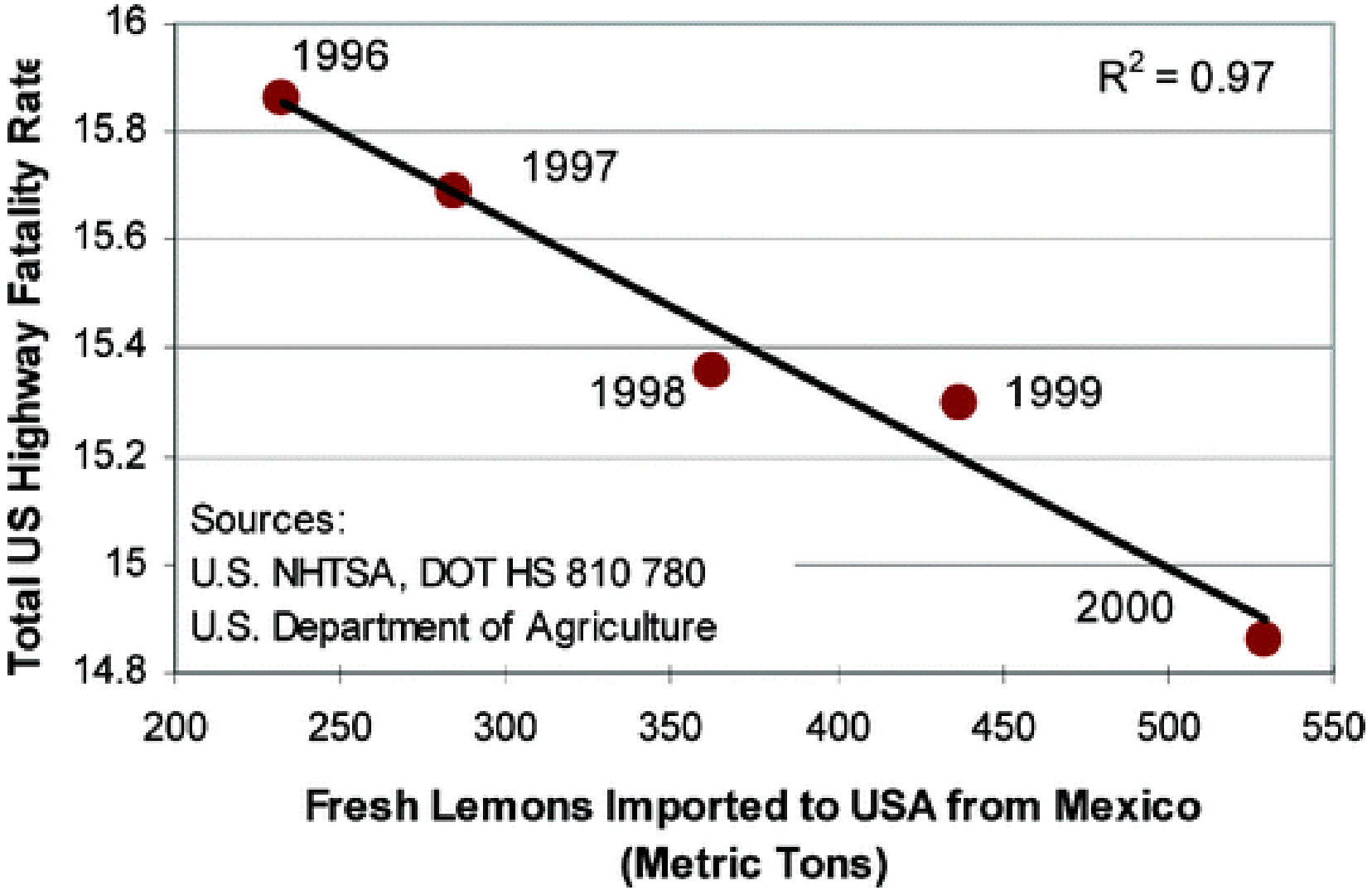
- **Association is not causation**
- **Regression to the mean**

**Confounders:**

- **Secular trends in before-after studies**
- Hawthorne and Checklist effects
- Simpson's paradox

# Do lemons from Florida cause US highway fatalities ?

UNIVERSITY OF Southampton




**Source:**
www.cqeacademy.com/cqe-body-of-knowledge/continuous-improvement/quality-control-tools/

# Association vs. causation: Rochester library study

UNIVERSITY OF Southampton

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

**Study method**: compare LOS in patients of Drs who used library often vs. patients whose Drs do not (case-control design)

**Result:** LOS significantly less in library-using Drs

**Interpretation:**

- Is library use the *cause* of reduced LOS ?
- Is library use a *marker* of doctors who keep patients in hospital less?
- Is library use the *result* of doctor keeping patients in hospital less ?!

**A better question**:

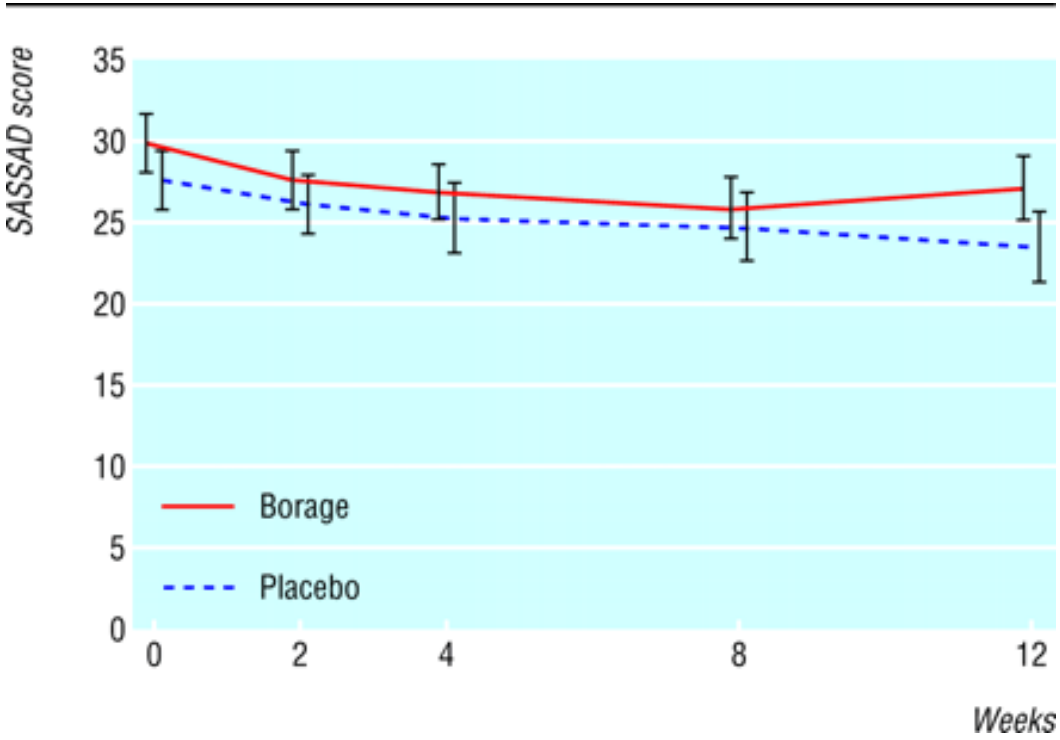
What is the impact on LOS of providing a sample of doctors with access to a library ?

# Regression to the mean

If you chose individuals with extreme values for a variable, it will be closer to the mean the next time you measure it

Happens because sample is not randomly selected

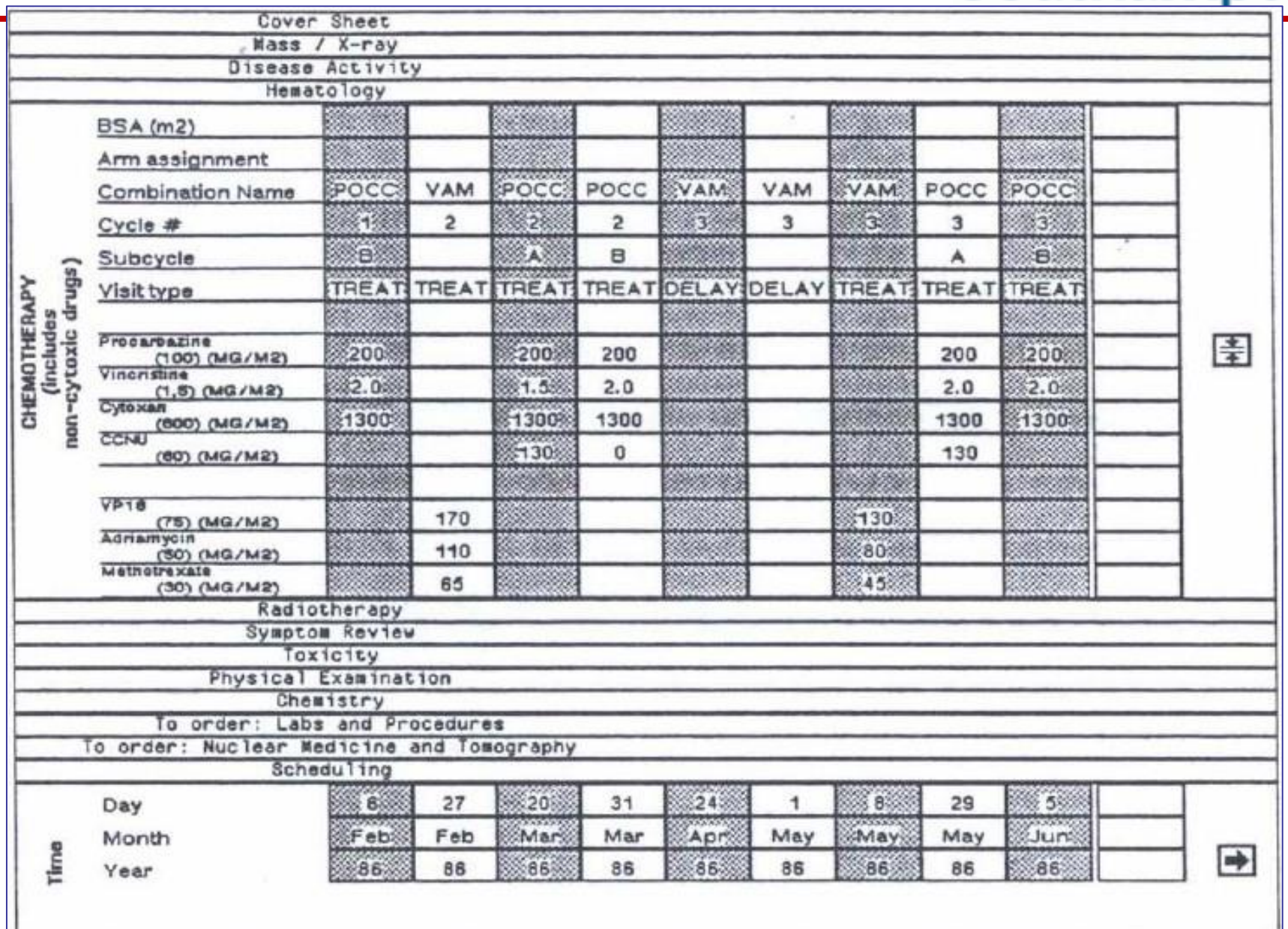
Beware of this in before-after studies !




Trial of borage oil in people with atopic eczema (Takwale et al., 2003). Example from Martin Bland, York University.

# Oncocin clinical workstation – Stanford 1980s

UNIVERSITY OF Southampton



Cover Sheet
Mass / X-ray
Disease Activity
Hematology

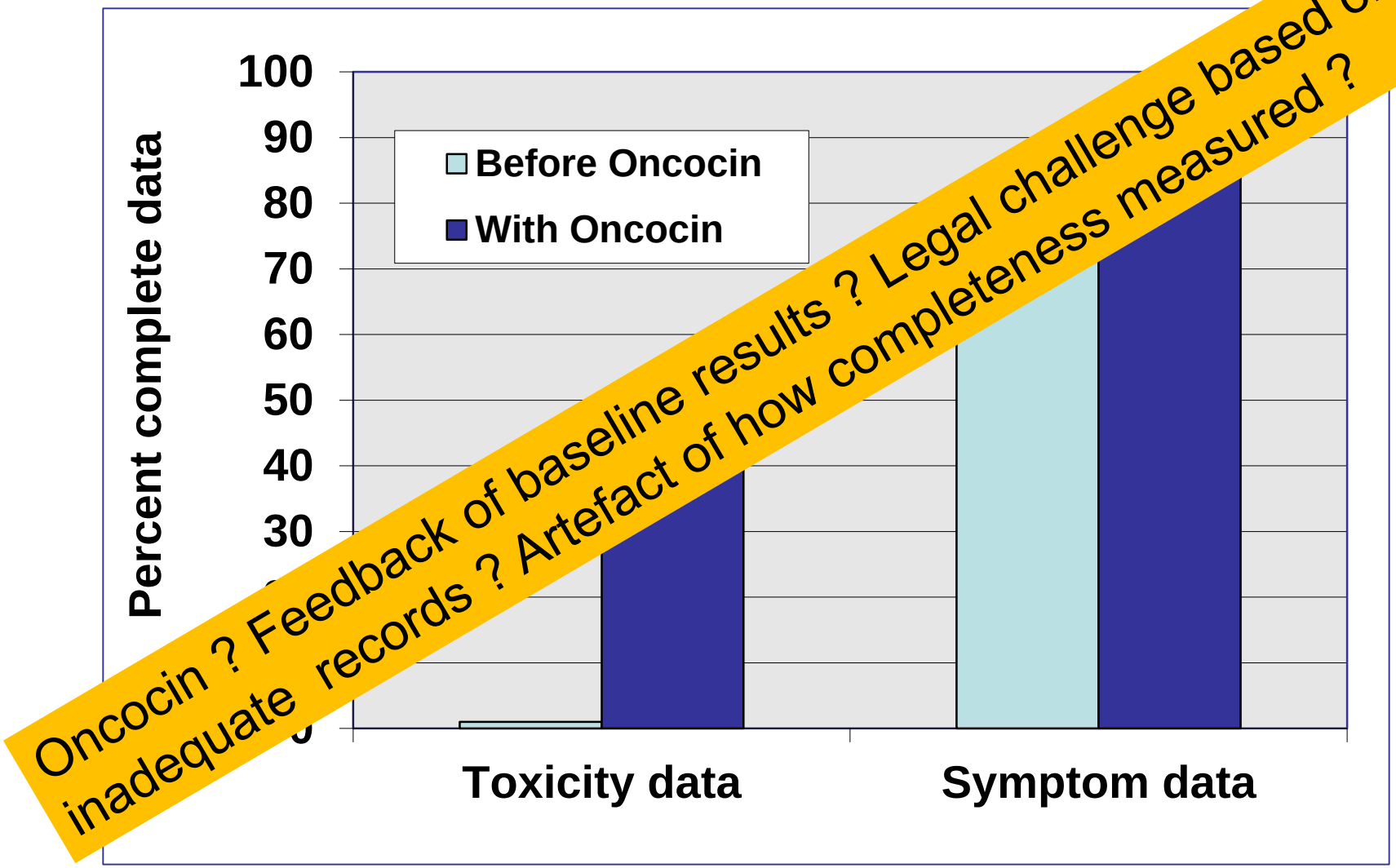
CHEMOTHERAPY (includes non-cytoxic drugs)

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| BSA (m2) | | | | | | | | | | |
| Arm assignment | | | | | | | | | | |
| Combination Name | POCC | VAM | POCC | POCC | VAM | VAM | VAM | POCC | POCC | |
| Cycle # | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | |
| Subcycle | B | | A | B | | | | A | B | |
| Visit type | TREAT | TREAT | TREAT | TREAT | DELAY | DELAY | TREAT | TREAT | TREAT | |
| | | | | | | | | | | |
| Procarbazine (100) (MG/M2) | 200 | | 200 | 200 | | | | 200 | 200 | |
| Vincristine (1.5) (MG/M2) | 2.0 | | 1.5 | 2.0 | | | | 2.0 | 2.0 | |
| Cytoxan (600) (MG/M2) | 1300 | | 1300 | 1300 | | | | 1300 | 1300 | |
| CCNU (60) (MG/M2) | | | 130 | 0 | | | | 130 | | |
| | | | | | | | | | | |
| VP16 (75) (MG/M2) | | 170 | | | | | 130 | | | |
| Adriamycin (30) (MG/M2) | | 110 | | | | | 80 | | | |
| Methotrexate (30) (MG/M2) | | 65 | | | | | 45 | | | |

Radiotherapy
Symptom Review
Toxicity
Physical Examination
Chemistry
To order: Labs and Procedures
To order: Nuclear Medicine and Tomography
Scheduling

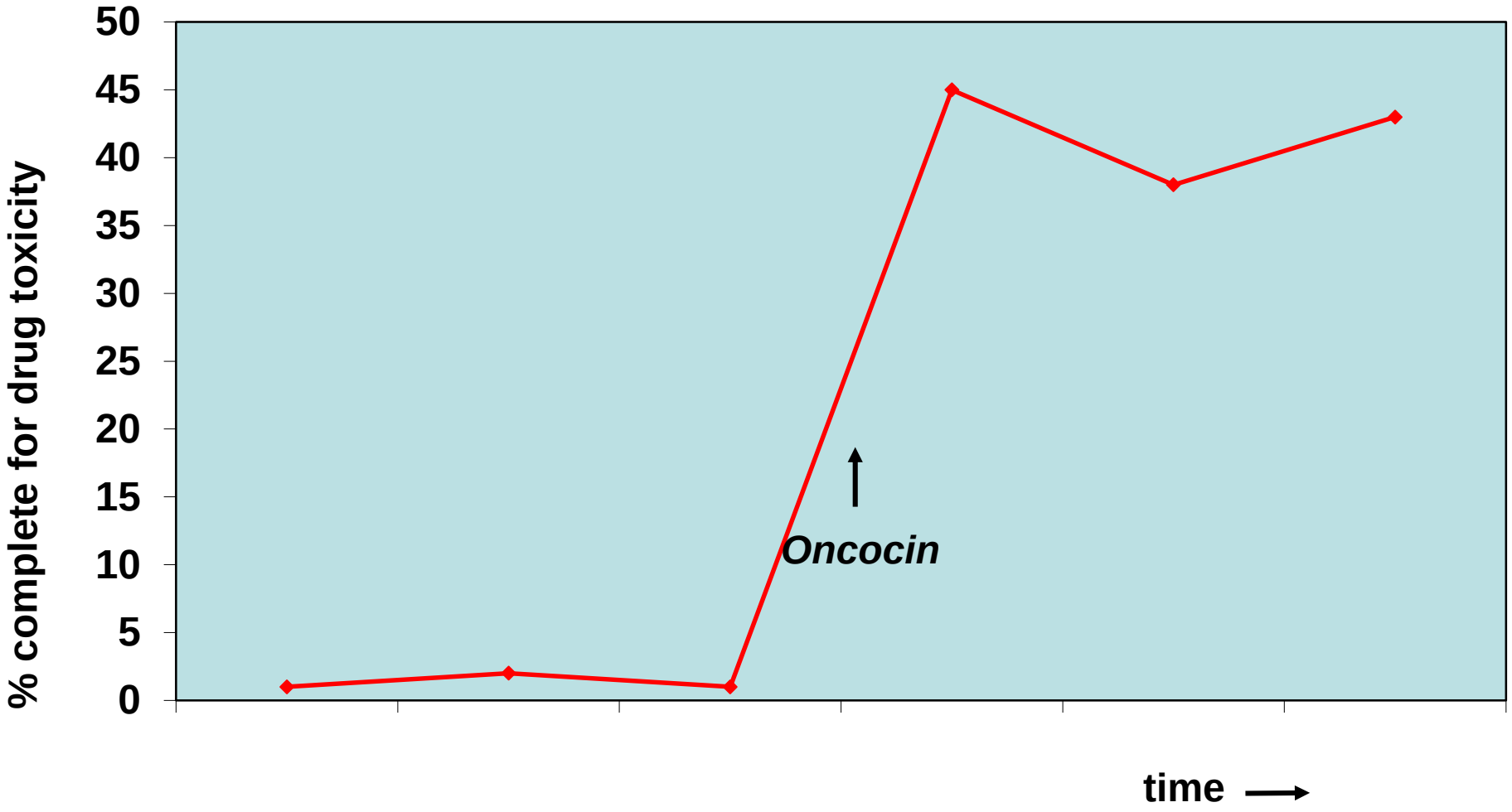
| Time | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Day | 6 | 27 | 20 | 31 | 24 | 1 | 8 | 29 | 5 | |
| Month | Feb | Feb | Mar | Mar | Apr | May | May | May | Jun | |
| Year | 86 | 86 | 86 | 86 | 86 | 86 | 86 | 86 | 86 | |

# Did Oncocin improve data quality ?

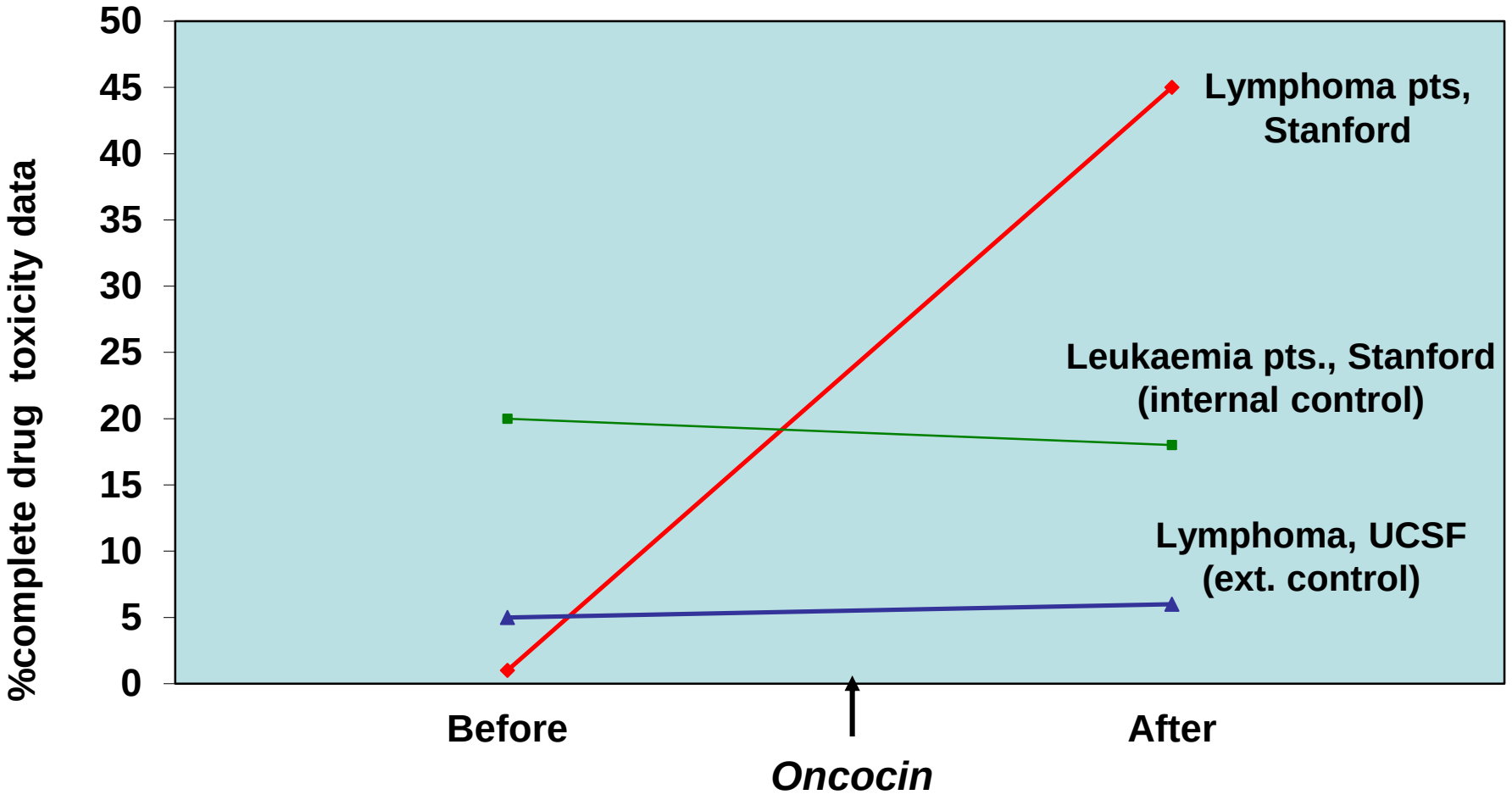



Kent & Shortliffe 1986

# Interrupted time series study design



# Controlled before-after design



# PhD project: RCT of GP teledermatology to prevent unnecessary referrals in 560 patients

University of Southampton


**Results: nearly 1/3 of dermatology clinic visits could have been prevented**

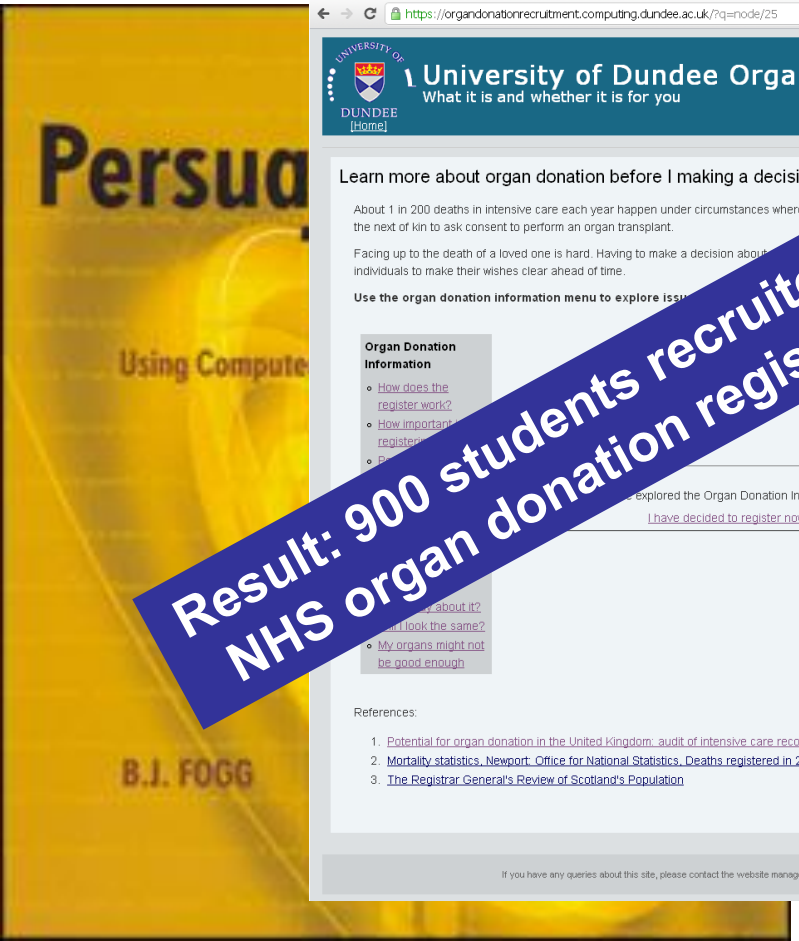
Eminovic et al. . Arch Dermatol. 2009;145(5):558-64.

***With Depts. of Medical Informatics and Primary Care, AMC Amsterdam***

# RCT of website design - does Fogg's theory help persuade people to donate organs for transplant?

UNIVERSITY OF Southampton

Work of Thomas Nind, PhD Student, Dundee





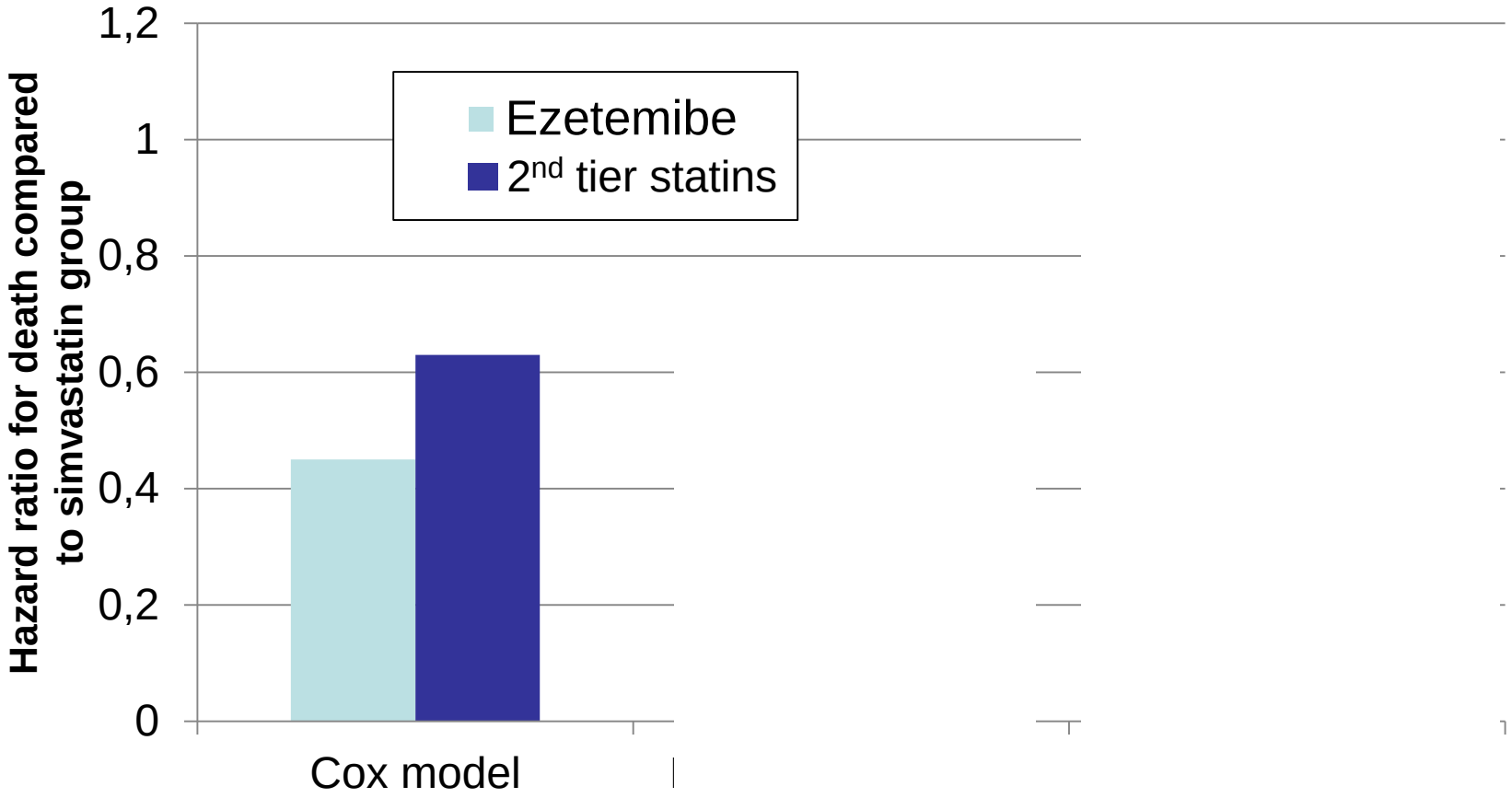


**Result: 900 students recruited to RCT in 5 days; no difference in NHS organ donation register sign-up rates (38% both groups)**

**Persuasive features:**

1. URL includes https, dundee.ac.uk
2. University Logo
3. No advertising
4. References
5. Address & contact details
6. Privacy Statement
7. Articles all dated
8. Site certified (W3C / Health on Net)

# RWE: bias estimating ezetimibe impact on mortality in 2233 post-MI deaths using CPRD




*Eg. First MI; missing cholesterols; medication covariates, immortal time bias…*

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

# Mismatch between routine self report and objective data

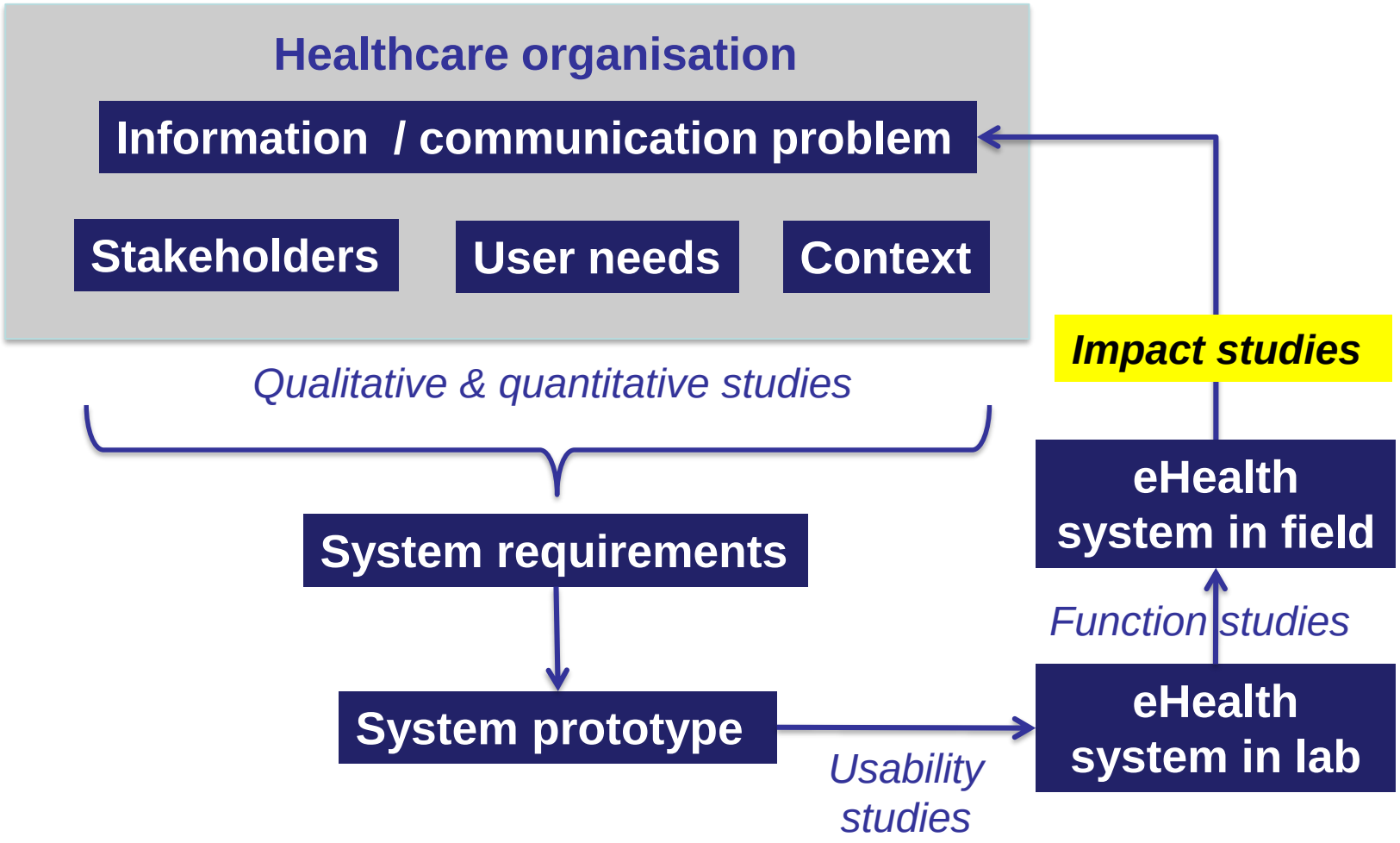
UNIVERSITY OF Southampton

Randomised trial of Text2Quit SMS programme in 503 US adults:

- **Self reported** cessation at 6 months: 20% in SMS group, 10% control group (effect size 2, NNT 10)
- **Biochemically confirmed** smoking cessation (saliva cotinine levels) at 6 months: 11% SMS group, 5% control group (effect size 2, NNT 18)
- Possible explanation: social response bias

Abroms et al, Am J Prev Med 2014

# Evaluation studies in the digital health product lifecycle



# Big Health Data (or “Real World Evidence”) & evaluation

# Possible responses

UNIVERSITY OF Southampton

Yes, this is the intended benefit

- Oncocin required data before doctor could prescribe, other toxicity data entered from lab reports

No, it's an artefact of measurement methods

- Easier to check if data complete in database than paper record
- Definition of "complete data" changed (for paper records, no mention = no toxicity present)

No, it's an indirect impact via changes in staff

- New staff coincided with introduction of Oncocin
- Hawthorne effect, stimulated by presence of Oncocin in clinic
- Feedback of baseline results raised motivation

Numerous other possible explanations:

- Legal case, poor data quality, letter from chief executive
- New, toxic drug introduced
- Chance effect: small numbers…

# Asthmopolis

UNIVERSITY OF Southampton

# Confounding by indication

UNIVERSITY OF Southampton

40% of cancer patients treated with new drug [illegible] 5 years versus 30% of patients treated with old [illegible]

Difference persist despite taking acco[illegible] differences in age, baseline cancer severity, [illegible] markers…

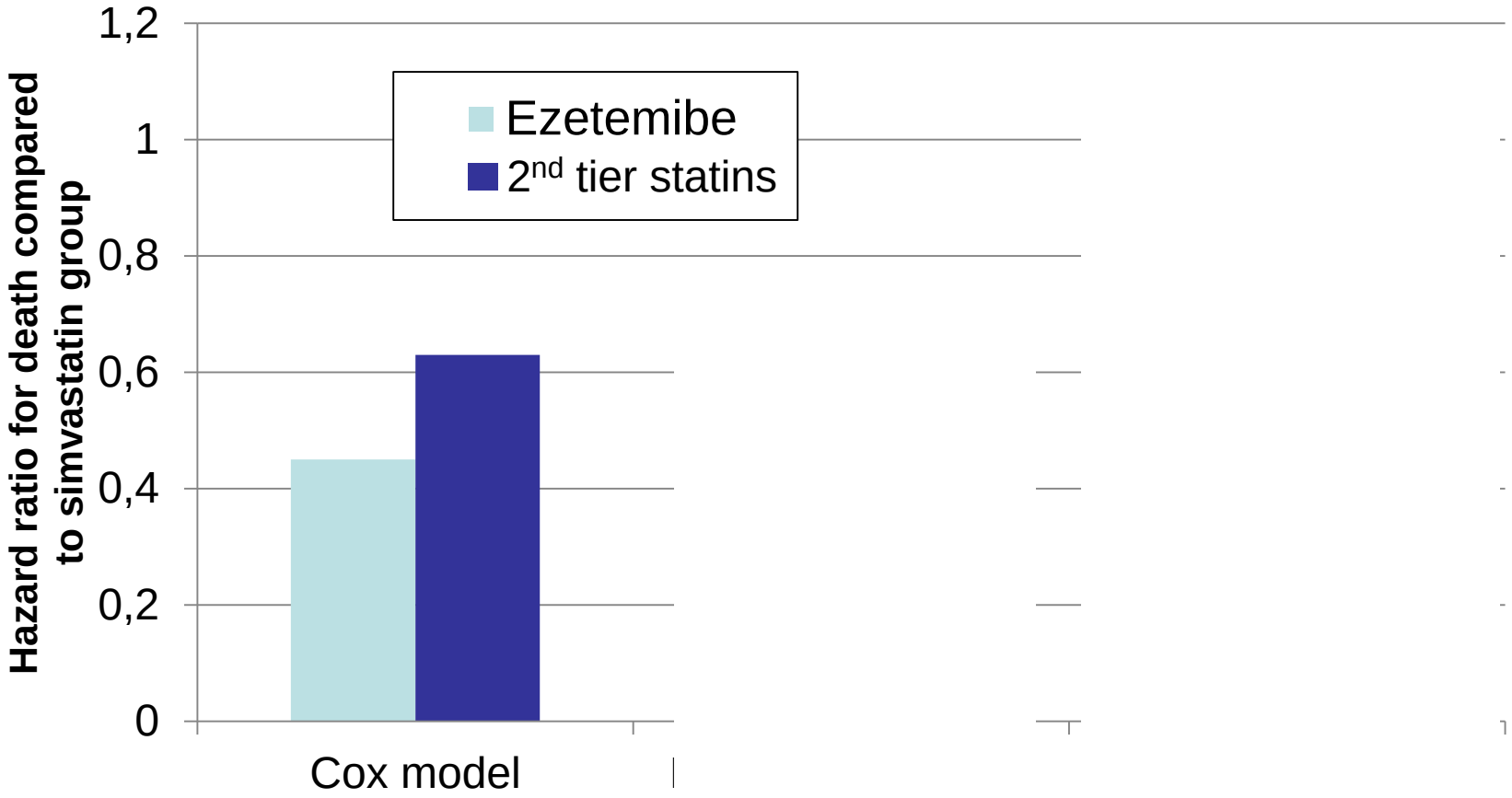
Conclusion: the new drug re[illegible] mortality by 10%

But ***maybe*** allocation to [illegible]ew drug depends on the doctor's intuition on who will survive (subtle predictive feature not recorded in a[illegible]tabase)

So, receipt [illegible] new drug is a ***marker*** of better outcome - not t[illegible]use

**Propensity scoring as a potential solution to this**

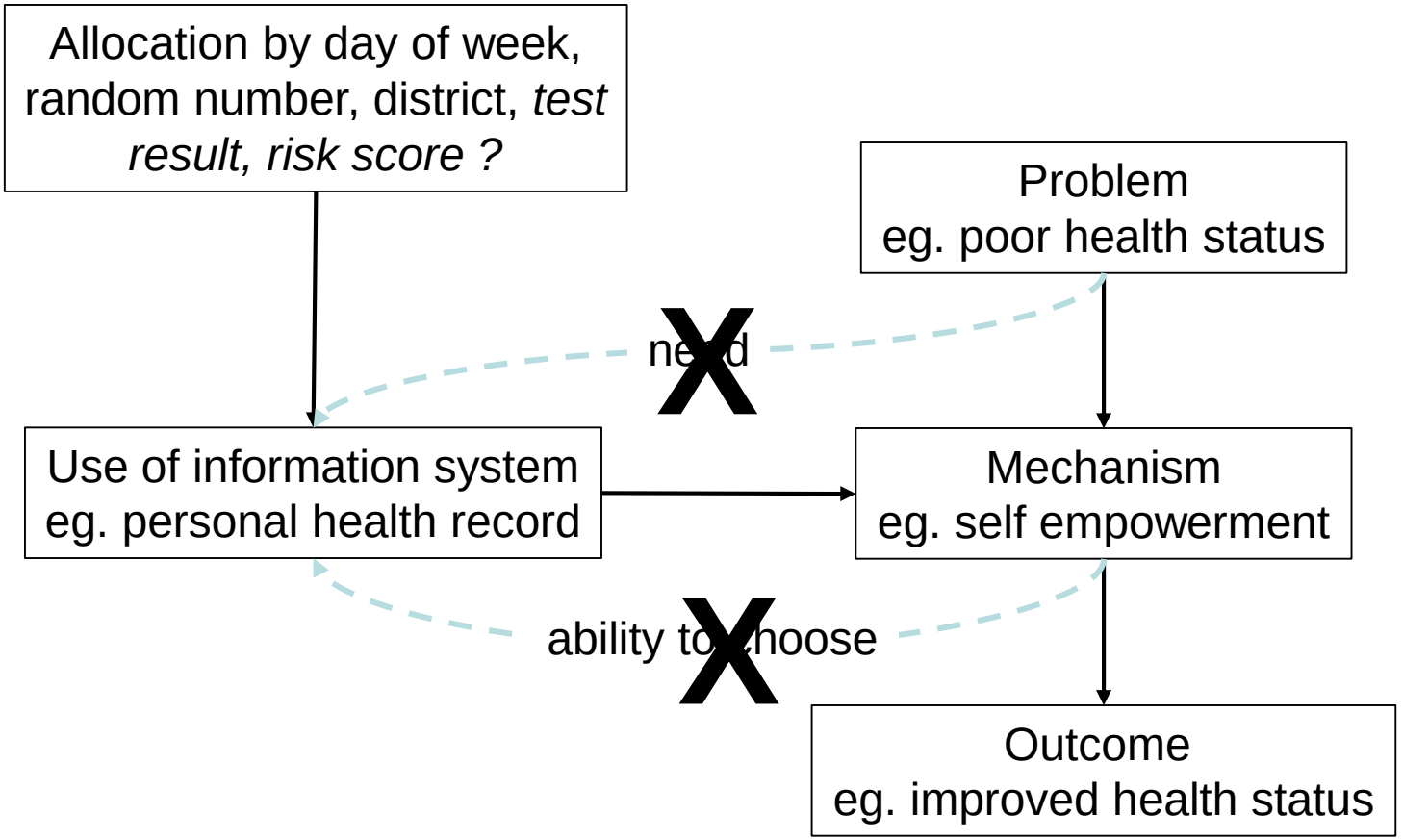
# The impact of bias on estimating mortality for ezetimibe in 2233 post-MI deaths (all cause mortality)




*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

# Graphical models



# Regression discontinuity design

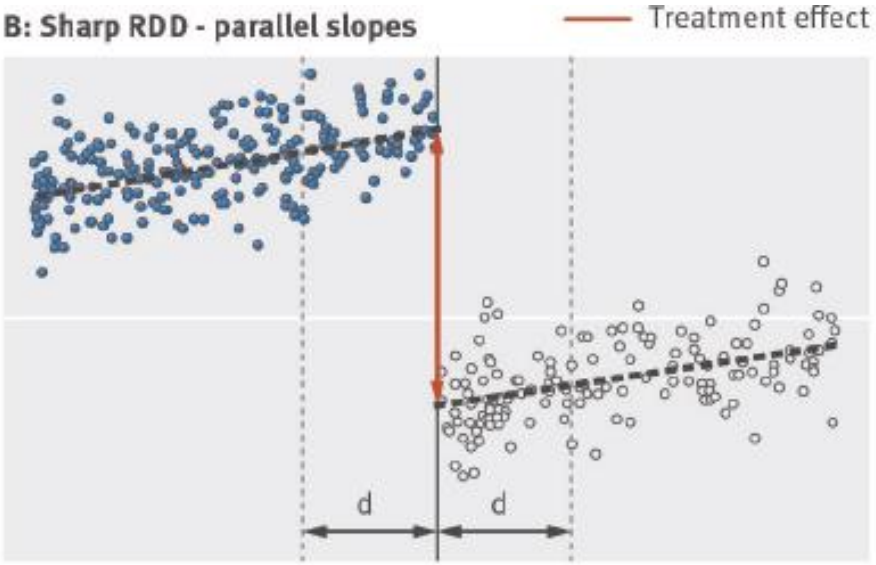
Some drugs / procedures are applied according to a test result or predictive model

People just above & just below an allocation threshold are very similar

If you have enough people to compare, you can *estimate* the impact of the intervention

Eg. chemotherapy on older women – RCT failed to recruit

Thistlethwaite & Campbell, 1960



# Scenarios when RDD may be useful

UNIVERSITY OF Southampton

- When routine data are available
- Treatment has already become established
- “Randomisation is unethical”
- Rare diseases with reluctance to refer to single centre
- When RCTs recruit unrepresentative samples

# Some concerns about the "data bit"

UNIVERSITY OF Southampton

Can we re-use data captured for one purpose to inform another ? (1st law of MI - Johan van der Lei, Lancet 1991)

"Anonymisation" and privacy protection – EU GDPR 2018 – 700,000 opt outs from Care.data in UK…

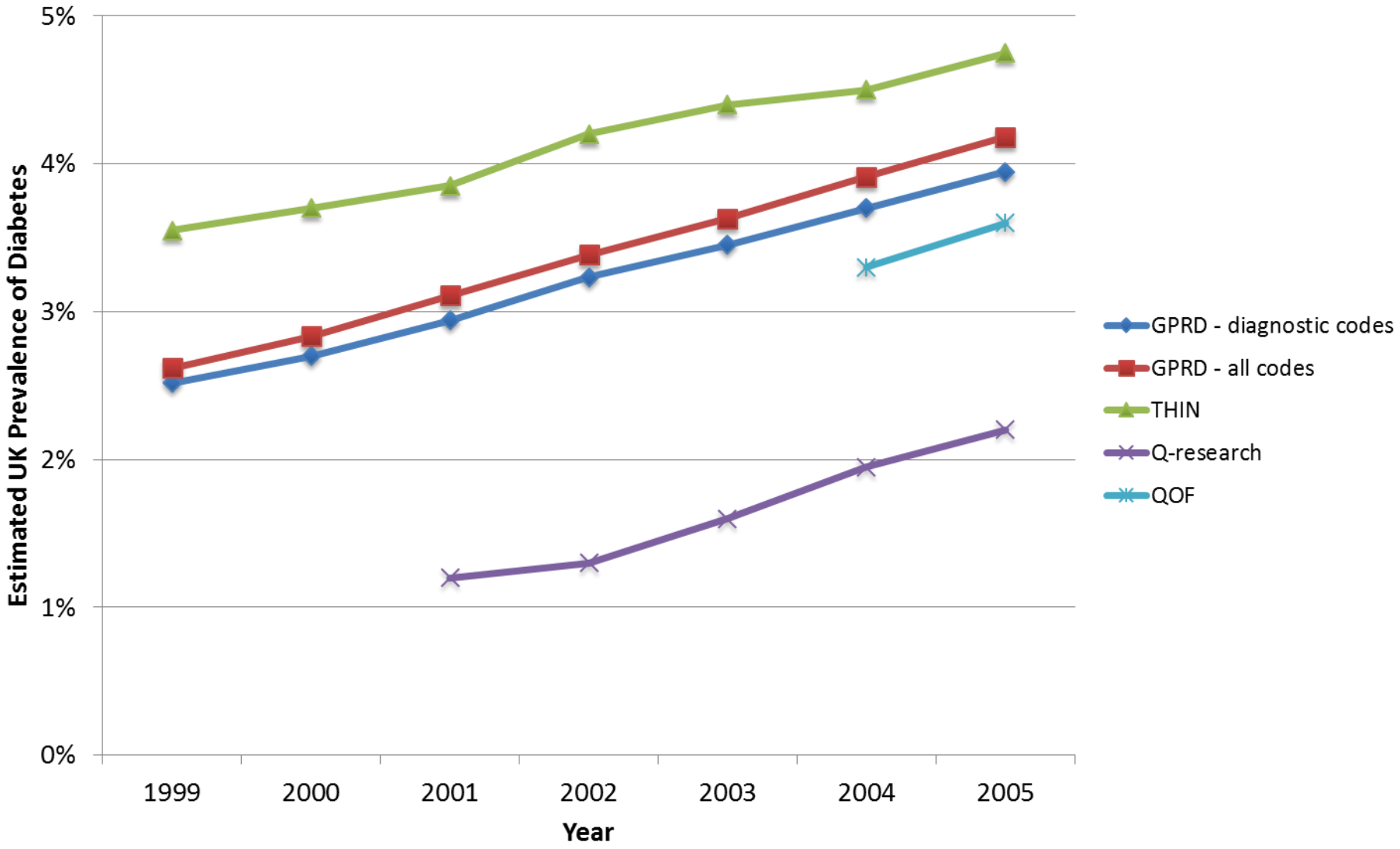
Differing usage of common clinical codes in general practice [eg. circa 45 codes to find asthma patients - www.clinicalcodes.org]

Variable / poor quality of routine health system data

Is our clinical data ontology sufficiently robust to drive a clinical data semantic web / SOA ? [the bioinformatics ontology is]

# Diabetes prevalence depends on which database you check, & how

UNIVERSITY OF Southampton




*Slide from Niels Peek*

# Potential advantages of RWE

UNIVERSITY OF Southampton

More power so can examine subgroups: datasets 100-1000 times larger than for RCTs

Can answer more questions - wider variety of data

More representative - data captured from routine care, cf. studies

Quicker and cheaper to answer questions: use existing datasets

Can use data-driven quality improvement to build continuous, rapid learning cycles – “Learning Health System”

Sherman et al – FDA view on RWE - NEJMed 2016

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

# Potential disadvantages of RWE

UNIVERSITY OF Southampton

- Poor quality data due to misclassification bias (poor co-morbidity coding; identity or linkage problems) – so exclude many cases, or do manual case note review
- Patient-relevant outcome data eg. PROMs missing
- Cost of data capture systems is high (Addenbrookes' EPIC EPR: £200M + drop in CQC rating)
- Hidden biases, eg. confounding by indication
- Unclear inclusion criteria – datasets may be limited by differing case definition, geography, income, education…
- Temptation to data dredge & all associations stat. significant leads to frequent false positive results
- Worse publication bias than RCTs

Byar, Why databases should not replace trials, Biometrics 1980

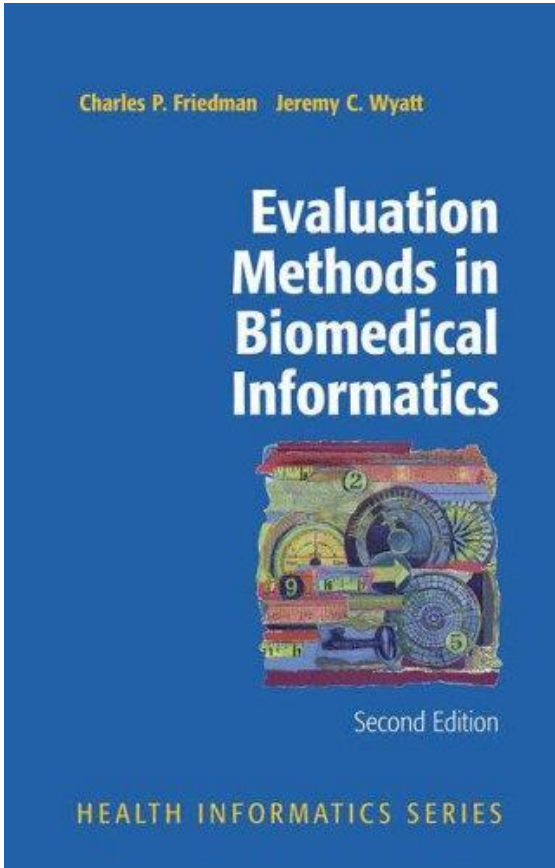
Lars Hemkens, Ioannidis et al. CMAJ 2016

# Good evaluation practice for eHealth interventions

UNIVERSITY OF Southampton

1. Know why you are evaluating: who are the stakeholders, what decision do they face ?
2. Understand stakeholder questions and the level of evidence they need to answer them
3. Design your impact study with:
   - Enough participants of the right kind
   - The right intervention
   - The right control
   - Validated outcome measures
4. Check for biases and confounders, that you will learn something if study is negative
5. Run the study & report your results

See: Murray E et al. Design & evaluation of digital interventions. Am J Prev Med Nov 2016



# Intervening in the app lifecycle

| Stage in app lifecycle | Stakeholder | Quality improvement process | Example |
|---|---|---|---|
| Development | Developer | Involve clinicians / experts<br>Refer to engineering standards<br>Understand quality criteria<br>Develop & evaluate app using appropriate framework | BSI app standard PAS 277<br>HON code, RCP checklist<br>13 questions (Murray, 2016) |
| Uploading to app store | App store | Check technical aspects<br>Check privacy<br>Check developer qualifications | iPhone store excludes drug-related apps unless developer is product licence holder |
| App rating | Raters | Wisdom of the crowd<br>Use explicit criteria | Can fail: Abroms 2014<br>RCP checklist |
| Selection from the app store | User | Consider risks<br>Check reviews<br>Check quality<br>Check CE mark, intended user, training needed etc. | Risk checklist<br>iMedicalApps<br>RCP checklist, CE mark<br>Euroseal label (Rigby 2003) |
| Usage for self management | User | Notify regulator of errors, near misses | RCP guidance 2014 |
| Removal from app store | Regulators | Respond to adverse events, lack of data to support claims | Acne apps |