Institute and Faculty of Actuaries

University of East Anglia

Obtaining public data for public purposes

Prof Elena Kulinskaya (UEA)

Content

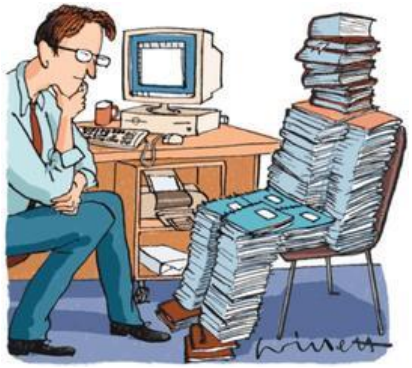
- What are the sources of public data
- What other data would be useful
- Can private companies help public research?

Institute and Faculty of Actuaries

Data Sources: Randomised Clinical Trials and Meta-analyses of RCTs

- Randomised clinical trials are the ‘gold standard’ for evidence of effectiveness of health interventions
  - Confounders randomized equally to both groups (in theory)
- Generalisability from trial participants to general population
  - Exclusion on grounds of age, comorbidity, intolerance to intervention
- Short follow-up (Maximum 5 years)
- Commercial trial data not available for individual scrutiny
  - Lack of transparency
- Large observational datasets can fill these gaps with robust statistical analyses

Data Sources: Observational Data



Administrative Data

- Primary care CPRD and THIN
- Hospital Data HES
- Social care data: care and nursing homes, older people in the community, etc.

Specialised registries

- Cancer Registries
- National Joint Registry

Institute and Faculty of Actuaries

Survey and Census data

- Health Surveys for England, Wales, Scotland and Northern Ireland (yearly) A total of 8,795 adults and 2,185 children were interviewed in 2013.
- Census Data (every 10 years from 1971, 1% of the population of England and Wales, together with records for other people in their households.
- The ONS Longitudinal Study (LS) is a linked set of individual census responses. Results from the censuses of 1971, 1981, 1991, 2001 and 2011 and life event information. Linked to health data, Cancer Registry, http://calls.ac.uk/
- The Health and Social Care Information Centre (HSCIC) commissions and holds many data collections, and is supposed to provide linkages.

Longitudinal Cohort Studies

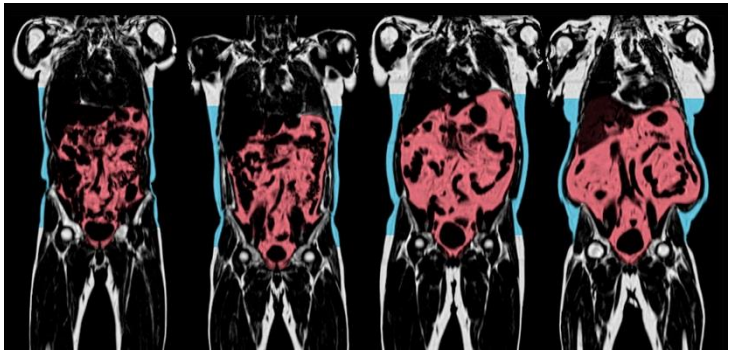
- An estimated 3.5 % of the population have taken part in one or more UK cohort studies (MRC, 2014).
- **Birth cohorts**
- MRC National Survey of Health and Development, a representative sample (N=5362) of men and women born in England, Scotland or Wales in March 1946
- Aberdeen Children Of the Nineteen Fifties (N=12,150) [all Aberdeen primary school children born between 1950 and 1956]
- The 1970 British Cohort Study (BCS70) (N=17,000) follows the lives of more than 17,000 people born in England, Scotland and Wales in a single week of 1970

Representative Population Studies

- English Longitudinal Study of Ageing (ELSA): Waves 0-7, 1998-2015 (over 50+ in 2002, N=12,000, adding extra samples every 2 years)
- Cognitive Function and Ageing Studies (CFAS(N=7635), CFAS II(N=7762))
- European Prospective Investigation of Cancer (EPIC) (N=30,000 in Norfolk, recruitment 1993-97, 40-79 yo, longitudinal information on diet+health)
- Dementias Platform combines 30 cohort studies

UK Biobank

- 500,000 volunteers, aged 40-69, recruited 2006-2010
- biological samples
- detailed demographic and lifestyle
- MRI and DEXA scanning, n=10,000
- Diet questionnaire longitudinally
- Genetic information
- linked to ONS, Cancer Register, HES data, diagnostic imaging and primary care data by 2017.
- At the moment, there are about 10,000 deaths among the participants
- Self-selected sample, so may be biased.



Institute and Faculty of Actuaries

Use of Biobank Data for longevity research

nature COMMUNICATIONS

ARTICLE

Received 28 Sep 2015 | Accepted 29 Feb 2016 | Published 31 Mar 2016

DOI: 10.1038/ncomms11174 OPEN

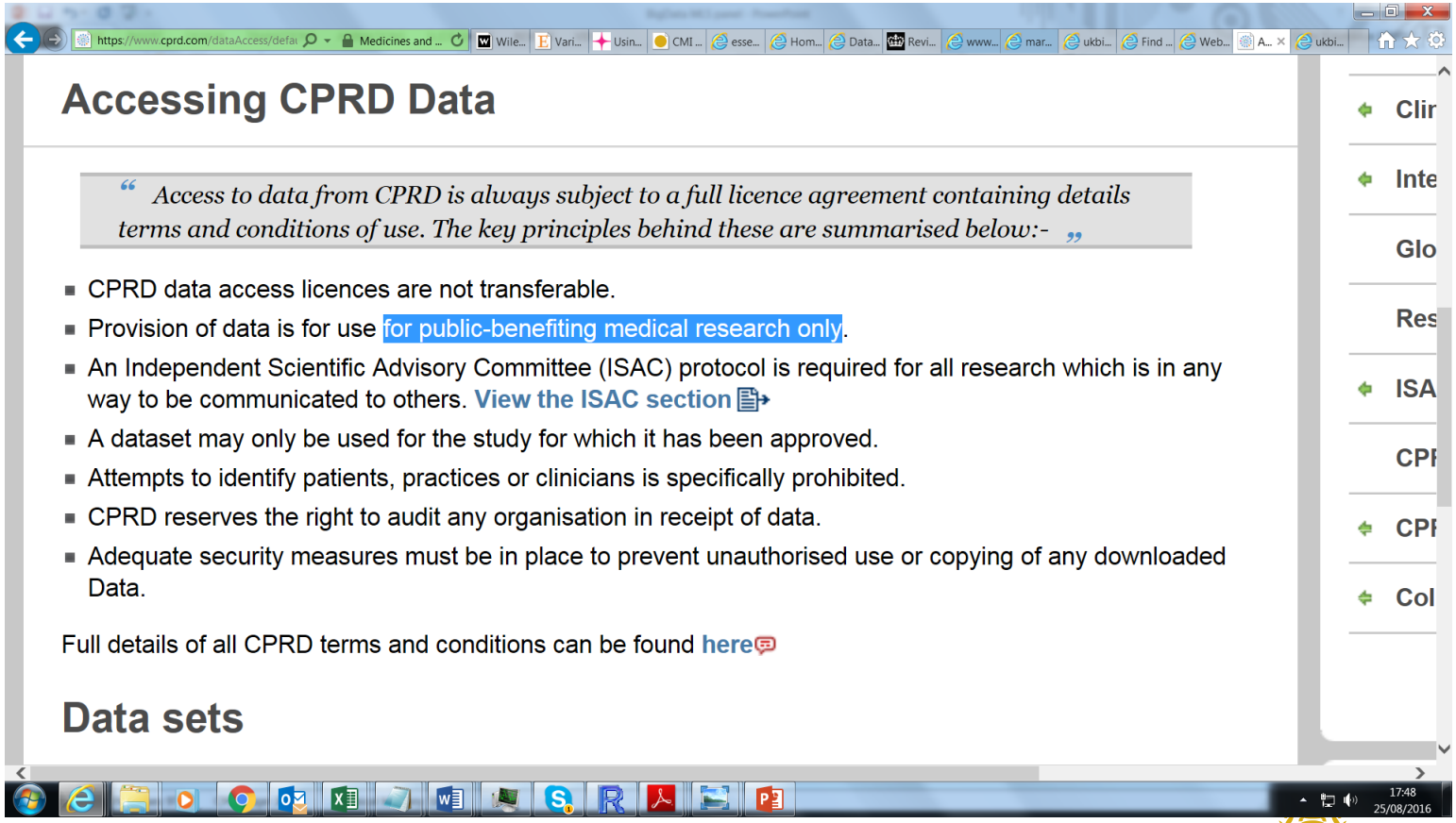
Variants near *CHRNA3/5* and *APOE* have age- and sex-related effects on human lifespan

Peter K. Joshi[1], Krista Fischer[2], Katharina E. Schraut[1,3], Harry Campbell[1], Tõnu Esko[2,4,5,6] & James F. Wilson[1,7]

Lifespan is a trait of enormous personal interest. Research into the biological basis of human lifespan, however, is hampered by the long time to death. Using a novel approach of regressing (272,081) parental lifespans beyond age 40 years on participant genotype in a new large data set (UK Biobank), we here show that common variants near the apolipoprotein E and nicotinic acetylcholine receptor subunit alpha 5 genes are associated with lifespan. The effects are strongly sex and age dependent, with *APOE* ε4 differentially influencing maternal lifespan ($P = 4.2 \times 10^{-15}$, effect −1.24 years of maternal life per imputed risk allele in parent; sex difference, $P = 0.011$), and a locus near *CHRNA3/5* differentially affecting paternal lifespan ($P = 4.8 \times 10^{-11}$, effect −0.86 years per allele; sex difference $P = 0.075$). Rare homozygous carriers of the risk alleles at both loci are predicted to have 3.3–3.7 years shorter lives.

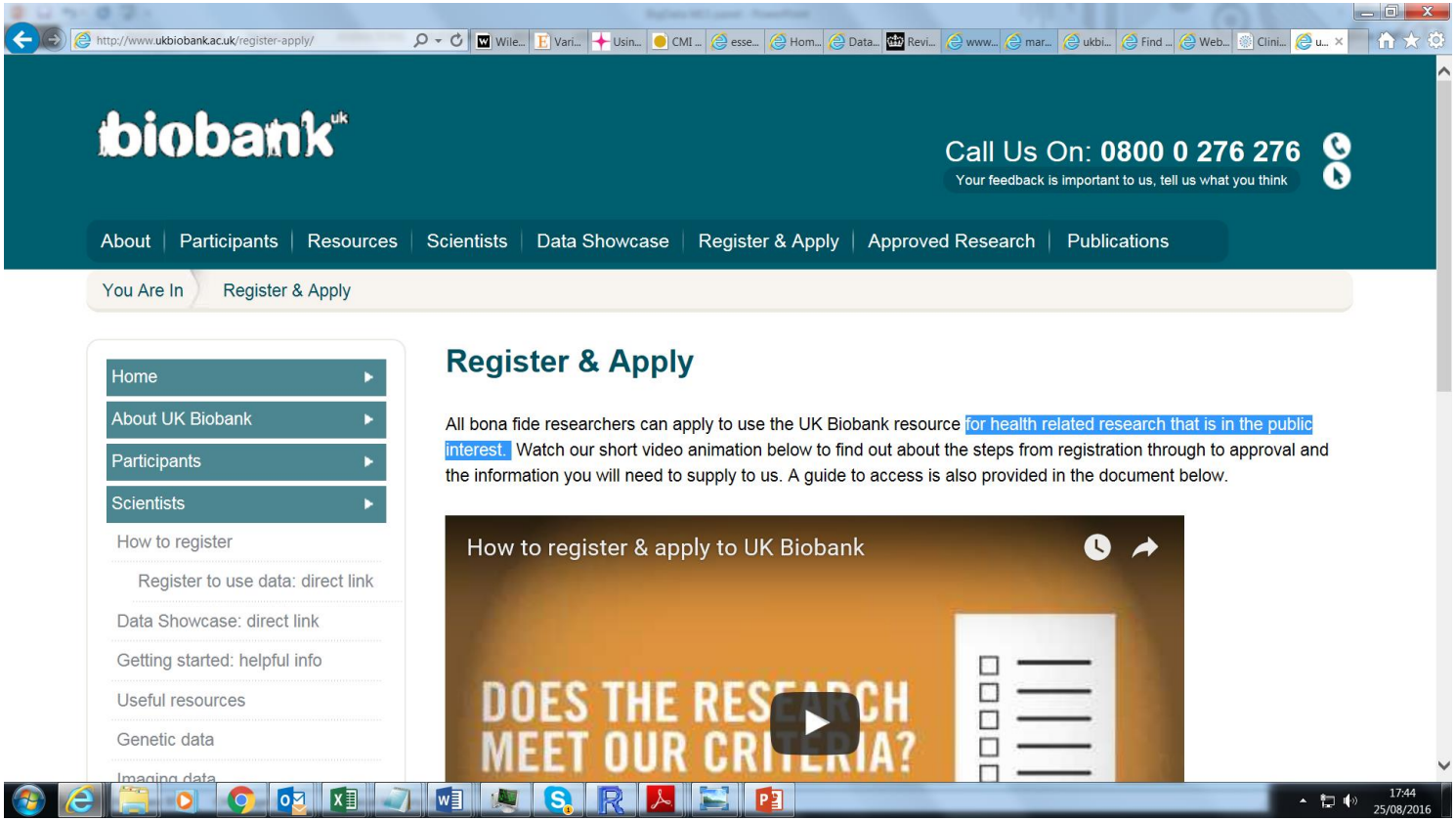
Institute and Faculty of Actuaries

Difficulties: Health Research only



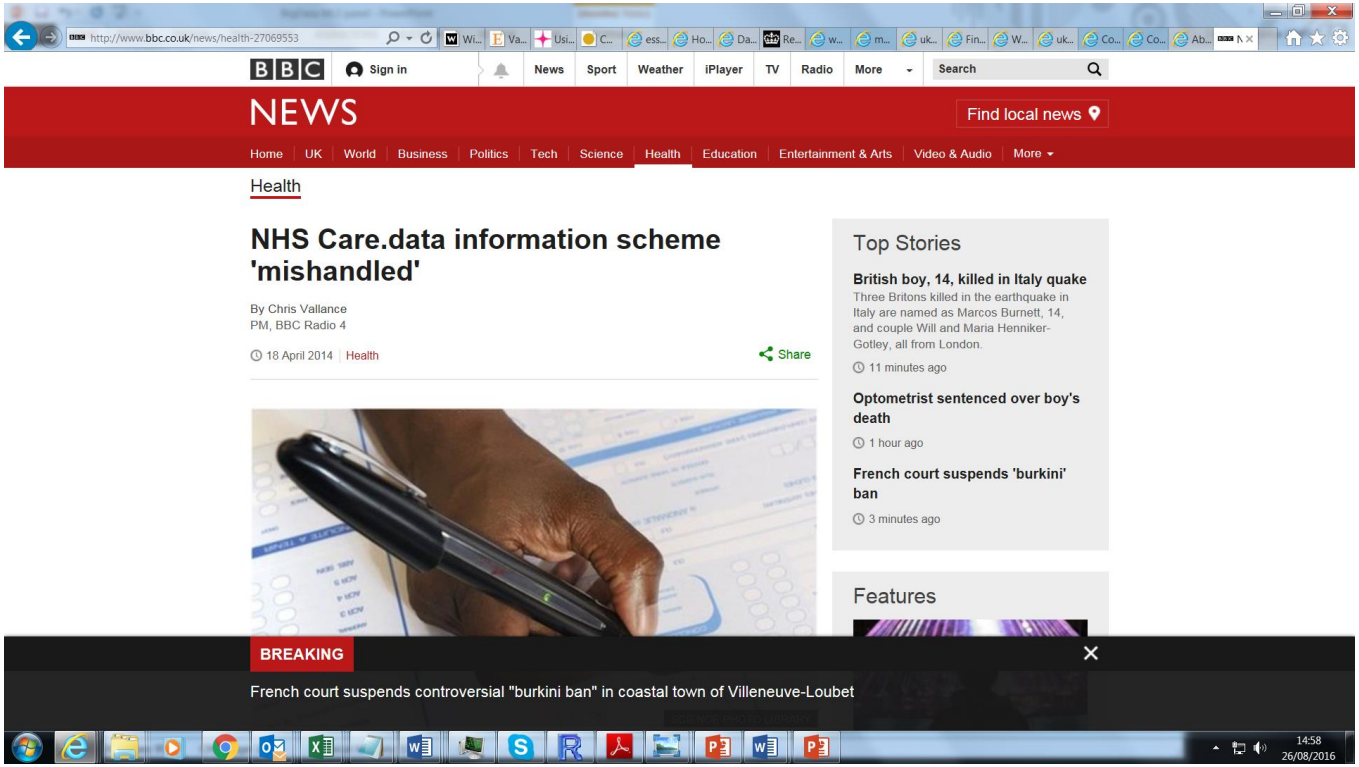
Institute and Faculty of Actuaries

Biobank: not available for actuarial research?



Institute and Faculty of Actuaries

Difficulties: confidentiality, data sharing, data linkage



"The National Data Guardian proposes a New consent / opt-out model for consultation to enable people to opt out from their personal confidential data being used for purposes beyond their direct care."

Institute and Faculty of Actuaries

ESRC Business and Local Government Data Research Centre (BLG DRC)

- Funded under the ESRC's Big Data Network, £5m, 2014-19
  - An Eastern ARC (Essex, UEA & Kent) partnership
- Exploitation of data to benefit researchers, data owners and society
- Highest ethical standards, anonymised data used non-disclosively
- Data can be safely accommodated at the UK Data Archive at Essex, and safe access is possible in different formats, including through the safe rooms at UEA and Essex.
- Discussions with potential data providers to build trust and refine research questions of common interest
- Some nervousness about data sharing although route seems to be through well-specified research projects, subject to normal Research Governance arrangements
- Different LAs hold different data, in different ways
- Lack of a common approach to linking health and social care data

ESRC Business and Local Government Data Research Centre

University of Essex

UEA University of East Anglia

University of Kent

LSE

Institute and Faculty of Actuaries

Other sources of data

http://www.dailymail.co.uk/sciencetech/article-2838025/Fitbit-

Wiley: Methodological Deve... | Variance stabilizing transfor... | Using Big Health and Actuar... | Inside the bizarre French co... | Fitbit data is now being ...

Cookie Policy | Feedback

Mail Online — Science & Tech

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Money | Video | Travel | Fashion Finder

Latest Headlines | Science | Pictures — Login

Fitbit data is now being used in COURT: Wearable technology is set to revolutionise personal injury and accident claims

- Law firm in Calgary, Alberta is using a Fitbit to show effects of an accident
- Device collects steps taken, sleep patterns and other data about the victim
- Analytics firm Vivametrica has launched service to compare the victims' activity with the national average to determine the strength of a claim
- Data will be used alongside other evidence, and evaluated by a doctor

By SARAH GRIFFITHS FOR MAILONLINE
PUBLISHED: 16:56, 17 November 2014 | **UPDATED:** 16:57, 17 November 2014

Share — 149 shares — 1 View comments

In personal injury lawsuits, it can be notoriously difficult to establish whether a

Site / Web — Enter your search — Search

Like Daily Mail | Follow @dailymailtech | Follow Daily Mail | +1 Daily Mail | Download our iPhone app | Download our Android app

Today's headlines | Most Read

- The rare, deep-Earth tremor that shook the planet: Waves from north Atlantic 'weather bomb' are measured in...
- CIA reveals Spacenet 'AI in the sky' that could constantly monitor activity on Earth via high resolution...
- The world's largest pyramid is hidden inside a mountain in Mexico: Early settlers built a church on top of...
- Does YOUR iPhone 6 have 'Touch Disease'? A design defect is rendering devices useless - and Apple can't fix...
- How did that get there? Divers find a century-old STEAM TRAIN at the bottom of Lake Superior
- Can being vegan make you go EXTINCT?

17:25 26/08/2016

Institute and Faculty of Actuaries

Google Flu example

"Big data hubris" is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis.

Laser et al, *Science*, 2014

Why Google Flu Is A Failure

Steven Salzberg, CONTRIBUTOR

Fighting Pseudoscience FULL BIO

Opinions expressed by Forbes Contributors are their own.

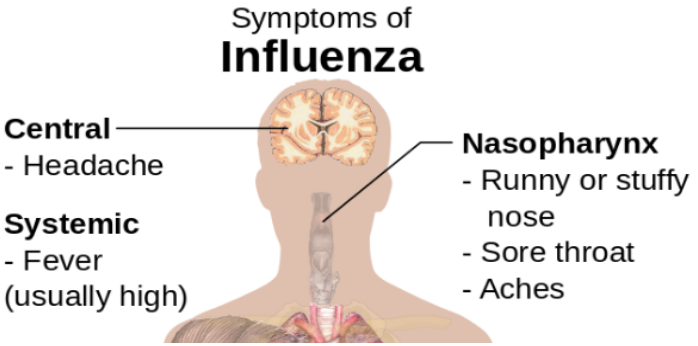
It seemed like such a good idea at the time.

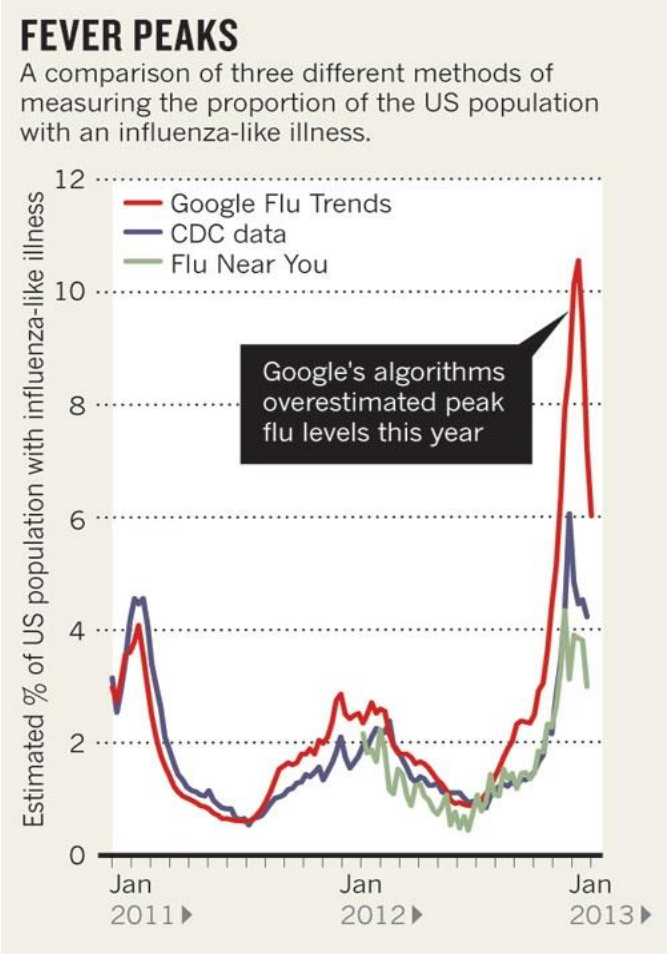
People with the flu (the influenza virus, that is) will probably go online to find out how to treat it, or to search for other information about the flu. So Google GOOG +0.58% decided to track such behavior, hoping it might be able to predict flu outbreaks even faster than traditional health authorities such as the Centers for Disease Control (CDC).

Instead, as the authors of a new article in *Science* explain, we got "big data hubris." David Lazer and colleagues explain that:

> "Big data hubris" is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis.

The folks at Google figured that, with all their massive data, they could outsmart anyone.





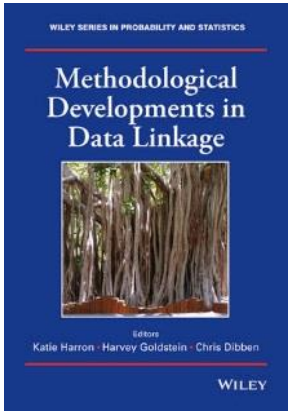
Institute and Faculty of Actuaries

Statistical Problems in Big Data

- Big Data is observational data!
- Methodology of observational data analysis/ meta-analysis:
- False positives arising from multiple exploratory analyses
- Biases due to peculiarities of units, outcomes or settings
- Missing data
- Inadequate linkage strategies
- Data gathered at varied levels such as transaction, person, organization, community, and state
- Causal Inference from (mostly) correlational data
- Modelling heterogeneity

Institute and Faculty of Actuaries

Need for linkages between private and public data



- **Basis risk and Longevity**
- Longevity and morbidity risks are evaluated on the general population data.
- CMI data, link to primary care data, ONS and to Mosaic codes!
- **Does private medical insurance increase longevity? HLE?**
- Health Insurance data, link to primary care data, ONS, Mosaic
- **Is a private care home better than an LA home?**
- LA care home census, BUPA care home census, link to primary care data, ONS
- ***Need for a trusted intermediary able to provide such linkages***

Institute and Faculty of Actuaries

Questions

Comments

Expressions of individual views by members of the Institute and Faculty of Actuaries and its staff are encouraged.

The views expressed in this presentation are those of the presenter.

Institute and Faculty of Actuaries