Infectious disease epidemiology: slowly moving towards open science

Niel Hens

universiteit hasselt | I-BioStat | KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Jonge Academie

Health Economics & Modelling Infectious Diseases
Vaccine & Infectious Disease Institute
University of Antwerp

'Embracing Data Management - Bridging the Gap Between Theory and Practice',
FOSTER, 4 June 2015, Brussels

Outline

Opinions are contagious, or not?

LaCour, M. and Green, D. Science, 2014 - retracted

When contact changes minds: an experiment on transmission of support for gay equality

→ Outcome: Contact with minorities coupled with discussion of issues pertinent to them is capable of producing a cascade of opinion change.

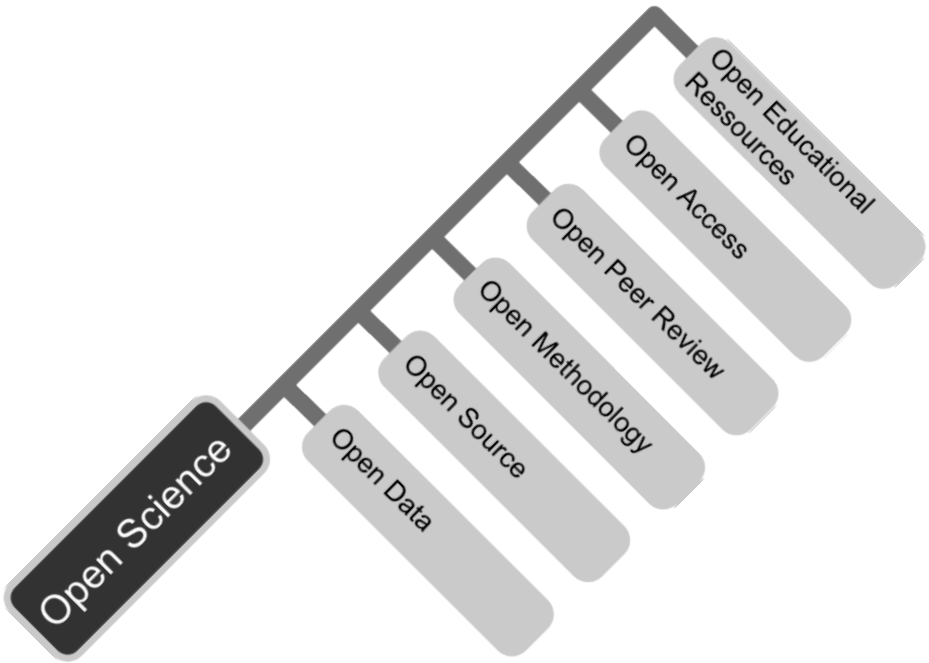
→ Retraction because, among other issues: 'LaCour has not produced the original survey data from which someone else could independently confirm the validity of the reported findings.'

The Chronicle, 21 May 2015: How 2 Persistent Grad Students Upended a Blockbuster Study

Compliance with data sharing is challenging.

- Allison, D. Science, 2009: The movement toward open data has begun: NIH, Science, Nature journals.
- Nature Genetics requires authors to publicly deposit microarray gene expression data:
  - reproduction of published work for only 2 of 18 papers.
  - data were either not found or incomplete or original analysis descriptions were unclear
- Public Library of Science journals require data sharing
  - until recently this didn't happen
  - now they insist

Open Science



Statistics: clinical trials

- Clinical trials are the best regulated studies
  - https://clinicaltrials.gov
  - https://www.clinicaltrialsregister.eu
- JAMA: clinical trials sponsored by drug companies have to be reanalyzed by 'academic statisticians' before acceptance for publication in the journal
- Keiding, N. Biostatistics, 2010
  - There is usually much more to 'research' than the statistical analysis.
  - A 'reasonable choice of statistical model'
- Clinical trials have changed data management
- Unfortunately not all guidelines and recommendations make sense and require constant re-evaluation.

Statistics: sharing code

- Editorial board meeting @Biometrics:

  issue: mandatory submission of R-package (stats software)

  AEs voted against: burden & risk of packages with limited lifespan
- Sharing code: 'available upon request'
- The Journal of Statistical Software: 1st rank

Infectious Disease Epidemiology: sharing data?

- Mostly observational studies or surveys:
  - outbreak data: SARS, H1N1, Ebola, MERS, ...

    - no clear data format
  - social contact survey (EU FP6 project POLYMOD)
- Emerging trend: syndromic surveillance
  - the great influenza survey
  - funding issues
  - little political support

Infectious Disease Epidemiology: sharing data?

The Good, the Bad and the Ugly

The good, ...

- Benjamin Cowling and colleagues: influenza research:
  - https://sites.google.com/site/bencowling88/datasets
  - shares raw data & code according to NIH guidelines
    - promote reproducibility of results
    - allow others to conduct their own analyses
    - allow others to compare other data
    - allow others to plan their own studies
- John Edmunds and colleagues: Ebola
  - http://cmmid.lshtm.ac.uk/research/ebola/
  - shares code and results in realtime on website
  - publications follow later
- ...

The good, . . .

- Jombaert et al. (Epidemics, 2014): OutbreakTools - Hackathon - Git repository

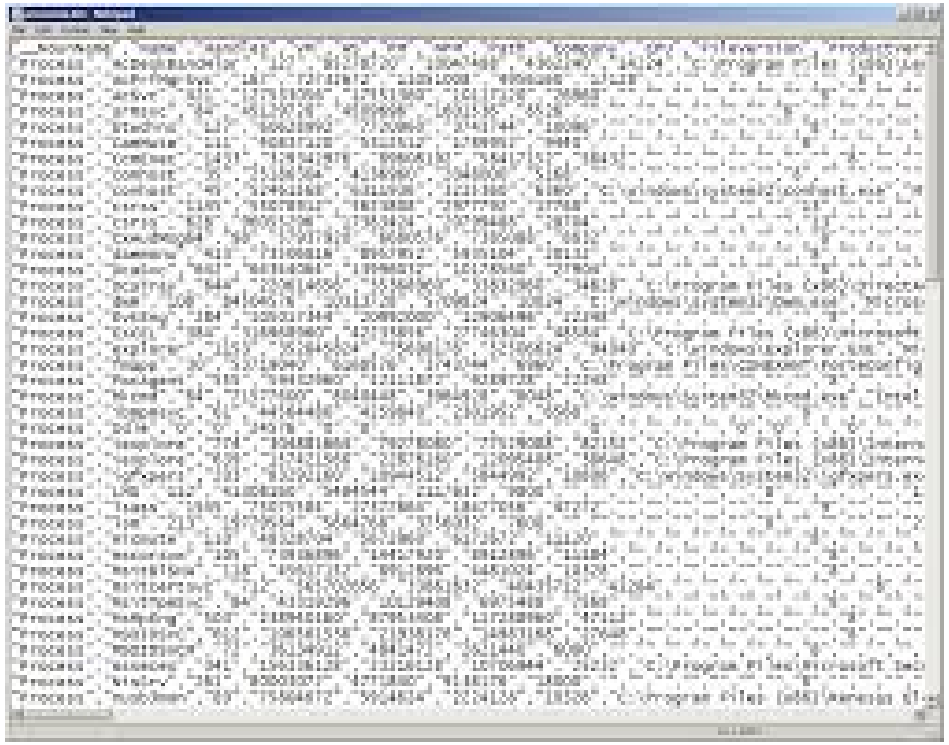
- Hens et al. (2012): ± 9800 e-chapter downloads - all data & code available

the bad, . . .

- violation of scientific integrity
- tax payer's money
- human health

and the ugly.

- Sharing of raw data without documentation:



- Even with documentation it can go wrong . . .

Social Contact Survey

Goal:

Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents.

Objectives

- Disentangle contact behaviour from transmission process
- Get insights in predictiveness of social contact data
- Get new insights in the transmission process

Social Contact Survey

- Wallinga et al. (2006): Utrecht
- POLYMOD
  - pilot study: Beutels et al. (2006)
  - main study: Mossong et al. (2008)

OPEN ACCESS Freely available online PLOS MEDICINE

Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases

Joël Mossong[1,2*], Niel Hens[3], Mark Jit[4], Philippe Beutels[5], Kari Auranen[6], Rafael Mikolajczyk[7], Marco Massari[8], Stefania Salmaso[8], Gianpaolo Scalia Tomba[9], Jacco Wallinga[10], Janneke Heijne[10], Malgorzata Sadkowska-Todys[11], Magdalena Rosinska[11], W. John Edmunds[4]

1 Microbiology Unit, Laboratoire National de Santé, Luxembourg, Luxembourg, **2** Centre de Recherche Public Santé, Luxembourg, Luxembourg, **3** Center for Statistics, Hasselt University, Diepenbeek, Belgium, **4** Modelling and Economics Unit, Health Protection Agency Centre for Infections, London, United Kingdom, **5** Unit Health Economic and Modeling Infectious Diseases, Center for the Evaluation of Vaccination, Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium, **6** Department of Vaccines, National Public Health Institute KTL, Helsinki, Finland, **7** School of Public Health, University of Bielefeld, Bielefeld, Germany, **8** Istituto Superiore di Sanità, Rome, Italy, **9** Department of Mathematics, University of Rome Tor Vergata, Rome, Italy, **10** Centre for Infectious Disease Control Netherlands, National Institute for Public Health and the Environment, Bilthoven, The Netherlands, **11** National Institute of Hygiene, Warsaw, Poland

(WoK: 447 citations)

Social Contact Survey

Belgian Contact Survey

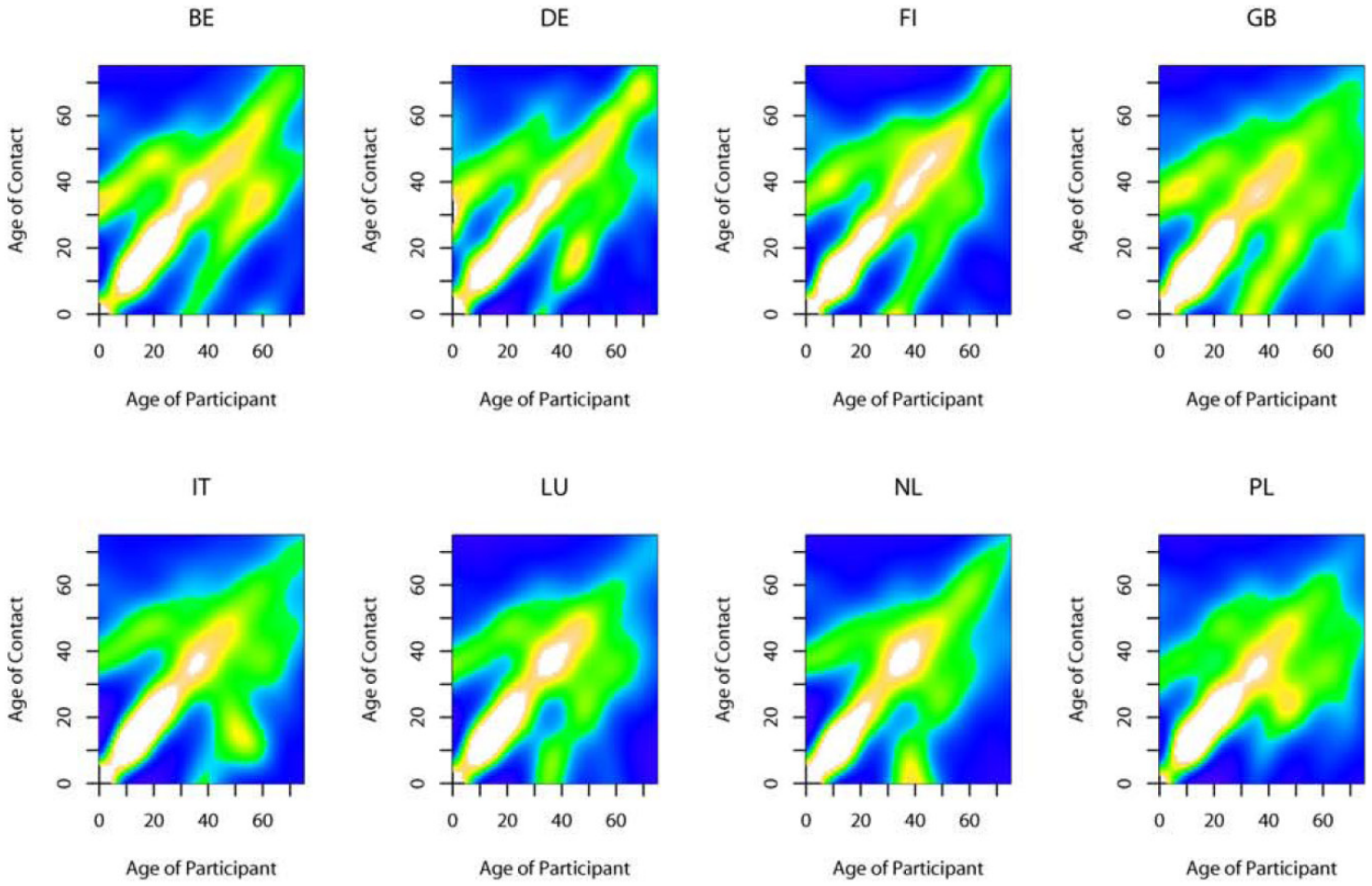
- Part of POLYMOD project
- Period March - May 2006
- 750 participants, selected through random digit dialing
- Diary-based questionnaire
- Two main types of contact: non-close and close contacts
- Total of 12775 contacts ($\approx$ 16 contacts per person per day)

Leeftijd (of leeftijdscategorie)	Geslacht		Plaats van contact (meerdere antwoorden mogelijk)					
	♀ vrouw	♂ man	thuis	werk	crèche, peutertuin, school, hogeschool universiteit	onderweg (auto, trein bus, ...)	vrije tijd	andere
9 (-)		X	X			X		
25 (-30)	X							X

Hoe dikwijls ziet u deze persoon?					Hebt u hem/haar aangeraakt? (bv hand geven, kus geven, sport)		Tijd doorgebracht met deze persoon				
(bijna) elke dag	enkele keren per week	enkele keren per maand	enkele keren per jaar of minder	eerste keer	ja	neen	Minder dan 5 min	5-15 min	15 min -1u	1-4u	4u of langer
X					X					X	
			X			X			X		

Hens et al. (2009a,b)

Mixing patterns in EU



Mixing patterns in EU

- Data were shared, via research gate, the VENICE platform etc
- Data were misused
  - because they were made available
  - because researcher did not check the supplementary material
  - 'supplementary'?
  - ...?

  Examples:
  - hepatitis A
  - impact of weather on social contact behaviour
  - reciprocity

Mixing patterns in EU

Supplementary Information Table S1. Details of survey methodology in each country

Country	**Belgium BE**	**Germany DE**	**Finland FI**	**Great Britain GB**	**Italy IT**	**Luxembourg LU**	**Netherlands NL**	**Poland PL**
Recruitment method	Random digit dialling on fixed telephone lines	Door-to-door household visits	Stratified sample from population register, mobile or land lines	Door-to-door household visits	Random digit dialling on fixed telephone lines	Random digit dialling on fixed telephone lines	Sampling from a register	Household visits
Survey field work conducted by	FQC, www.fqc.be	TNS Infratest, www.tns-infratest.com	TNS Gallup, www.tns-gallup.fi	FML, www.fm-ltd.co.uk	CRA Research, www.cra-research.com	TNS-ILRES, www.tns-ilres.lu	Inhouse, RIVM	OPINIA Agencja Badan Rynku, www.badania-rynku.com.pl
Financial incentives	5€	None	None	£3	None	5€	10€ for participation in the larger study	5€
Sampling method	Quota sampling by age, sex, geographical region, rural/urban	Quota sampling by age	Quota sampling by age and sex	Quota sampling by age, sex, geographical region	Quota by age, sex, geographical region, rural/urban and day of the week	Quota sampling by age and sex	Quota sampling by age and geographical region	Quota sampling age, sex, geographical region, rural/urban
How were the diaries delivered and collected?	Delivered and collected by mail	Face-to-face	Delivered and collected by mail	Delivered by mail, collected in person	Delivered and collected by mail	Delivered and collected by mail	Delivered in person and collected by mail	Delivered and collected in person

Mixing patterns in EU

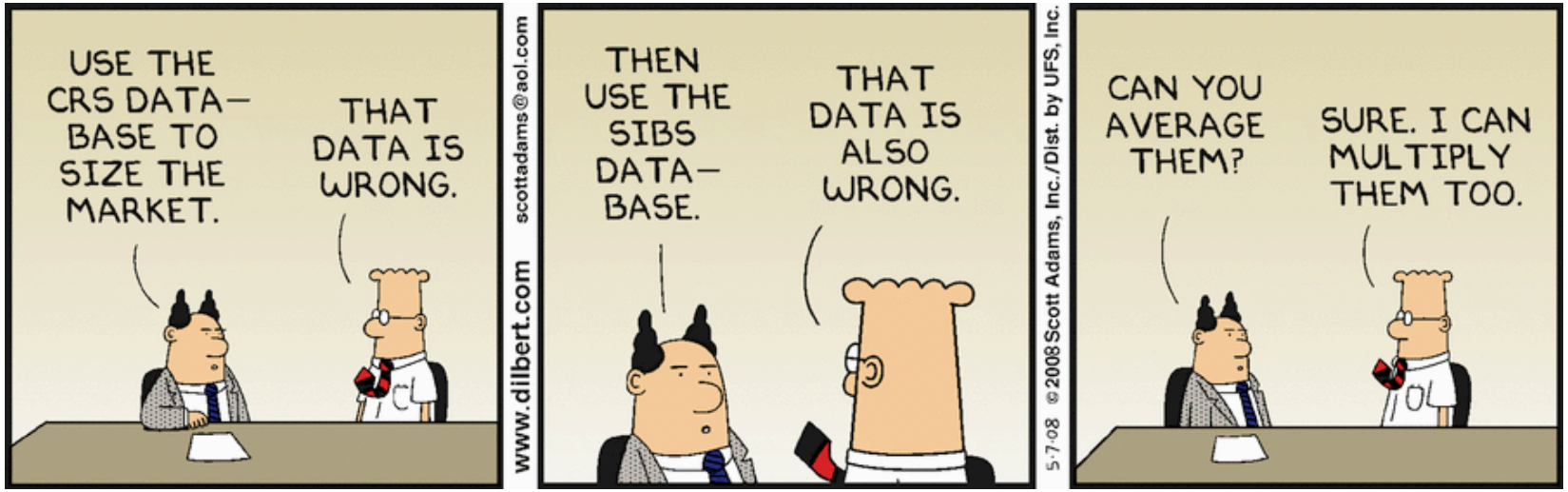
Country	**Belgium BE**	**Germany DE**	**Finland FI**	**Great Britain GB**	**Italy IT**	**Luxembourg LU**	**Netherlands NL**	**Poland PL**
Follow-up of participants • Reminder call before assigned day (RC) • Follow-up call after assigned day (FU1) • Follow-up call if diary unreturned (FU2)	RC, FU1, max. three FU2	No	FU1		RC, FU1	RC, FU1, FU2	No	
Was the survey part of a larger study?	No *	Yes, participants of a large multi-theme survey	No	No	No	No	Yes, large national serological study on vaccine preventable infections	No
Data quality checks	all manually checked and double entered	100 diaries double entered	Not known	double entry	double entry	single entry, all diaries checked manually	10% double entry	single entry, all diaries checked manually
Over what time period was the survey conducted?	March-May 2006	Feb-Mar, Jun-Jul 2006	April-June 2006	April 2006	May 2006	May-June 05, April-May 06	Feb-June 2006	March-April 2006
Age groups used for different diary types: young children (parental proxy reporting), older children and teens (self-reporting), adults (self-reporting)	0-8 parental proxy reporting, 9-17 self-reporting children and teenagers, 18+ self reporting adults	0-13 parental proxy reporting 14+ self-reporting No younger children self-reporting	0-12 parental proxy reporting 13-17 self-reporting teenagers, 18+ self reporting adults	0-7 parental proxy reporting, 8-17 self-reporting children and teenagers, 18+ self reporting adults	0-8 parental proxy reporting, 9-17 self-reporting children and teenagers, 18+ self reporting adults	0-8 parental proxy reporting, 9-17 self-reporting children and teenagers, 18+ self reporting adults	0-8 parental proxy reporting, 9-17 self-reporting children and teenagers, 18+ self reporting adults	0-8 parental proxy reporting, 9-17 self-reporting children and teenagers, 18+ self reporting adults
Were participants instructed not to record professional contacts (eg, with clients) in the diary?	Yes, if estimated at more than 20	Yes, if estimated at more than 10	Yes, if estimated at more than 10	No	No	No	Yes, if estimated at more than 10	No

Mixing patterns in EU

Country	**Belgium BE**	**Germany DE**	**Finland FI**	**Great Britain GB**	**Italy IT**	**Luxembourg LU**	**Netherlands NL**	**Poland PL**
Maximum number of contact entries in the diary ?	90	73	34	29	45	55	45	45
What languages were used?	French & Dutch**	German	Finnish and Swedish	English	Italian	German, French & Portuguese	Dutch	Polish

* Participants were assigned to collect diary data on two days of the week, one weekend and one workday per respondent. For the purpose of the current comparative multicountry analysis, a random selection was made from the Belgian data of one day per respondent, evenly distributed over all days of the week.
** Different contextual terminology between Belgium and The Netherlands (Dutch) or Luxembourg (French), mainly related to education, employment and migration

Mixing patterns in EU



Discussion & Recommendations

- The Good, the Bad and the Ugly:
  - → trade-off: sharing raw data requires documenting the data and sharing code
  - → worthwhile the effort
- There is hope: moving towards a unifying format for outbreaks (FF100, UK, 2009)
- Much more work needs to be done though!

"I think you'll find that mine is bigger..."

References

Hens, N., Ayele, G. M., Goeyvaerts, N., Aerts, M., Mossong, J., Edmunds, J. W., and Beutels, P. (2009a). Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. *BMC Infectious Diseases*, 9:187.

Hens, N., Goeyvaerts, N., Aerts, M., Shkedy, Z., Damme, P. V., and Beutels, P. (2009b). Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infectious Diseases*, 9:5.