

Open data: Balancing transparency with resilience

Stephan Lewandowsky

University of Bristol and University of Western Australia



stephan.lewandowsky@bristol.ac.uk

Transparency and Open Data are Essential



- Because if we (as a community) don't act, others will do it for us
- Recent UK Parliament inquiry hinted at government action [\(Video of hearing\)](#)
- U.S. NIH currently seeks to reclassify all basic research as clinical trials [\(Washington Post\)](#)

Transparency and Open Data are Essential

- But we need to consider their full implications
- Transparency and open data are political
- Transparency can be weaponized:
 - intentions of people who request data
 - consent and ethics
 - competence of people who request data
 - communication platforms
 - cherry-picking

Open Data is Political

- U.S. data access act 1998 (and 2000)
 - all data from federally funded projects available
 - data disseminated by government must adhere to act's definition
 - citizens can challenge
 - “influential data” reanalysis by “qualified” individuals
 - privately funded data disclosure

- The acts were drafted by the tobacco industry and allies
- Implementation was overseen by tobacco industry
- Most challenges launched by lobbyists or industry not public
- Many regulations delayed

(Baba et al., 2005, *American Journal of Public Health*)

“A delicate diplomatic situation”: tobacco industry efforts to gain control of the Framingham Study

Janine K. Cataldo^{a,*}, Lisa A. Bero^b, Ruth E. Malone^c

^a*Department of Physiological Nursing—Gerontology, University of California San Francisco, San Francisco, CA, USA*

^b*Department of Clinical Pharmacy and Health Policy, University of California San Francisco, San Francisco, CA, USA*

^c*Department of Social and Behavioral Sciences, University of California San Francisco, San Francisco, CA, USA*

Accepted 30 January 2010

- Tobacco industry used access to raw data for re-analysis by industry consultant
- Tobacco-related morbidity disappeared
 - Lead (paint, gasoline)
 - PVC
 - Any type of pollution

Open Data may Never be Open Enough

- Lamar Smith (R-TX), Chair of the House of Representatives Science Committee
- Issued subpoena in 2016 to National Oceanic and Atmospheric Administration (NOAA) ...
- ... for climate data that is not available (via Google Earth)
- The catchphrase “secret science” to advocate for data disclosure discussed by the tobacco industry as early as 1998 (Gianelli, 1998)

When open data is a Trojan Horse: The weaponization of transparency in science and governance

Karen EC Levy¹ and David Merritt Johns²

Big Data & Society
January–June 2016: 1–6
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/2053951715621568
bds.sagepub.com



- Transparency can be weaponized:
 - intentions of people who request data
 - consent and ethics
 - competence of people who request data
 - communication platforms
 - cherry-picking

I. Do Intentions Matter?

- Open Data advocates: No
- Tobacco industry: No, why?
- Public health researchers: Absolutely yes!

Nefarious Intentions and Science

- Dr. Smith publishes a paper suggesting interference causes forgetting.
 - Dr. Jones requests data from Dr. Smith and discovers that it is actually leprechauns that caused the forgetting.
 - Dr. Jones is Dr. Smith's ex-spouse and recently lost a bitter custody battle for their children and a hamster.
 - Ultimately the scientific community ignores the issue.
- Egos are bruised
 - Careers hampered
 - But no (not much?) lasting harm done

Nefarious Intentions and Public Policy

- Dr. Smith publishes a paper suggesting that Product X causes cancer.
- Dr. Jones, who works for Manufacturer X, requests data and purports to discover that cancer is caused by “dispositional factors” rather than Product X.
- Ultimately the scientific community resolves the issue.

But massive
harm is done

Scientific Debate \neq Public Debate

nature
climate change

ARTICLES

PUBLISHED ONLINE: 28 OCTOBER 2012 | DOI:10.1038/NCLIMATE1720

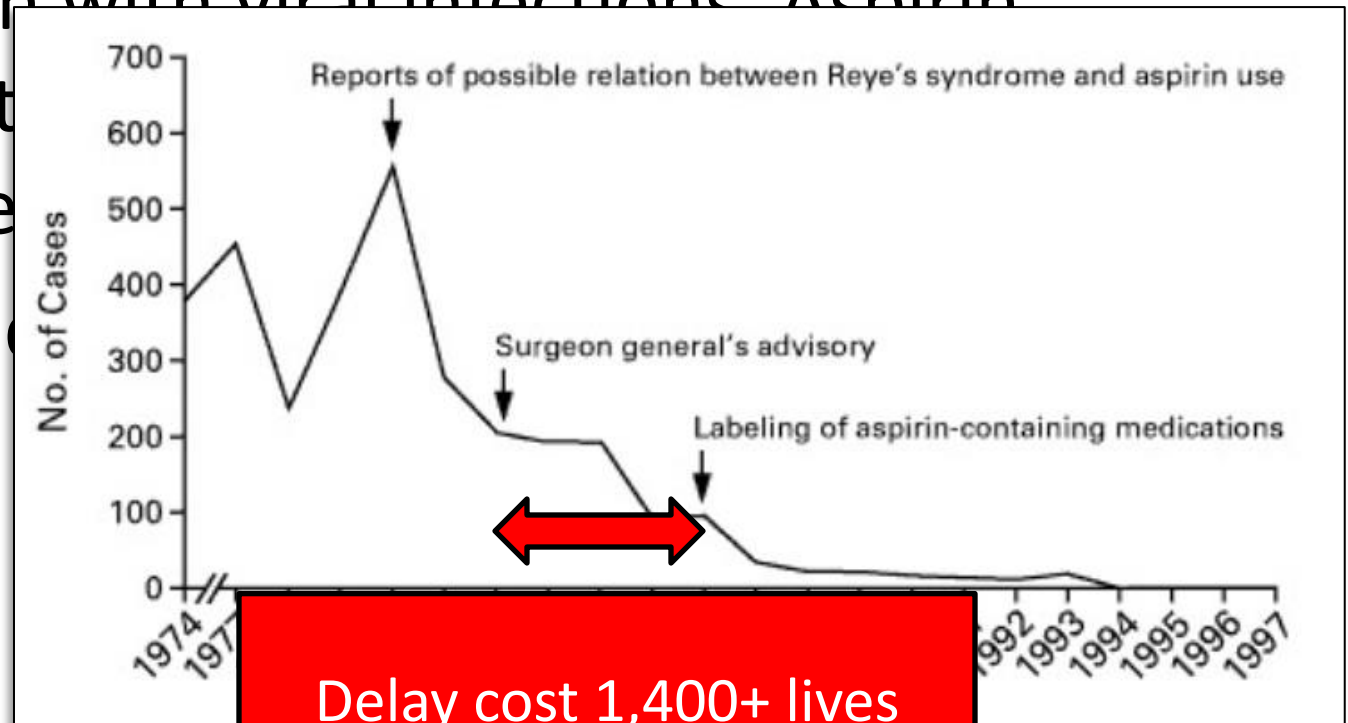
The pivotal role of perceived scientific consensus in acceptance of science

Stephan Lewandowsky[★], Gilles E. Gignac and Samuel Vaughan

- The appearance of a scientific debate, whether real or not, prolongs public indecisiveness.
- Tobacco control legislation was delayed by decades due to *appearance* of scientific debate.

Aspirin and Reye's Syndrome (e.g., Michaels & Monforton, 2005)

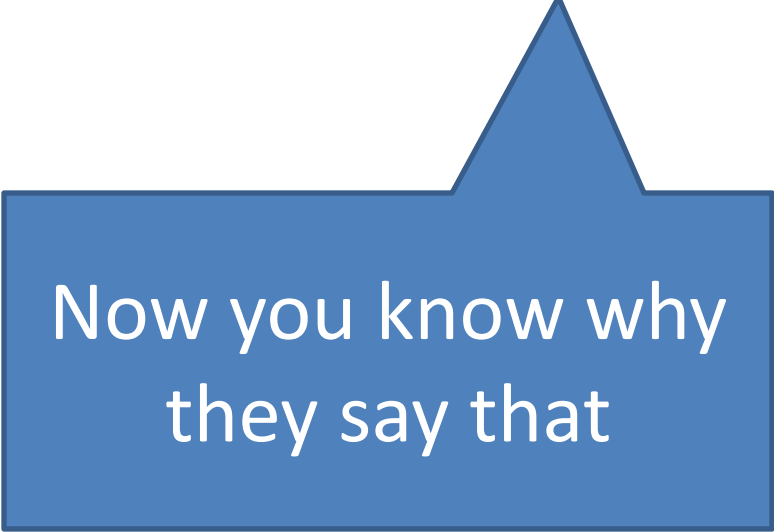
- In children with viral infections, Aspirin consumption is associated with Reye's syndrome
- 1 in 3 children with Reye's syndrome die



Delay cost 1,400+ lives
(Author of *Data Quality Act* a key figure in delay)

I. Do Intentions Matter?

- Open Data advocates: No
- Tobacco industry: No, why?
- Public health researchers: Absolutely yes!



Now you know why
they say that

II. Consent and Ethics

- Medical or clinical research
 - patient confidentiality
 - anonymization can be difficult
 - even de-linking is insufficient unless the linking key has been destroyed or is held by another institution (U.K. data protection act)
 - consent may have been given for one stated purpose of a study only

Consent: Exploring Implications

- Ms. Jones consents to participate in an experiment that observes the effect of WM training on IQ
- The experimenter collects the usual covariates and demographics such as gender and ethnicity
- The Ku-Klux-Klan reads the results and discovers that blacks have higher IQ overall but that training
- The *Kluxer's Trumpet* headline: “No matter how much you train them, they cannot get smarter”

Consent: Exploring Implications

- Ms. Jones consents to participate in an experiment that observes the effect of WM training on IQ
- Suppose Ms. Jones was black
 - did she realize what she consented to?
 - would she have given consent if she knew this could happen?
 - given what happened, would she ever consent again?
 - note that anonymity is not

If data are open, they
are open to abuse

Concerns are Being Articulated

Sydes et al. *Trials* (2015) 16:104
DOI 10.1186/s13063-015-0604-6



METHODOLOGY

Open Access

Sharing data from clinical trials: the rationale for a controlled access approach

Matthew R Sydes^{*}, Anthony L Johnson, Sarah K Meredith, Mary Rauchenberger, Annabelle South and Mahesh KB Parmar

Abstract

Background: The move towards increased transparency around clinical trials is welcome. Much focus has been on under-reporting of trials and access to individual patient data to allow independent verification of findings. There are many other good reasons for data sharing from clinical trials. We describe some key issues in data sharing, including the challenges of open access to data. These include issues in consent and disclosure; risks in identification, including

III. Does Competence Matter?

- Researchers operate in an institutional context
 - ethics boards
 - data management plans

–

–

–

• M

Ce

–

–



The screenshot shows the MailOnline website. At the top is the 'MailOnline' logo. Below it is a navigation bar with links: Home, News, U.S., Sport, TV&Showbiz, Australia, Femail, Health, Science, Money. Below that is a red bar with links: Columnists Home, Polls. The main content area features a large headline: 'MMR safe? Baloney. This is one scandal that's getting worse'. Below the headline is a timestamp: 'Last updated at 10:24 31 October 2005'.

MailOnline

[Home](#) | [News](#) | [U.S.](#) | [Sport](#) | [TV&Showbiz](#) | [Australia](#) | [Femail](#) | [Health](#) | [Science](#) | [Money](#)

[Columnists Home](#) | [Polls](#)

MMR safe? Baloney. This is one scandal that's getting worse

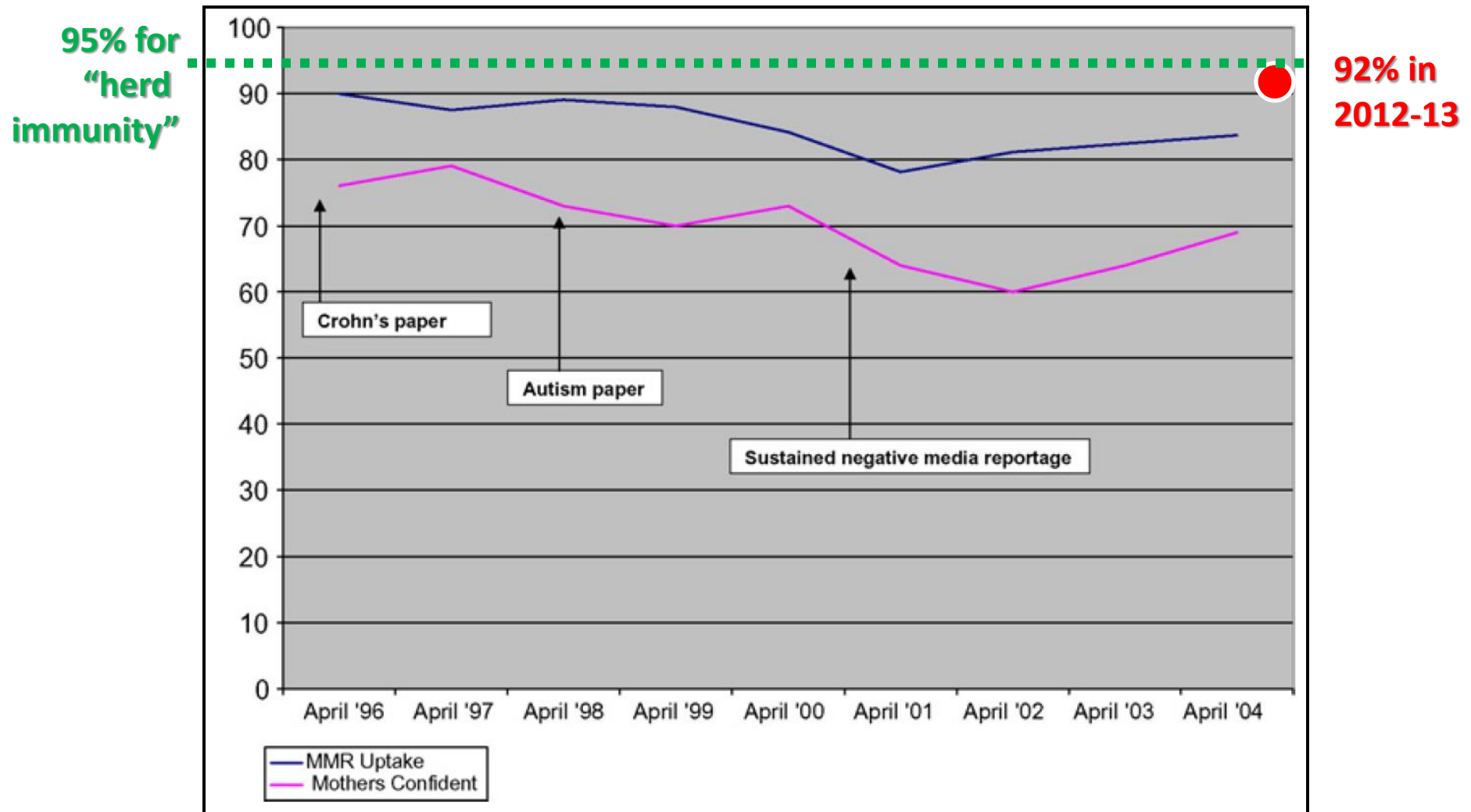
Last updated at 10:24 31 October 2005

Consideration of the Platforms

- Dr. Smith publishes a paper suggesting that Product X causes cancer.
- Dr. Jones, who works for Manufacturer X, requests data and discovers that cancer is caused by “dispositional factors” rather than Product X.
- Dr. Jones and Manufacturer X publish analysis on blogs and Twitter. The Daily Mail picks it up.
- Ultimately the scientific community resolves the issue.

U.K. MMR Vaccination Rates

(Smith et al., 2007)



IV. Cherry-Picking

- We urge scientists to preregister hypotheses and analysis plans to guard against cherry-picking of results or outcome measures.
- We do not keep track of the Ku-Klux-Klan requesting 90 data sets with a racial-identifier variable
- So they can trumpet the one result that yields the “desired” racial differences

What Does this Add up to?

- Science should be open and transparent
- But there is a distinction between science on the one hand, and noise, nonsense, commercial interests, or political propaganda on the other
- Openness and transparency facilitate science, but they also aid in the dissemination of noise, nonsense, commercial interests, and political propaganda

Solutions? Symmetry

- People who request data must be competent and must operate in an institutional context of accountability
- People who request data must preregister their intentions (and conform to them)
- Participants' consent must be considered
- Data availability (and limits) should be enshrined in peer-review record at the time of publication to avoid later controversy

Thank you



Don't let transparency
damage science

Stephan Lewandowsky and Dorothy Bishop explain (*Nature*, **2016**, 529, 459-461)

Importance of Competence

- U.K. Medical Research Council's guidelines:
“The custodian [of the data] must ensure that the group [receiving the data] accepts a duty of confidence and protects confidentiality through training procedures, etc, **to the same standards as the custodian.**”