



OPINION ARTICLE

Time to really share real-world data? [version 1; referees: 2 approved, 1 approved with reservations]

Sophie Graham¹, Laura McDonald², Radek Wasiak¹, Michael Lees³,
Sreeram Ramagopalan ²

¹Real-World Evidence, Evidera, London, W6 8DL, UK

²Centre for Observational Research and Data Sciences, Bristol-Myers Squibb, Uxbridge, UB8 1DH, UK

³Worldwide Health Economics and Outcomes Research, Bristol-Myers Squibb, Paris, F-92500, France

v1 First published: 11 Jul 2018, 7:1054 (doi: [10.12688/f1000research.15517.1](https://doi.org/10.12688/f1000research.15517.1))
Latest published: 11 Jul 2018, 7:1054 (doi: [10.12688/f1000research.15517.1](https://doi.org/10.12688/f1000research.15517.1))

Abstract

Data other than that from clinical trials are important for healthcare decision making. However, unlike the vocal calls seen for more open access to trial data, there are limited efforts being made to ensure that agencies that collect real-world data (RWD) share this, despite its importance. There are many RWD sources across the world that could be readily exploited for research once shared. There are policy and privacy questions that need to be tackled, but opening up and sharing RWD offers remarkable potential for improvements in care for individuals and more effective use of limited healthcare resources. Open science should become the standard for RWD as well as clinical trials, especially those that have a high likelihood to influence practice.

Keywords

real-world data, sharing, clinical data



This article is included in the **Science Policy Research gateway**.

Open Peer Review

Referee Status:

	Invited Referees		
	1	2	3
version 1 published 11 Jul 2018	report	report	report

- 1 **Sajan Khosla**, AstraZeneca Academy House, UK
- 2 **David Madigan**, Columbia University, USA
- 3 **Laurie A. Tomlinson** , London School of Hygiene & Tropical Medicine, UK

Discuss this article

Comments (0)

Corresponding author: Sophie Graham (sophie.graham@evidera.com)

Author roles: **Graham S:** Writing – Original Draft Preparation; **McDonald L:** Writing – Original Draft Preparation; **Wasiak R:** Writing – Review & Editing; **Lees M:** Conceptualization, Writing – Review & Editing; **Ramagopalan S:** Conceptualization, Writing – Review & Editing

Competing interests: SG and RW are employees of Evidera. LM, ML and SR are employees of Bristol-Myers Squibb.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2018 Graham S *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Graham S, McDonald L, Wasiak R *et al.* **Time to really share real-world data? [version 1; referees: 2 approved, 1 approved with reservations]** *F1000Research* 2018, 7:1054 (doi: [10.12688/f1000research.15517.1](https://doi.org/10.12688/f1000research.15517.1))

First published: 11 Jul 2018, 7:1054 (doi: [10.12688/f1000research.15517.1](https://doi.org/10.12688/f1000research.15517.1))

Introduction: Real-world data and its importance

Real-world data (RWD) are data collected in the course of routine health care delivery or otherwise generated without constraints, as in the ‘real-world’^{1,2}. These data are used to understand disease epidemiology, patterns of care and patient need as well as provide valuable insight into treatment effectiveness and safety in day to day clinical practice.

Due to their potential to provide insight into questions not effectively addressed from clinical trials, RWD are increasingly influencing health care decision making, including regulatory assessment, clinical practice and policy. For example, the 21st Century Cures Act in the United States has required the Food and Drugs Administration (FDA) to draw up guidelines for the role of RWD in drug approvals (https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe_white_paper_2017.09.06.pdf), and in the United Kingdom (UK) the Academy of Medical Sciences and the Association of the British Pharmaceutical Industry have recently prioritised supporting the inclusion of RWD in regulatory and health technology assessment (HTA) processes (<https://acmedsci.ac.uk/more/news/next-steps-for-using-real-world-evidence>). These data are also influencing policy; for example, in the UK the controversial decision to increase hospital staffing levels on the weekend, was driven by analysis from inpatient National Health Service (NHS) hospital data from 2009–10 showing a significantly higher risk of mortality in the 30-day follow up period from patients admitted on the weekend, compared to those admitted midweek³.

Why share data?

The benefits of sharing data are evidenced with the increasing transparency of clinical trial data. Access to clinical trial data has allowed the validation of initially reported trial results⁴, but perhaps more importantly openly available data has also been used to answer new questions about disease. For example, the Dialogue for Reverse Engineering Assessment and Methods (DREAM) challenge aimed at identifying a prognostic model for overall survival in patients with metastatic prostate cancer using the raw data from the comparator arms of four clinical trials led to the identification of aspartate aminotransferase as an important prognostic factor for overall survival⁵. Similarly, the SPRINT competition challenged entrants to re-analyse data from a randomized trial of blood-pressure control and this led to the development of a new decision-making tool for clinicians to determine whether patients should receive intensive hypertensive treatment or not based on patient characteristics⁶.

In genetics, the sharing of data have been a major enabling factor in the identification of disease associated genes. Genome-wide association studies often require data from tens of thousands of patients to have the statistical power to detect variants implicated in disease⁷. Initiatives such as the UK Biobank⁸ have enabled richly genotyped and phenotyped data on many thousands of patients to be available to the international scientific community. In doing so, these data have been used to better understand the genetic architecture underlying a number of complex diseases including Alzheimer’s, major depression and atrial fibrillation, among many others^{8–10}. Given the often prohibitively high time and cost burden associated with

establishing a cohort needed to unpick the genetic underpinnings of disease, the model employed by the UK Biobank exemplifies the benefits of data sharing not only by enabling reproducibility, but critically in the acceleration and advancement of scientific research.

A number of European clinical databases provide access to RWD. For example, in the UK, the **Clinical Practice Research Datalink** (CPRD), an electronic medical record database that covers general practitioner encounters, has been available to researchers and has led to the generation of hundreds of peer-reviewed publications, with notable contributions including validation of the safety of the measles, mumps and rubella vaccine¹. Further, CPRD data are being increasingly used in English national clinical guidelines and guidances¹¹. The Nordic countries also have a long tradition of collecting patient medical information in the form of national population-based and prescription registries. The high quality of the data recorded and the good coverage of the sources has contributed to a vast number of pharmacoepidemiological insights^{12,13}, for example in realising risks associated with serotonin reuptake inhibitors in pregnancy¹⁴.

Combining accessible RWD sets is of particular importance in Post-Authorisation Safety Studies (PASS), which are studies carried out after a medicine has been approved for use to obtain further information on its safety in the real-world. In such studies, large patient numbers are required to increase power to detect true rare adverse outcomes associated with a particular treatment. To meet the FDA’s Post Marketing Requirements for a post-marketing safety surveillance system, the **Mini-Sentinel Initiative** was launched in 2009, and was eventually expanded to the full Sentinel Initiative in February 2016. The initiative includes electronic data from different healthcare data holders that is automatically collected on an ongoing basis and merged into a common data format, which the FDA can query at any point to quickly and securely monitor drug safety issues. It now includes over 300 million person years of unduplicated data from over 17 different data partnerships. The Mini-Sentinel Initiative (pilot project) identified a 1.5 (95% CI, 0.2–3.2) increased risk of intussusception per 100,000 infants administered the rotavirus vaccination¹⁵, highlighting the power of sharing data to identify adverse events that occur at very low rates.

Lack of data sharing has consequences

There is a considerable lack of data sharing in observational research. This was shown in a recent review of 237 observational studies published in the BMJ from 1st January 2015 to 31st August 2017. This work found that 63% of studies reviewed during this period did not share the raw data upon which the analyses were conducted¹⁶. There are also important examples of large medical record data databases that do not readily share data with external researchers. These include data from the Spanish regional general practitioner database Base De Datos Para La Investigacion Farmacoepidemiologica en Atencion Primaria (BIFAP) and the Secure Anonymised Information Linkage (SAIL) databank, which includes primary and secondary care data from the Welsh population.

Lack of data sharing is potentially not without consequence. For example, the Infections in Oxfordshire Research Database (IORD) showed that any apparent aforementioned UK ‘week-end effect’ on mortality arises from patient-level differences at admission^{3,17,18}. Data from IORD has been available for many years and if this was openly accessible for researchers to analyse the negative impacts of policy changes could perhaps have been prevented. Rosiglitazone, a glitazone used in the treatment of type II diabetes, was initially approved by the European Medicines Agency in 1999, despite there being limited evidence to support during the approval process. In July 2010, the FDA convened an expert panel to discuss removing rosiglitazone from the market because of arising evidence that rosiglitazone was associated with increasing risk of myocardial infarction. The original evidence was based on the RECORD trial, which was an unblinded, open label trial¹⁹. It then took two additional clinical trials and over 10 years of post-marketing surveillance to detect an increased risk in myocardial infarction²⁰. Had a real-world PASS been carried out immediately after market approval utilising a number of different data sources along the lines of the aforementioned Sentinel initiative, unwarranted clinical outcomes could have been robustly detected much earlier.

Barriers to data sharing; combating privacy concerns

A fundamental barrier associated with the sharing of RWD are the concerns over patient confidentiality. This was illustrated when the UK governmental Care.data programme, which aimed to create a national combined primary and secondary care database, was halted over general public concerns for patient privacy (<https://blogs.biomedcentral.com/on-medicine/2016/08/19/need-care-closure-care-data/>). Stricter regulatory frameworks introduced as part of the General Data Protection Regulation (GDPR) within the EU aim to tackle many privacy issues involved in the storage, processing and management of data in a digital era. New

regulations will not only tighten organisational laws around the handling of data, but will also work to increase the rights of individuals, giving them more control over their personal information including the right to transparency, access and erasure. Importantly, despite privacy concerns, studies demonstrate that patients are generally willing to share their healthcare data in the context of contributing to public health as long as the potential benefits are appropriately communicated^{20,21}.

Conclusion

Even though there are significant potential benefits that improved access to RWD could provide patients, the progress to increase data sharing is slow. There are policy and privacy questions that need to be tackled²², but opening up and sharing healthcare data offers remarkable potential for improvements in care for individuals as well as potential for more effective use of limited healthcare resources. In some instances, for example with the [National Quality Registries in Sweden](#), this potential is already being realised. It is hoped that sharing of RWD becomes standard practice, especially in those data-sets that have a high likelihood of influencing public health and clinical practice.

Data availability

No data are associated with this article.

Competing interests

SG and RW are employees of Evidera. LM, ML and SR are employees of Bristol-Myers Squibb.

Grant information

The author(s) declared that no grants were involved in supporting this work.

References

- McDonald L, Lambrelli D, Wasiak R, *et al.*: **Real-world data in the United Kingdom: opportunities and challenges.** *BMC Med.* 2016; **14**(1): 97. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jarow JP, LaVange L, Woodcock J: **Multidimensional Evidence Generation and FDA Regulatory Decision Making: Defining and Using “Real-World” Data.** *JAMA.* 2017; **318**(8): 703–704. [PubMed Abstract](#) | [Publisher Full Text](#)
- Freemantle N, Richardson M, Wood J, *et al.*: **Weekend hospitalization and additional risk of death: An analysis of inpatient data.** *J R Soc Med.* 2012; **105**(2): 74–84. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ebrahim S, Sohani ZN, Montoya L, *et al.*: **Reanalyses of Randomized Clinical Trial Data.** *JAMA.* 2014; **312**(10): 1024–1032. [PubMed Abstract](#) | [Publisher Full Text](#)
- Guinney J, Wang T, Laajala TD, *et al.*: **Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: development of a prognostic model through a crowdsourced challenge with open clinical trial data.** *Lancet Oncol.* 2017; **18**(1): 132–142. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ledford H: **Open-data contest unearths scientific gems — and controversy.** *Nat News.* 2017; **543**(7645): 299. [PubMed Abstract](#) | [Publisher Full Text](#)
- Visscher PM, Wray NR, Zhang Q, *et al.*: **10 Years of GWAS Discovery: Biology, Function, and Translation.** *Am J Hum Genet.* 2017; **101**(1): 5–22. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wray NR, Ripke S, Mattheisen M, *et al.*: **Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression.** *Nat Genet.* 2018; **50**(5): 668–681. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Marioni RE, Harris SE, Zhang Q, *et al.*: **GWAS on family history of Alzheimer’s disease.** *Transl Psychiatry.* 2018; **8**(1): 99. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nielsen JB, Fritsche LG, Zhou W, *et al.*: **Genome-wide Study of Atrial Fibrillation Identifies Seven Risk Loci and Highlights Biological Pathways and Regulatory Elements Involved in Cardiac Development.** *Am J Hum Genet.* 2018; **102**(1): 103–115. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Oyinlola JO, Campbell J, Kousoulis AA: **Is real world evidence influencing practice? A systematic review of CPRD research in NICE guidances.** *BMC Health Serv Res.* 2016; **16**: 299. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Reigstad MM, Larsen IK, Myklebust TÅ, *et al.*: **The Nordic Health Registries: an important part of modern medical research.** *Hum Reprod.* 2016; **31**(1): 216–217. [PubMed Abstract](#) | [Publisher Full Text](#)
- Furu K, Wettermark B, Andersen M, *et al.*: **The Nordic countries as a cohort for pharmacoepidemiological research.** *Basic Clin Pharmacol Toxicol.* 2010; **106**(2):

- 86–94.
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Kieler H, Artama M, Engeland A, *et al.*: **Selective serotonin reuptake inhibitors during pregnancy and risk of persistent pulmonary hypertension in the newborn: population based cohort study from the five Nordic countries.** *BMJ.* 2012; **344**: d8012.
[PubMed Abstract](#) | [Publisher Full Text](#)
 15. Yih WK, Lieu TA, Kulldorff M, *et al.*: **Intussusception risk after rotavirus vaccination in U.S. infants.** *N Engl J Med.* 2014; **370**(6): 503–512.
[PubMed Abstract](#) | [Publisher Full Text](#)
 16. McDonald L, Schultze A, Simpson A, *et al.*: **Lack of data sharing in observational studies.** *BMJ.* 2017; **359**: j4866.
[PubMed Abstract](#) | [Publisher Full Text](#)
 17. Walker AS, Mason A, Quan TP, *et al.*: **Mortality risks associated with emergency admissions during weekends and public holidays: an analysis of electronic health records.** *Lancet.* 2017; **390**(10089): 62–72.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 18. McDonald L, Ramagopalan SV, Lees M: **Real-world data really matter.** *CMAJ.* 2017; **189**(41): E1293.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 19. Home PD, Pocock SJ, Beck-Nielsen H, *et al.*: **Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial.** *The Lancet.* 2009; **373**(9681): 2125–2135.
[PubMed Abstract](#) | [Publisher Full Text](#)
 20. Mello MM, Lieou V, Goodman SN: **Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing.** *N Engl J Med.* 2018; **378**(23): 2202–2211.
[PubMed Abstract](#) | [Publisher Full Text](#)
 21. Weitzman ER, Kelemen S, Kaci L, *et al.*: **Willingness to share personal health record data for care improvement and public health: a survey of experienced personal health record users.** *BMC Med Inform Decis Mak.* 2012; **12**: 39.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 22. Warren E: **Strengthening Research through Data Sharing.** *N Engl J Med.* 2016; **375**(5): 401–403.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 15 October 2018

doi:10.5256/f1000research.16919.r38761



Laurie A. Tomlinson 

Department of NonCommunicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK

Should real-world data be more freely available? Of course it should, it is impossible to disagree. But, "I think you'll find it's a bit more complicated than that".

Use of observational data for drug safety (among other things) is crucially important. However, many things are critical to make this happen. High quality research training with access to excellent statisticians. A research system that doesn't depend on impact factor and no benefit for getting an 'interesting' result. Quality peer-review. Lots of funding to do important work of this kind. An educated audience capable of critical analysis of complex study designs. Brexit not gutting the MHRA. Close links with the EMA. Affordable access for academics to access the data. And yes open data would help but who has the time to rerun the analyses of other groups who have proven themselves trustworthy when there are probably more important things to do? There are also risks. Anti-vaxxers on a mission to identify spurious drug-effects. And vitally loss of public trust. Yes there are surveys showing public support for access to their data but we have already made a mess of this with care.data and can we risk that again? Especially when pharma or profit-driven companies are promoting it? In a world ever more driven by conspiracy theories and paranoia this seems a big step at the moment.

Yes real-world data should be more open. But at the moment it is only one of many important battles to fight.

Is the topic of the opinion article discussed accurately in the context of the current literature?

Partly

Are all factual statements correct and adequately supported by citations?

Partly

Are arguments sufficiently supported by evidence from the published literature?

Partly

Are the conclusions drawn balanced and justified on the basis of the presented arguments?

Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 26 July 2018

doi:10.5256/f1000research.16919.r36415



David Madigan

Department of Statistics, Columbia University, New York City, NY, USA

This is an interesting short opinion piece about RWD. I worry that it exhibits the kind of unbridled enthusiasm that I fear ultimately undermines efforts to appropriately incorporate RWD into healthcare decision making. There are real limitations to RWD studies and the authors should acknowledge these. See, for example, many of Ioannidis's papers or Madigan, D., et al. (2013)¹.

The CPRD is not a particularly good example since the data are available to researchers under severe restrictions, e.g., "A dataset may only be used for the study for which it has been approved."

As I understand it, Sentinel queries require multi-week/month turnaround times.

The notion that an observational study could have uncovered the rosiglitazone problem early is speculative.

The section on privacy concerns pays lip service to some really complex issues. These data are de-identified in superficial ways and identification of specific individuals is entirely possible. The implications for health insurance, employment, etc. are highly complex.

The paper should discuss common data models. Shared data with idiosyncratic data models are of somewhat limited value. In a similar vein, the article should discuss OHDSI.

References

1. Madigan D, Ryan PB, Schuemie M, Stang PE, Overhage JM, Hartzema AG, Suchard MA, DuMouchel W, Berlin JA: Evaluating the impact of database heterogeneity on observational study results. *Am J Epidemiol.* 2013; **178** (4): 645-51 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the topic of the opinion article discussed accurately in the context of the current literature?

Partly

Are all factual statements correct and adequately supported by citations?

Partly

Are arguments sufficiently supported by evidence from the published literature?

Partly

Are the conclusions drawn balanced and justified on the basis of the presented arguments?

Partly

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Referee Report 18 July 2018

doi:10.5256/f1000research.16919.r35957



Sajan Khosla

AstraZeneca Academy House, Cambridge, UK

I would suggest adding commentary on the reproducibility of RWD studies by better sharing of code and data. Those practitioners of RWD are all aware of the intricate nuances that lead to variable definitions and sharing of these may enable better reproducibility. Perhaps the better use of technology to better develop transparency and reproducibility would be beneficial in these areas, leading the clinical trial setting¹.

The importance of RWD is a combination of the potential regulatory use within the context of the 21st Century Cures act, being able to reinforce clinical decision making with information at the patient bedside and to also support policy making decisions. Having been part of the team on the weekend mortality paper, this analysis was less influential but more supportive to larger initiatives at hand. I would also support that with the outputs from international disease registries in oncology which highlight how poor/late presentation of cancer within the UK system leads to the significant gap in cancer outcomes and led to better public health initiatives around recognising symptoms of cancer².

References

1. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong IC, Rijnbeek PR, van der Lei J, Pratt N, Norén GN, Li YC, Stang PE, Madigan D, Ryan PB: Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform.* 2015; **216**: 574-8 [PubMed Abstract](#)
2. Rachet B, Maringe C, Nur U, Quaresma M, Shah A, Woods LM, Ellis L, Walters S, Forman D, Steward J, Coleman MP: Population-based cancer survival trends in England and Wales up to 2007: an assessment of the NHS cancer plan for England. *Lancet Oncol.* 2009; **10** (4): 351-69 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the topic of the opinion article discussed accurately in the context of the current literature?

Partly

Are all factual statements correct and adequately supported by citations?

Yes

Are arguments sufficiently supported by evidence from the published literature?

Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?

Partly

Competing Interests: No competing interests were disclosed.

Referee Expertise: Real World Evidence, Health Informatics, Data science

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research