



“Fake News” in the Rush to Inform on COVID-19: How Research Retractions Can Inform Best Practices for Real-World Data Quality and Reporting

Rachael L. DiSantostefano, PhD¹, Erica A Voss, MPH,^{1,2} Clair Blacketer, MPH^{1,2}

¹Janssen Research and Development, Titusville, NJ; ²OHDSI collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York, NY

Disclosure: All authors are full-time employees of Janssen R&D, hold pension rights, and are shareholders of Johnson & Johnson. Contact: Rachael DiSantostefano, rdisanto@its.jnj.com

Background

Introduction

- COVID-19 publications may represent the largest growth of scientific literature ever, with >~16-23,000 papers and >~6,000 preprints through April 2020 with projected significant growth [1,2].
- This proliferation of publications includes two high-profile retractions (e.g., Surgisphere data), which have the potential to erode trust in real-world data (RWD) [3,4].

Objectives

- Analyze manuscript retractions with a focus on COVID-19 and
- propose RWE best practices based on insights from these retractions

Methods

Database

- RetractionWatch® database (RWDB), largest curated database of manuscript retractions [5].

Analysis

- Retractions in 2020 were identified (search terms: COVID-19, coronavirus, SARS-Cov-2, 2019-nCov).
- We described characteristics of COVID-19 retractions reported in RWDB and extracted additional information from the publications relating to context, data sources & impact of retraction (**Table 1**).
- Based on this assessment, we 1) compared to historical retractions & 2) identified best practices to address retraction reasons, focusing on transparency and data quality assessment using OHDSI.

Results

Table 1: Characteristics of COVID-19 Retractions through September 23, 2020 from RWDB

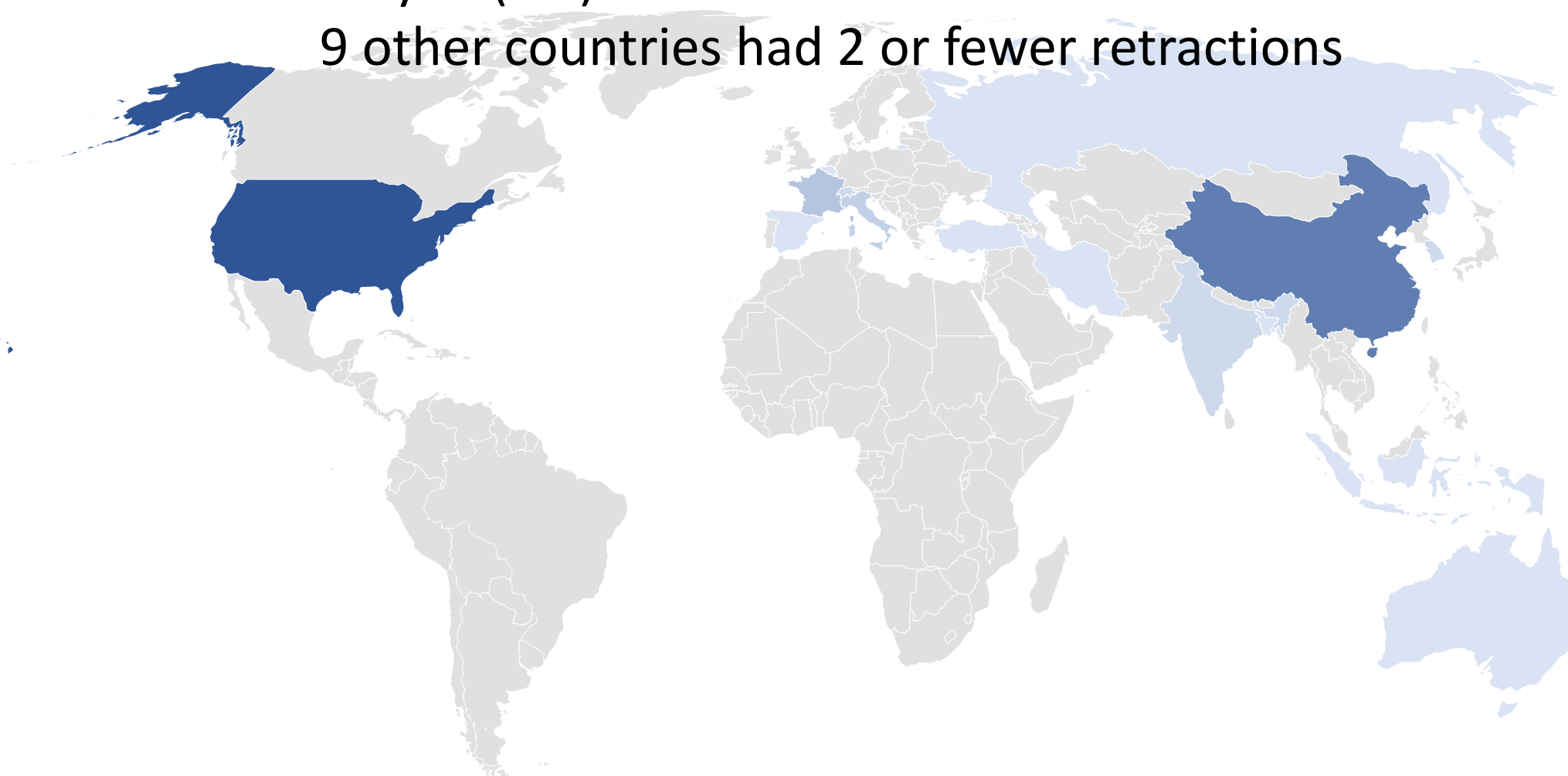
Characteristic	n=36 Retractions	
Time to Retraction (Days)	Statistics	
Min, Max	0, 94	
Subset of Retraction Reasons (multiple responses allowed)	Frequencies	
	Count	Percent
Data issue*	8	22
Analysis, result, or conclusion issue*	7	19
Limited or no withdrawal information	6	17
No notice of withdrawal	5	14
Lacking IRB, 3 rd party, author approval	5	14
Duplicate publication in error	3	8
Ethical concerns	1	3
Status at Withdrawal	Count	Percent
In press	3	8
Not Prepared	1	3
Preprint	12	33
Published	20	55
Publisher	Count	Percent
Elsevier	12	33
Cold Spring Harbor Laboratory Press	12	33
Other publishers combined	12	33

% for Status at Withdrawal and Publisher do not add to 100 due to rounding.

*Issues include concerns, reliability, and errors.

16 Countries Represented in Retractions

Number of papers with authors from
USA=15 (42%)
China=11 (31%)
France=4 (11%)
Italy=3 (8%)
9 other countries had 2 or fewer retractions



Historical Retraction Information [6]

- Retraction rate: 4 per 10,000 publications
- Retractions can take a few years on average
- About half of retraction reasons relate to fabrication or plagiarism
- Many publications offer few specifics about retraction reasons
- Journals with higher impact factors do more policing

- COVID-19 articles were reviewed, published, and retracted more quickly than historical retractions [6].
- 5 retractions involved hydroxychloroquine (not shown). A 3rd paper using Surgisphere data was withdrawn on off-label use of ivermectin in COVID-19 but resulted in use (>350,000 people) & inclusion in guidelines [7].
- Limitations: COVID-19 retractions varied considerably in data/study type, and future work should isolate RWD in pharmacoepidemiology. A longer time horizon is needed to compare against historical trends.
- Retraction reasons (**Table 1**) can be partially addressed by requiring transparent, unambiguous reporting of data preparation & analysis. This allows robust assessment of data quality & analysis validity (**Table 2**).

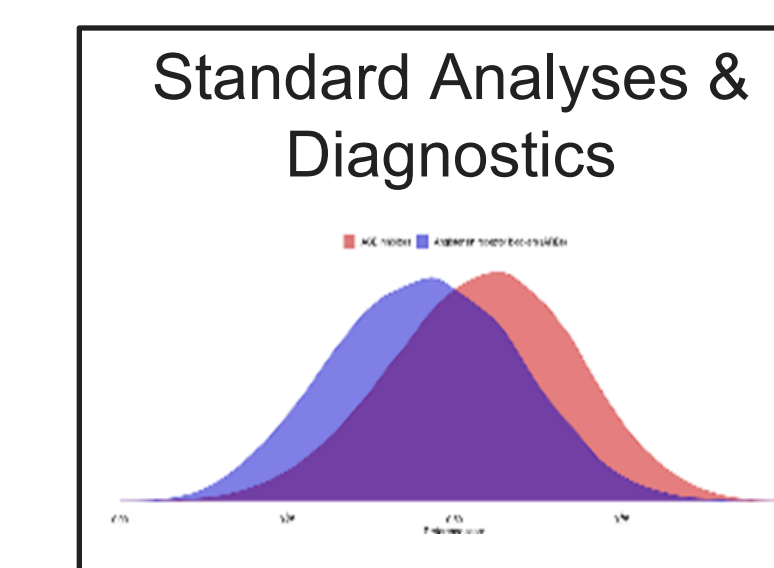
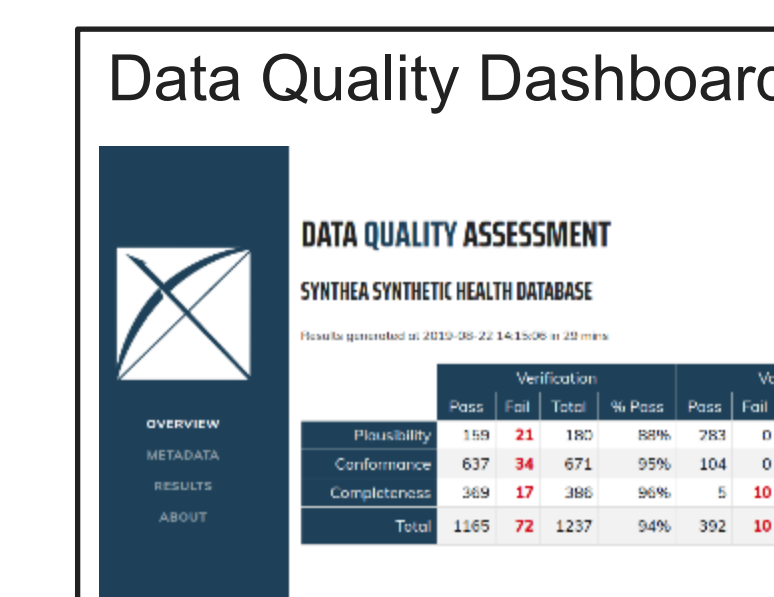
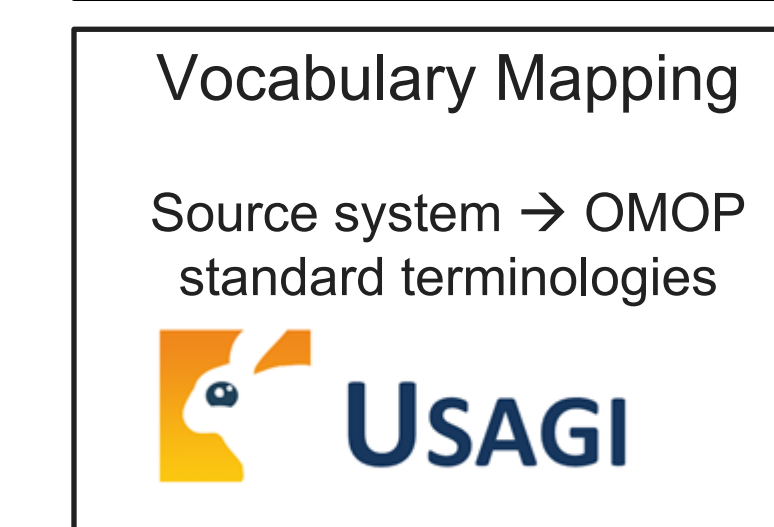
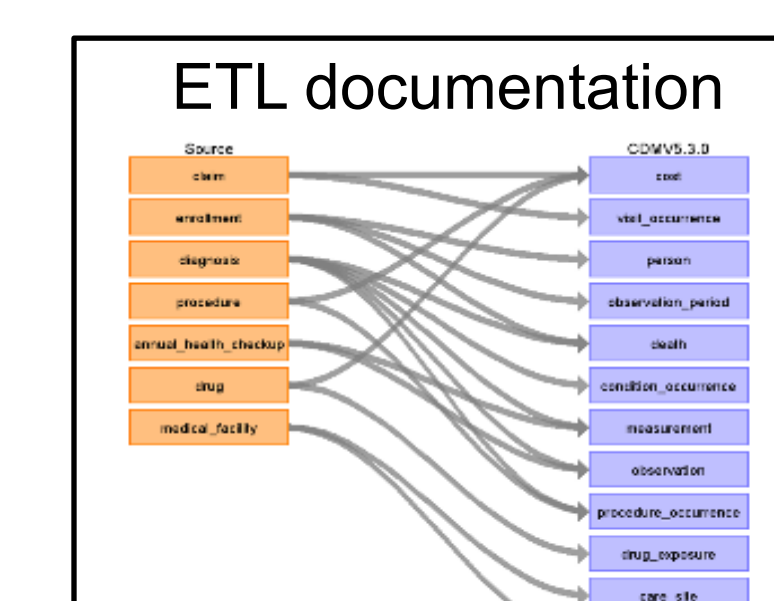


Table 2: Recommendations for Documentation of Data Quality and Analysis Reporting

Best Practice: Publicly available documentation or documentation provided upon request	Source or Tool to Facilitate Implementation
ETL Logic Documentation	White Rabbit Rabbit in a Hat
Local vocabulary mappings	Usagi
Data quality report	Data Quality Dashboard
Metadata about the source data	https://github.com/ohdsi/etl-lambda-builder ETL provides build process information.
Structured reporting of analysis decisions and diagnostics	RECORD-PE checklist [8] STaRT-RWE template [9] Tripod statement (prediction models)
Pharmacoepidemiology analysis programming code and diagnostics	Post to GitHub or provide upon request

Conclusions

- Standards for RWD quality, analysis, and reporting exist, but may need amending as use of RWD grows substantially.
- Transparency in reporting is paramount to evaluating RWD quality and analysis validity. Data use agreements and/or confidentiality considerations may impinge upon publicly available data and documentation; however, reviewers and journal editors must be provided with some minimal set of documentation to ascertain data provenance and quality for new sources.
- OHDSI tools and corresponding best practices will allow the opportunity for us to lead in both defining and applying best practices in RWD quality and analysis reporting.

References

- Brainard J. Science Magazine. Scientists are drowning in COVID-19 papers. Can new tools keep them afloat? May 13, 2020. (Accessed June 23, 2020). <https://www.sciencemag.org/news/2020/05/scientists-are-drowning-covid-19-papers-can-new-tools-keep-them-afloat>
- Fraser N, et al. Preprinting a pandemic: the role of preprints in the COVID-19 pandemic. (accessed June 23, 2020). Biorxiv. https://www.biorxiv.org/content/10.1101/2020.05.22.111294v1#disqus_thread
- Mehra MR, et al. Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. Lancet. 2020 May 22:S0140-6736(20)31180-6. doi: 10.1016/S0140-6736(20)31180-6. Online ahead of print.
- Mehra MR, et al. Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. N Engl J Med. 2020 May 1;NEJMoa2007621. doi: 10.1056/NEJMoa2007621. Online ahead of print.
- Retraction Watch Database (2020) Available at: <http://retractiondatabase.org/RetractionSearch.aspx> (Accessed 22 Sept 2020).
- Brainard J, et al. Science Magazine. October 18, 2020. What a massive database of retracted papers reveals about science publishing's 'death penalty'
- Chaccour C. Ivermectin and COVID-19: How a Flawed Database Shaped the Pandemic Response of Several Latin-American Countries //www.isglobal.org/en/healthisglobal/-/custom-blog-portlet/ivermectin-and-covid-19-how-a-flawed-database-shaped-the-covid-19-response-of-several-latin-american-countries/2877257/0 (Accessed 22 Sept 2020)
- Langan SM, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). BMJ 2018;363:k3532
- Wang SV, et al. A structured template and reporting tool for real-world evidence studies to facilitate unambiguous communication of methods and validity assessment (STaRT-RWE), <https://doi.org/10.7910/DVN/6R1KCA> (Accessed 24 July 2020).