

# **“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, PhD<sup>1</sup>, Erica A Voss, MPH,<sup>1,2</sup> Clair Blacketer, MPH<sup>1,2</sup>

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

[rdisanto@its.jnj.com](mailto:rdisanto@its.jnj.com); [mblacke@its.jnj.com](mailto:mblacke@its.jnj.com); [evoss3@its.jnj.com](mailto:evoss3@its.jnj.com)

Submission Category: Observational Data Management

## Abstract

**Introduction:** COVID-19 publications may represent the largest growth of scientific literature ever, with an estimated >~16-23,000 papers including >~6,000 preprints through April and projected to double every few weeks [1,2]. Unfortunately, this analysis proliferation in the advance of well-controlled trials includes high-profile COVID-19 manuscript retractions in the Lancet and NEJM (i.e., Surgisphere data concerns), which has the potential to erode trust in real-world data (RWD) [3,4].

**Objectives:** Study objectives were to: 1) analyze manuscript retractions with a focus on COVID-19 and 2) propose RWE best practices based on insights from these retractions.

**Methods:** Using the RetractionWatch® database, we described reasons for COVID-19 manuscript retractions [5]. Based on our assessment, we identified modifications to current practices that address retraction reasons, with a focus on transparency and data quality assessment using OHDSI tools.

**Results:** There were 22 retracted COVID-19 publications through June 23, 2020, with the most common retraction reasons relating to data and analysis issues (Table 1). For the symposium, COVID-19 retraction summaries will be updated with new retractions and in-depth assessment of RWD-related retractions.

**Table 1. Characteristics of COVID-19 Retractions through June 23, 2020**

Characteristic	N=22 Retractions Descriptive Statistics	
<b>Time: Publication to Retraction (Days)</b>		
Mean, Median	19.8, 9	
Min, Max	0, 92	
Total Non-Missing	21	
Description of Retraction or Retraction Reason (multiple responses allowed)	Frequencies	
	Count	Percent
Has a data issue*	7	32
Has an analysis, result, or conclusion issue*	7	32
Limited information available	7	32
No notice of withdrawal	5	23
Duplicate publication (in error)	3	14
Authorship concerns	2	9
Ethical concerns	1	5
Status at Withdrawal	Count	Percent
In press	2	9
Not Prepared	1	5
Preprint	8	36
Published	11	50
Country	Count	Percent
China	8	36
United States	6	27
Others (twice: India, France; once: Bangladesh, Indonesia, Iran, S. Korea)	8	36

Note: Percentages may not add to 100 for mutually exclusive categories due to rounding.

\*Issues for data includes concerns, reliability and errors. Issues for analysis, results, and conclusions include concerns and errors.

Generally, COVID-19 articles were reviewed, published, and retracted more quickly than historical retractions [6]. Information about some retractions was sparse. Best practices motivated by this retraction analysis include publicly-available documentation to achieve unambiguous reporting of data transformations, data quality assessment, analysis decisions, and programming code (Table 2). These practices allow for more thorough assessment of RWD quality and RWE validity.

**Table 2. Documentation and Data Quality and Analysis Reporting Recommendations**

Best Practice: Publicly available documentation	Source: Example	Source or Tool to Facilitate Implementation												
ETL Logic Documentation: available publicly or upon request by reviewers	<a href="https://ohdsi.github.io/ETL-LambdaBuilder/docs/Optum_Clinformatics">https://ohdsi.github.io/ETL-LambdaBuilder/docs/Optum_Clinformatics</a>	White Rabbit Rabbit in a Hat												
Local vocabulary mappings	<a href="https://github.com/OHDSI/ETL-LambdaBuilder/tree/master/docs/Optum%20Clinformatics/Vocab%20Updates">https://github.com/OHDSI/ETL-LambdaBuilder/tree/master/docs/Optum%20Clinformatics/Vocab%20Updates</a>	Usagi												
Data quality report	<a href="https://data.ohdsi.org/DataQualityDashboard/">https://data.ohdsi.org/DataQualityDashboard/</a>	Data Quality Dashboard												
Metadata about the source made available to reviewers	Example: Count of persons removed from CDM <table border="1"> <thead> <tr> <th>Reason</th> <th>Number Persons removed</th> <th>Date</th> </tr> </thead> <tbody> <tr> <td>Unknown gender</td> <td>22296</td> <td>2020-04-2</td> </tr> <tr> <td>Implausible year of birth - past</td> <td>642</td> <td>2020-04-2</td> </tr> <tr> <td>Implausible year of birth - post earliest observation period</td> <td>1868</td> <td>2020-04-2</td> </tr> </tbody> </table>	Reason	Number Persons removed	Date	Unknown gender	22296	2020-04-2	Implausible year of birth - past	642	2020-04-2	Implausible year of birth - post earliest observation period	1868	2020-04-2	<a href="https://github.com/ohdsi/etl-lambdabuilder">https://github.com/ohdsi/etl-lambdabuilder</a>  ETL information provides information about the build process.
Reason	Number Persons removed	Date												
Unknown gender	22296	2020-04-2												
Implausible year of birth - past	642	2020-04-2												
Implausible year of birth - post earliest observation period	1868	2020-04-2												
Structured reporting of analysis decisions and diagnostics	<a href="https://www.tripod-statement.org/about/">https://www.tripod-statement.org/about/</a>	RECORD-PE checklist [7] STaRT-RWE template [8] Tripod statement (prediction models)												
Pharmacoepidemiology analysis programming code posted or upon request by reviewers	<a href="https://data.ohdsi.org/OhdsiStudies">https://data.ohdsi.org/OhdsiStudies</a>	Post to GitHub or provide to editor upon request												

**Conclusions:** Standards for RWD quality, analysis, and reporting exist [9-12], but amendments are critical as use of RWD grows substantially. 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

1. 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>
2. Fraser N, Brierley L, Dey G, Polka JK, Pálffy M, Coates JA. 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](https://www.biorxiv.org/content/10.1101/2020.05.22.111294v1#disqus_thread)
3. Mehra MR, Desai SS, Ruschitzka F, Patel AN. 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.
4. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. 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.
5. Retraction Watch Database (2020) Available at:  
<http://retractiondatabase.org/RetractionSearch.aspx> (accessed 17 June 2020).
6. Brainard J, You J. Science Magazine. October 18, 2020. What a massive database of retracted papers reveals about science publishing's 'death penalty'  
<https://www.sciencemag.org/news/2018/10/what-massive-database-retracted-papers-reveals-about-science-publishing-s-death-penalty> (Accessed June 23, 2020)
7. Langan SM, Schmidt SAJ, Wing K, Ehrenstein V, Nicholls SG, Filion KB, Klungel O, Petersen I, Sorensen HT, Dixon WG, Guttman A, Harron K, Hemkens LG, Moher D, Schneeweiss S, Smeeth L, Sturkenboom M, von Elm E, Wang SV, Benchimol EI. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). BMJ 2018;363:k3532
8. Wang SV, Pinheiro S, Hua W, Arlett P, Uyama Y, Berlin JA, Bartels D, Kahler K, Bessette LG, Schneeweiss S, 2020, 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 July, 2020).
9. International Society for Pharmacoepidemiology. Guidelines for Good Pharmacoepidemiology Practices (GPP). 2015. (Accessed 17 June 2020).  
<https://www.pharmacoepi.org/resources/policies/guidelines-08027/>
10. Hall GC, Sauer B, Bourke A, Brown JS, Reynolds MW, Lo Casale R. Guidelines for good database selection and use in pharmacoepidemiology research. Pharmacoepidemiol Drug Saf. 2012 Jan;21(1):1-10. doi: 10.1002/pds.2229. Epub 2011 Nov 8.
11. Duke Margolis White Paper. Characterizing Real-World Quality and Relevancy for Regulatory Purposes. October 1, 2018. (Accessed 17 June 2020).  
[https://healthpolicy.duke.edu/sites/default/files/atoms/files/characterizing\\_rwd.pdf](https://healthpolicy.duke.edu/sites/default/files/atoms/files/characterizing_rwd.pdf)
12. Duke Margolis White Paper. Determining Real-World Data's Fitness for Use and the Role of Reliability. September 26, 2019. (Accessed 17 June 2020).  
[https://healthpolicy.duke.edu/sites/default/files/u31/rwd\\_reliability.pdf](https://healthpolicy.duke.edu/sites/default/files/u31/rwd_reliability.pdf)