
Rachael L. DiSantostefano, PhD; Erica A Voss, MPH; Claire Blacketer, MPH
Janssen Research and Development, Titusville, NJ. OHDSI 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

1) Analyze manuscript rejections with a focus on COVID-19 and
2) propose RWE best practices based on insights from these rejections

Methods

Database

• RetractionWatch® database (RWDB), largest curated database of manuscript rejections [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

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Retractions</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to Retraction (Days)</td>
<td>0-94</td>
<td>Min, Max</td>
</tr>
<tr>
<td>Subset of Retraction Reasons</td>
<td>Count</td>
<td>Percent</td>
</tr>
<tr>
<td>Data issue*</td>
<td>8</td>
<td>22</td>
</tr>
<tr>
<td>Analysis, result, or conclusion issue*</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>Limited or no withdrawal information</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>No notice of withdrawal</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>Lacking IRR, 3rd party, author approval</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>Duplicate publication in error</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Ethical concerns</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Status at Withdrawal</td>
<td>Count</td>
<td>Percent</td>
</tr>
<tr>
<td>In press</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Not pressed</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Preprint</td>
<td>12</td>
<td>33</td>
</tr>
<tr>
<td>Published</td>
<td>20</td>
<td>55</td>
</tr>
</tbody>
</table>

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 primarily addressed by requiring transparent, unambiguous reporting of data preparation & analysis. This allows robust assessment of data quality & analysis validity (Table 2).

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