How to use the regulatory data from Health Canada for secondary analyses on new drugs, biologics and vaccines

Isaac Bai, Peter Doshi, Matthew Herder

Abstract
Incorporating clinical data held by national health product regulatory authorities into secondary analyses such as systematic reviews can help combat publication bias and selective outcome reporting, in turn, supporting more evidence-based decisions regarding the prescribing of drugs, biologics and vaccines. Owing to recent changes in Canadian law, Health Canada has begun to make clinical information—whether it has been previously published or not—publicly available through its ‘Public Release of Clinical Information’ (PRCI) online database. We provide guidance about how to access and use regulatory data obtained through the PRCI database for the purpose of conducting drug and biologic secondary analyses.

Introduction
Secondary analyses such as systematic reviews and trial reanalyses are usually conducted using publicly reported data (especially journal publications). However, journal publications often contain relatively limited information, for example, about the design of a clinical trial. Further, factors such as publication bias and selective outcome reporting often lead to overestimating efficacy and understating the harms.1 2 Statistical methods exist to detect the likelihood of publication bias, but no single test can reliably rule it out.3 Direct access to unpublished data is arguably the only way to fully overcome publication bias. Selective reporting bias (biases that distort the proper interpretation of published trials) generally can only be detected by comparing two or more sources of information (eg, journal publication vs ClinicalTrials.gov vs regulatory data).4 Yet, registration on ClinicalTrials.gov is only mandatory for trials of FDA-regulated products with at least one trial site in the USA,5 and compliance with registration and results reporting requirements remains imperfect. For instance, one study showed that 59% of trials had not posted results on ClinicalTrials.gov 2 years after completion.6 Another study of 400 trials registered on ClinicalTrials.gov found only 58% had their primary outcomes adequately described.7

In this environment, health product regulators have emerged as an important source of data for secondary analyses.8 9 Since 1998, the US Food and Drug Administration (FDA) has published its scientific reviews on drug approval10 11 through its Drugs@FDA database. These reviews frequently contain more data than published literature, including key details of trial designs and outcomes, and the regulator’s interpretations of a drug’s safety, efficacy and quality. More recently, other regulators, namely, the European Medicines Agency (EMA), the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan and Health Canada (HC), have been empowered to publicly disclose portions of the regulatory dossier submitted by manufacturers when seeking regulatory approval. While Drugs@FDA often contains important elements from a regulatory dossier, such as trials’ summary results statistics and additional FDA-initiated analyses,12 the EMA, PMDA and HC publish the original industry documents submitted to the regulator—for example, Clinical Summaries, Clinical Study Reports (CSRs), adverse event narratives, trial protocols and statistical analysis plans—essentially in their entirety although with some redactions (table 1).

Despite the wealth of data that can now be secured from regulators, there is little sign that researchers engaged in systematic reviews and other secondary analyses routinely incorporate...
This likely stems from insufficient awareness of document availability as well as practical barriers to using regulatory-sourced data. Reports obtained from regulators can range up to thousands of pages in length, which may require tremendous time and resources from systematic reviewers. However, these barriers can be mitigated, in part, with guidance about how to access and use these data. In addition, when research misconduct is identified by regulators, there does not appear to be communication between the regulators and journals, as one study found no subsequent corrections to 57 studies published in the literature that FDA identified as having research conduct issues. In another instance, the publication of a major cardiovascular drug trial went uncorrected even after FDA reviewers judged the data unreliable a decade prior. These are additional reasons to have access to regulatory data, even for published studies.

To date, efforts have been made to provide experience and guidance on the use of regulatory data that were obtained through the Drugs@FDA database and EMA’s Policy 0070. Apart from HC’s own highly technical guidance, however, there is no published literature that explains the scope and holdings of HC’s ‘Public Release of Clinical Information’ (PRCI) online database. This is a critical gap in the literature because the PRCI database has important comparative advantages. First, Canada is releasing not only approved products but also ones rejected from market entry (table 1). Second, Canada’s holding include medical devices, not just drugs and biologics. First, Canada is releasing not only approved products but also ones rejected from market entry (table 1). Second, Canada’s holding include medical devices, not just drugs and biologics. Third, Canada’s holding include medical devices, not just drugs and vaccines. This Research Methods and Reporting article aims to fill this gap in the literature, detailing what the Canadian database contains and how it can be accessed.

**Methods**

The PRCI portal was comprehensively reviewed in March 2022, and 203 clinical information packages pertaining to drugs and biologics (excluding medical devices) were identified that were publicly released between March 2019 and March 2022 via the HC PRCI database. Information contained in the drug/biologic packages was charted based on their contents, submission types, authorisation types, initial regulatory decision dates and public release dates (table 2). Relevant policy materials such as HC’s PRCI guidance document and the terms and conditions page from the PRCI database was reviewed and summarised to provide further
context to the use of the information contained in the database. This article will explain the scope, terms of use and steps for accessing regulatory data from the database.

### Results

**What is contained in the database (what is the scope of PRCI)?**

Table 3 provides a list of the typical documents released on PRCI. This means a wealth of clinical data is available for approved, rejected and withdrawn products, and anyone can access clinical information through the database without any sort of formal registration or application process. To access the data, a user must accept the terms of use, but the architects of the portal have made this a smooth and easy process in an effort to ‘enable independent reanalysis of data and generate new research questions’.

PRCI’s data availability is not restricted to a specific timeframe. Products approved, rejected or withdrawn prior to 2019 when the portal was launched can be disclosed following a request made via the portal. Approximately a third of the products listed on the database predates 2019, with one (metronidazole) dated as early as 1960. However, like submissions post-March 2019, once they are released, anyone anywhere in the world can access them for free via HC’s portal website. Currently, more packages are available for post-2019 decisions, automatically published on reaching a decision. For example, new active substances versus a generic drug submission.

Table 2 Characteristics of the 203 clinical information packages released between March 2019 and March 2022

<table>
<thead>
<tr>
<th>Number of packages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total clinical information packages</td>
</tr>
<tr>
<td>By product type</td>
</tr>
<tr>
<td>Small molecule drug</td>
</tr>
<tr>
<td>Biologic</td>
</tr>
<tr>
<td>Vaccine</td>
</tr>
<tr>
<td>Radiopharmaceutical</td>
</tr>
<tr>
<td>By decision date</td>
</tr>
<tr>
<td>Before 2003</td>
</tr>
<tr>
<td>2003–2018</td>
</tr>
<tr>
<td>2019–present</td>
</tr>
<tr>
<td>By submission type</td>
</tr>
<tr>
<td>New applications (NDS, NDS-NAS)</td>
</tr>
<tr>
<td>Supplemental applications (SNDS, SNDS-c)</td>
</tr>
<tr>
<td>Generic drug applications (ANDS)</td>
</tr>
<tr>
<td>Interim order</td>
</tr>
<tr>
<td>By authorisation type</td>
</tr>
<tr>
<td>NOC</td>
</tr>
<tr>
<td>Authorised with conditions</td>
</tr>
<tr>
<td>NOD-W</td>
</tr>
</tbody>
</table>

ANDS, abbreviated new drug submission; NDS, new drug submission; NDS-NAS, new drug submission-new active substance; NOC, notice of compliance; NOD-W, notice of deficiency-withdrawal; SNDS, supplement to a new drug submission; SNDS-c, supplement to a new drug submission containing confirmatory trials.

**What does each clinical information package contain?**

Generally, each package contains three main types of documents: a Clinical Overview, Clinical Summaries and CSRs. All three of these documents, especially CSRs, can be useful for secondary research and tend to provide more accurate and complete information about a drug’s safety and/or efficacy profile relative to corresponding studies in the published literature.

First, the Clinical Overview is usually a single PDF document containing a high-level summary of relevant information pertaining to the drug of interest, including but not limited to the product development rationale, biopharmaceutics, pharmacology, efficacy/safety studies and benefit/risk assessment. This section should provide a complete list of all clinical studies of the drug, since the initiation of premarketing studies almost always requires regulatory approval. Information presented in this module can be cross-referenced to more detailed information found in subsequent modules.

Second, Clinical Summaries are usually divided into four parts: the Summary of Biopharmaceutic Studies and Associated Analytical Methods; the Summary of Clinical Pharmacology Studies; the Summary of Clinical Efficacy; and the Summary of Clinical Safety. Each part typically contains one single PDF document, and provides a history of methods, results and conclusions of all included studies for the submission, at the level more detailed than the Clinical Overview. More detailed information is cross-referenced to CSRs.

Third, CSRs contain unabridged reports of individual studies pertaining to the indication submitted to HC for marketing approval. Each individual CSR may contain separate PDF files for the main report body, the protocol and amendments, statistical methods, sample case report forms, and various adverse event narrative files.

These documents can be expected to be the same as those submitted to other regulators such as FDA and EMA, particularly since 2003 after which submissions will typically be organised to follow the International Council on Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) Common Technical Document (CTD) and Electronic Common Technical Document (eCTD) format.

The organisation of each package varies, in part, because the amount of information submitted by manufacturers to HC itself varies due to varying regulatory requirements applicable to, for instance, new active substances versus a generic drug submission. The packages also vary because the format of product submissions has evolved over time; more variation is seen prior to the 2003 adoption of the ICH standards.

For drugs that received final regulatory decisions before 2003, the clinical information packages may present differently. Older packages tend to contain scanned PDF documents that may not be searchable without first performing optical character recognition (OCR). However, manual OCR’ing of documents may not be required as some newer computers and software automatically perform OCR on scanned PDF documents without special OCR software.

Information contained in the packages is subject to anonymisation and redaction, primarily to protect personal information and reduce the risk of reidentification of study participants. The anonymisation report is also disclosed as part of each package posted on the PRCI portal. There are two categories of clinical information that can be redacted by manufacturers (box 1). However, HC has the discretion to disclose these two categories of information as part of a submission packages unless
manufacturers provide adequate justification. It is worth noting that manufacturers are responsible for redaction and submitting the anonymised reports, and redactions may not be consistent across different manufacturers (box 2).

What information can be used for the purposes of secondary analysis?
Since the HC clinical information packages may contain multiple PDFs with thousands of pages, it is important to locate the information pertinent for the purpose of your secondary analysis. For researchers looking for information on efficacy and safety data, Summary of Clinical Efficacy and Summary of Clinical Safety, respectively, provide an overview and a list of all clinical studies (controlled and uncontrolled, and irrespective of whether the study has been published in the biomedical literature) submitted to HC for marketing decision. Therefore, it can serve as an excellent source when conducting a systematic search of literature. Additionally, these modules can be used to identify any potentially unpublished trials. For individual studies including their report bodies, protocols and amendments, and statistical methods, data can be found in CSRs (generally found under ‘Reports of Efficacy and Safety Studies’). CSR and their numerous appendices contain substantially more details than the Summary modules and are suitable for critical appraisal of individual studies including the statistical methods, obtaining data from the studies for syntheses, and assessing changes in the protocols such as selective outcome reporting. Some packages might contain Reports of Postmarketing Experience (listed as part of a ‘Post-Authorization Activity Table’), which can be used to assess long-term safety outcomes. However, at the time of writing (April 2022), there is only one package available through the portal (for the drug mescarmin) that contains this type of information. Typically, additional data from postmarketing studies and/or pharmacovigilance must be secured independently from a different part of HC’s website.

For biopharmaceutics and analytical methods, data can be found in the Summary of Biopharmaceutic Studies and Associated Analytical Methods and Reports of Biopharmaceutic Studies. For information on pharmacology, data can be found in the Summary of Clinical Pharmacology Studies, Reports of Studies Pertinent to Pharmacokinetics Using Human Biomaterials, Reports of Human Pharmacokinetic (PK) Studies and...
Reports of Human Pharmacodynamic (PD) Studies (table 3). These two sections may not be relevant for most researchers looking for clinical efficacy and safety data. However, the biopharmaceutics modules may contain data on drug interactions and the pharmacology modules may contain data on initial tolerability, which can both be used to supplement assessment on drug safety. Additionally, immune response data for biologics and vaccines can be found in the pharmacology modules.

Regardless of the type of data you are looking for, Reports of Analyses of Data from More than One Study and Other Study Reports may contain relevant information that are not found in other modules, including but not limited to data from non-pivotal trials, trials for a different population, or trials for a different indication.

How do I access the data?
To access and download clinical information packages on HC’s website:
1. Start by going to the Health Canada Public Release of Clinical Information (https://clinical-information.canada.ca) website (online supplemental figure 1).
2. The packages can be accessed either directly from the list (online supplemental figure 2, box) or through the search (online supplemental figure 2, circle). Either a brand name or generic name can be used for searches. Search results can be narrowed using the filters. In this case, remdesivir is used as an example.
3. There may be multiple search results for a single drug, with each representing a separate submission/indication to HC for marketing approval, denoted by a unique ‘submission control number’ (online supplemental figure 3).
4. Inside the packages, each link represents a separate PDF file that can be downloaded. If desired, all PDFs within the same package can be downloaded all at once as a compressed ZIP file under ‘Submission archive’—‘Download ZIP’ (online supplemental figure 4).

Clinical information for products regulated by HC but not on the website and not in progress of release (online supplemental figure 5) can be requested through the website, by providing the following information: drug name (either brand or generic name), manufacturer name (ie, the manufacturer that owns the Drug Identification Number or Notice of Compliance), indication(s)/use of the drug (can be more than one), request of Clinical Information from the initial marketing approval and/or postauthorization activities, types of studies (adult, paediatric, phase I/II/III studies, effectiveness/safety studies or any combination of the above), reason for request and any additional information and the name and email address of the requester. The request form can be accessed through the front page of the website (online supplemental figures 6 and 7). Importantly, researchers seeking data from the portal do not need to provide a detailed reason or explain their proposed research plans. Rather, they only need to indicate the broad purpose of the research to ensure that the data are not being sought for a commercial purpose.

What are the limitations of the clinical information packages?
While the clinical information packages may contain a large amount of summary-level data, they do not normally provide individual patient-level datasets (IPD) that could be readily used with statistical software, as HC generally does not hold IPD. Some IPD is present in the form of patient-level details found in PDFs, such as adverse event narratives, and concomitant medication lists, and some CSR appendices. This means that researchers

**Box 1** Historical context of the PRCI database

Historically, information supplied by a manufacturer to HC was, as a matter of practice, treated as ‘confidential business information’ (CBI) that could not be publicly disclosed. However, significant changes brought about by Vanessas’s Law, which Canadas’s Parliament enacted in 2014, gave Health Canada (HC) the power to prescribe information as falling outside the scope of CBI. In 2019, HC used this new power to create a set of regulations that deemed information that had been considered CBI to no longer fall into that category once a decision to approve, reject or withdraw a drug, or modification of the PRCI portal.

Notably, some information continues to be treated as CBI; namely, (1) clinical information that was not used in the drug submission for the claimed indication (eg, exploratory outcome data to support future trials in obtaining a new indication); and (2) clinical information that describes tests, methods or assays that were used exclusively by a given manufacturer. According to HC policies, manufacturers are responsible for preparing the redaction version of documents and to provide an anonymisation report. A recent study found that the redactions in materials released by both HC and EMA primarily pertained to identifying details of trial investigators and participant ID numbers; such redactions may prevent certain analyses but generally still allow for the reanalysis of trial data. It remains to be seen how widely or narrowly these exceptions from disclosure are being relied upon by manufacturers and/or granted by HC.

**Box 2** Discrepancy in redactions of similar information

Below is an example showing the discrepancy in redaction practices between Moderna and Pfizer for a similar adverse event narrative file. Moderna did not redact participant details (first picture) whereas Pfizer did (second picture).

1. Screenshot of an adverse event narrative for the Moderna COVID-19 vaccine (mRNA-1273).
2. Screenshot of an adverse event narrative for the Pfizer COVID-19 vaccine (BNT162b2).
aiming to conduct an IPD meta-analysis will generally find the clinical information packages to not be appropriate as a source of information. The only exception is if the data are available in PDFs, in which case the researchers would first have to convert those patient-level details into analysable datasets.

Additionally, as the packages only contain data submitted by manufacturers for marketing approval, detailed reports of studies conducted by parties other than the sponsor are not included. This limitation is especially important for researchers conducting secondary analyses on older drugs, as there may be numerous studies conducted by parties other than the sponsor that are not contained in the packages.

Overall, for studies that are in the regulator’s holdings, the packages provide a far more granular and reliable source of data than peer-reviewed literature.

Key challenges
While the PRCI database contains regulatory data that can be accessed freely by researchers, uptake appears low.2-3 Given its growing volume and advantages, it is important to increase the awareness of the PRCI portal. At the same time, the value of the portal (in terms of containing evidence from the premarket phase of a drug’s development) may be limited as regulators increasingly rely on postapproval evidence.27 The data contained in the PRCI database will need to be updated and/or connected to other sources of information held by HC as evidence evolves, to ensure data contained are complete and up-to-date when researchers access the data.

Contributors IB wrote the first draft with input from MH, and all authors (IB, PD, MH) made critical revisions to the manuscript. All authors (IB, PD, MH) read and approved the final version of the paper. IB obtained funding for this study. IB is the guarantor for the study.

Funding The study is funded by the Dalhousie Medical School Research in Medicine Studentship.

Competing interests IB has no competing interest to disclose.

PD reported running the RIA Support Center, which was funded by Laura and John Arnold Foundation (2017–2022), and supported part of his university salary. He also reported receiving US FDA funds (via University of Maryland M-CERSI) for supporting graduate students (2020), being an unpaid IMEDS steering committee member (2016–2020), and is an editor at The BMJ (ongoing). In the past 36 months, MH reported being a member of the Patented Medicine Prices Review Board, Canada’s national drug price regulator, and receiving honoraria from the Board for his service, and received grants from CIHR (PJT 156256) as well as CIHR and the Public Health Agency of Canada.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the context includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Isaac Bai http://orcid.org/0000-0003-4812-8428
Peter Doshi http://orcid.org/0000-0002-7804-4113

References
8. Loder E, Tovey D, Gidlow E. The Tamiflu trials. BMJ 2014;348:g2630.
21 Lexchin J, Herder M, Doshi P. Canada finally opens up data on new drugs and devices. *BMJ* 2019;365:k1825.
Figure 1. Main page of The Health Canada Clinical Information on Drugs and Medical Devices
Figure 2. Access clinical information packages that have been released either through drug search (circle) or from the list (box).
Figure 3. Search results for remdesivir
Figure 4. Table of contents for remdesivir (submission control number 240551).

Figure 5. Link to the list of clinical information packages in progress of release
Figure 6. Link to the request page for clinical information packages that have yet to be released
Figure 7. Request page for clinical information packages