Pharmaceutical clinical trials transparency and privacy

Sarah Lyons[1], **Vivien Fagan**[2]
[1] Privacy Analytics, Ottawa, Canada
[2] IQVIA, Livingston, West Lothian, UK

Correspondence to:
Sarah Lyons
General Manager, Privacy Analytics
251 Laurier Avenue, Suite 200
Ottawa, ON, Canada K1P5J6
slyons@privacy-analytics.com

Abstract
With the introduction of new clinical trial transparency regulations around the world, transparency functions have had to adapt to a range of reporting requirements. In 2007, the FDA Amendments Act (FDAAA) established requirements for trial sponsors to reveal trial results to participants and patients, physicians, and independent researchers. Since then, more requirements have emerged, including anonymisation and publication policies introduced by the EMA and Health Canada. Going beyond regulatory compliance, transparency leaders have launched voluntary data-sharing initiatives to enable researchers' access to structured individual patient data. With this move toward greater transparency and the drive for more data, transparency functions working with the clinical trials environment need a broader toolkit of capabilities, including anonymisation, to protect participants' privacy. The authors explored these emerging trends in a webinar for FDANews on July 23, 2020 (https://www.fdanews.com). FDANews has an 80,000-person database, mainly from the clinical trial space. This article summarises the webinar.

Clinical trial transparency goals
The clinical trial transparency landscape has evolved over the last decade, with rising expectations for openness and disclosure. The goals of greater clinical trial transparency are multiple, with benefits to trial participants, clinical trial sponsors, regulators, the scientific community, and, ultimately, patients. Examples of these benefits include:

- **Avoiding duplication.** Transparency can help ensure the right trials are conducted by informing funders and researchers on which trials are needed and avoiding research duplication.
- **Patient access.** Transparency can help potential clinical trial participants better understand their options to enrol in new trials.
- **Better decisions.** With more complete information available from trials, better decisions can be made by those using evidence from clinical trials.
- **Higher quality.** By enabling the scientific community to examine clinical research, engage in quality improvement, and identify gaps in data, a more robust quality-predicated system can be attained.
- **Trust.** Transparency can build trust between the public, sponsors, and regulators through greater openness and collaboration.

Transparency involves multiple points of disclosure prior to, during, and after the clinical trial:[1]

1. Registration on a publicly accessible registry, such as ClinicalTrials.gov
2. Posting of summary results in a timely fashion, using plain language
3. Making trial reports publicly accessible and publishing trial outcomes
4. Sharing individual participant data in a privacy-preserving manner

Several new regulations have emerged to enforce greater transparency, from trial registration to publication of the clinical study reports (CSRs).[2,3] In addition to regulatory shifts toward greater transparency, many sponsors are adopting discretionary measures, such as voluntarily sharing anonymised participant data with researchers.[4,5]

FDA requirements
The Food and Drug Administration Amendments Act of 2007[6] (known as FDAAA) established a requirement for certain (applicable) clinical trials to be registered at trial initiation and to report summary results after trial completion in the public registry and results database (ClinicalTrials.gov). This law was intended to facilitate enrolment in clinical trials, allow for tracking of the progress of such trials, and address problems with the lack of timely dissemination of research findings.

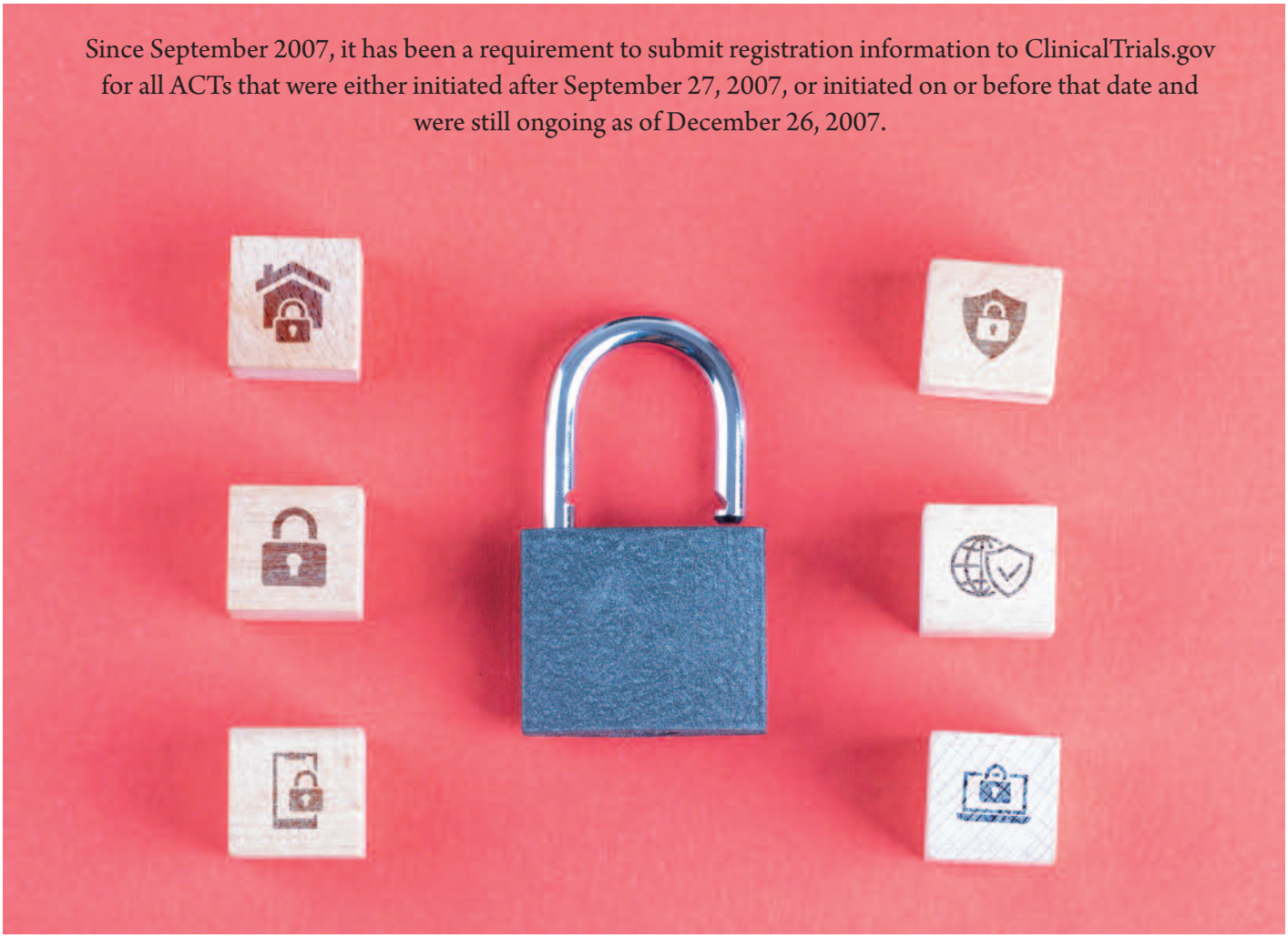
What is an applicable clinical trial (ACT)? ClinicalTrials.gov has a checklist for evaluating whether a clinical trial or study is an ACT.[7] In general, ACTs are trials of drugs and biologics: controlled clinical investigations, other than Phase 1 clinical investigations, of a drug, biological product, or medical device subject to FDA regulation, where a controlled clinical investigation generally includes interventional studies (with one or more arms) that meet one of the following conditions:

- have one or more sites in the US
- are conducted under an FDA investigational new drug application (IND) [or, in case of a device trial, investigational device exemption]
- involves a drug, biologic, or device that is manufactured in the US and is exported for research.

Since September 2007, it has been a requirement to submit registration information to ClinicalTrials.gov for all ACTs that were either initiated after September 27, 2007, or initiated on or before that date and were still ongoing as of December 26, 2007. This registration submission must be made no later than 21 days after enrolment of the first participant. Subsequent updates have included:[8]

- September 2008: the requirement to submit summary results for clinical trials of approved products within 12 months of the completion date (primary completion date [PCD]), where the PCD is the date of final data collection for the primary outcome measure(s) (OMs).
- September 2009: the requirement to include certain adverse event information in the summary results.
- September 2016: the Final Rule extended the requirement for results information submission to ACTs of a drug, biological product, or medical device that is *not* approved, licenced, or cleared by the FDA, thus alleviating concerns regarding bias in the

Since September 2007, it has been a requirement to submit registration information to ClinicalTrials.gov for all ACTs that were either initiated after September 27, 2007, or initiated on or before that date and were still ongoing as of December 26, 2007.



literature and possible selective publication of only those approved products.[9] The Final Rule, which took effect on January 18, 2017, also introduced the requirement to submit the full protocol and statistical analysis plan along with the final results posting.

- January 2019: the requirement to indicate whether there is a plan to make individual participant data collected in the study, including data dictionaries, available to other researchers (typically after the end of the study).
- January 2020: all results postings (trials with a start date on or after January 18, 2017, with first submitted results information on or after January 01, 2020) will be publicly posted within 30 days of submission, regardless of a completed review process by ClinicalTrials.gov (Protocol Registration and Results System [PRS] review), including any brief Quality Control (QC) comments that identify at least one major issue (major issues identified in the comments must be corrected or addressed), along with a note that the QC has not concluded. All versions of the QC reviewed record will then be posted until the review process concludes.

ClinicalTrials.gov common pain points

Within the results database, examples of common pain points include, but are not limited to:

- Changes made to the text in the treatment arm descriptions are not carried throughout the tabulated database, so any changes made to the descriptive text must be manually repeated throughout the database each time the treatment arm is presented.
- A similar issue exists if there is a need to repeat a statistical analysis: no option to copy.
- Unable to have multiple units of measure within an OM, e.g., for a table presenting pharmacokinetic results. The only option is to split the OM over multiple OMs, i.e., by unit.
- When a study has two or more periods: ClinicalTrials.gov dictates that all treatment arms are repeated for all periods, utility would be improved if there was the option to select different treatment arms for multiple periods.

With regard to the PRS review/QC, examples of common pain points include, but are not limited to:

- Contributors may use a "lesson learned" approach from a previous posting to guide addressing a similar scenario in a different posting; however, that does not necessarily mitigate for conflicting PRS comments. This lack of consistency is also apparent when results are resubmitted following updates due to PRS comments. It would be advantageous to have the same PRS reviewer for re-submitted results to avoid the scenario where a different reviewer raises a separate issue with the resubmitted results that was not raised at the initial submission.
- Despite the clarity of the definition between "major" and "advisory" issues, there is inconsistency across studies in the ranking of the identified issues. Each major issue must

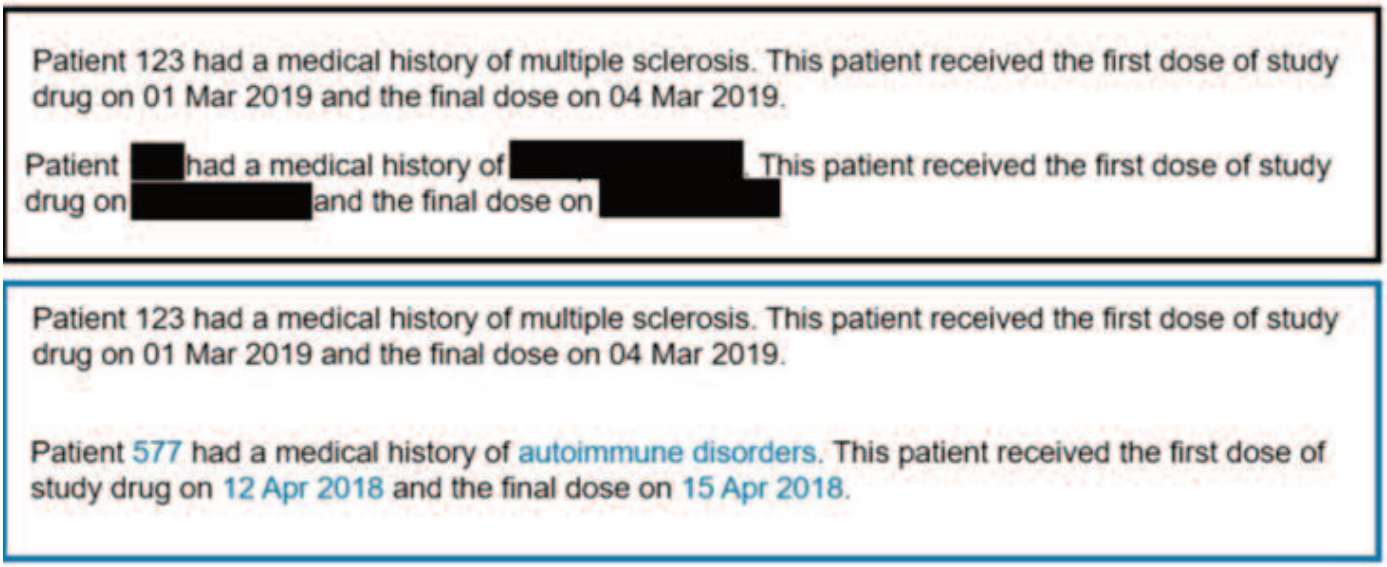


Figure 1. Visual comparison of redaction and anonymisation
Redaction of text shown above, where an opaque box obscures the information. Anonymisation of the same text shown below. A combination of pseudonymisation, generalisation, and date shifting are used to transform the text. The transformations are highlighted in blue in the figure for illustrative purposes only, whereas actual anonymisation benefits from the privacy concept of 'hiding in plain sight' whereby the information altered is not discernible.

be corrected or addressed, while advisory issues are suggestions only for improving the clarity of the record.

ClinicalTrials.gov best practices

In terms of clinical trial registration, submitting registration information to the PRS is relatively straightforward since there is typically limited interpretation required if the protocol is complete. The required content includes descriptive information, recruitment information, location and contact information, and administrative data elements.

The results section, on the other hand, while straightforward when it comes to entering appropriately compiled data, is often more complex, so it is important to assign the appropriate personnel. The preparation of clinical trial results posting is more than just an administrative task and is more suited to someone familiar with study protocols and CSRs and has experience in summarising clinical trial data. Lessons learned have suggested basic results entries have fewer errors and quality review comments from ClinicalTrials.gov when the appropriate person (e.g., medical writer) is tasked with preparing results information for submission.

Future considerations of clinical trial registration and results information submission

In 2019, the National Library of Medicine launched the ClinicalTrials.gov modernisation effort, which included a request for information from the public to guide efforts to enhance and better support the users of ClinicalTrials.gov.

All responses from the public were to be received by March 14, 2020, and were both published and shared via a public meeting in April 2020.[10] At the time this article went to press, the outcome of the modernisation effort was still awaited.

Other requirements around the globe

Other regulators around the world have introduced similar measures to promote transparency. Two have recently gone further by mandating the anonymisation and publication of CSRs: the EMA and Health Canada.

EMA Policy 0070

In 2016, the EMA implemented Policy 0070, which requires publication of the regulatory documents used in a successful marketing authorisation application. These documents include the CSR and selected appendices, as well as clinical overviews and clinical summaries. The CSR provides extensive details on the clinical trial, including the study objective, the investigational plan, and study design, the evaluation and analysis performed, as well as specific information about the trial subjects. This last item – detailed information about the participants' experience in the trial – creates potential privacy concerns and necessitates effective anonymisation of the CSR prior to publication.

The EMA's external guidance on the implementation of Policy 0070 encourages applicants to use quantitative methods to measure the risk of re-identification, recommending a risk threshold of 0.09.[11] The re-identification risk will depend upon indirectly identifying information in the documents, such as demographics and medical history information. The CSRs must be anonymised to mitigate the re-identification risk and reduce it to an acceptably low level (below 0.09). Applicants are asked by the EMA to justify alterations to the data and their choice of anonymisation techniques.

This updated guidance from the EMA in 2018 further reinforced a preference for

quantitative measures of risk over qualitative assessments and from redaction to anonymisation. Redaction, in this case, means complete concealment of patient data with an opaque box, such that all the inherent usefulness of the information is effectively removed. Anonymisation refers to the replacement of the original text with re-synthesised values selected to bring the re-identification of a given trial participant below the threshold. A visual comparison is presented in Figure 1.

While the EMA paused its Policy 0070 efforts due to a temporary closure following Brexit, it plans to resume efforts from its new headquarters in Amsterdam. During its June 2020 board meeting, the EMA affirmed its plan to resume publication for COVID-19 trial information, citing assurance needed by the public over the quality of evidence behind regulatory decisions.[12]

EMA's Policy 0070 contemplates a second phase in which the disclosure of participant-level data will be mandated, though timelines have not yet been announced.

Health Canada Public Release of Clinical Information

In 2019, Health Canada introduced its Public Release of Clinical Information (PRCI) requirements. PRCI mirrors EMA requirements, with the disclosure of CSRs now required for market authorisation. However, unlike EMA Policy 0070, PRCI also applies to historical submissions upon request.

Health Canada's guidance asks manufacturers to anonymise the clinical information using a risk-based statistical anonymisation process that is closely aligned to EMA guidance. Like EMA, Health Canada recommends a threshold of 0.09.[13]

While Health Canada has been explicit in their preference for a quantitative approach to re-identification risk measurement, during the early period of adoption, they have accepted submissions where a non-analytical or qualitative approach was taken, as well as those in which redaction was applied. However, there is a strong indication that Health Canada is encouraging a movement away from reliance on these methods. As an example, certain submissions that were heavily redacted were published with a notice from Health Canada:

NOTICE:
This clinical information package includes extensive redactions to the patient information. These redactions do not conform to Health Canada guidance, which encourages manufacturers to retain the analytical value of information by using other transformation methods (e.g., generalisation or randomisation), and to apply these methods to specific information that risks re-identifying an individual rather than to redact broad sections of information.

Health Canada encourages manufacturers to anonymise personal information according to the principles outlined in Guidance Document: Public Release of Clinical Information. Health Canada will continue to explore ways to help ensure all publications include anonymized personal clinical information.

If you require access to the redacted information, you may submit inquiries to the Information Science and Openness Division (hc.clinicaldata-donneescliniques.sc@canada.ca).

Health Canada's PRCI is very similar to the EMA's Policy 0070, with a few notable differences:

1. Health Canada's PRCI applies to device trials, in addition to drug and biologic trials.
2. In addition to proactive submissions for market authorisation, Health Canada's PRCI includes the publication of historical studies in response to access-to-information requests from the public.
3. Health Canada has not announced plans to enforce disclosure of individual participant-level data, whereas EMA has indicated its intent for a Phase 2 encompassing participant-level data.

Other minor variances exist, such as abnormal laboratory value listings being "in scope" for Health Canada PRCI but not for EMA Policy 0070. However, Health Canada has indicated its intent on accepting previously approved EMA packages to avoid effort duplication.

Statistical anonymisation

With the shifts to publishing the complete CSRs, several manufacturers have also shifted from using basic redaction methods to statistical anonymisation.[14] With the regulatory timelines, particularly for historical study requests in Canada, the need for scalable, efficient anonymisation capabilities has become more apparent.

By developing statistical anonymisation capabilities, organisations can safely share and reuse data for a variety of beneficial purposes, transforming data in a manner commensurate with the risk to protect privacy and achieve transparency.

With investment in anonymisation capabilities, new opportunities to reuse data for other purposes, including internal innovation, have become more evident. Organisations can use anonymisation to gain secondary benefits from trial data, such as gaining insights into the drug discovery process.

A statistical anonymisation approach and capability can be applied to a wide range of contexts and data types. The process of anonymisation evaluates the context of disclosure to understand potential threats and uses this to evaluate the identifiability of the data. The contextual evaluation should consider all means reasonably likely to be used to re-identify individual people. The data are then evaluated using generally accepted statistical techniques that measure whether individuals can be identified in the data. Finally, the data are transformed to the degree necessary to be rendered non-identifying.

When anonymising documents for transparency, a key first step in the process is detecting all the identifying variables associated with each individual data subject (i.e., trial participant) in order to measure identifiability. While directly identifying information (e.g., subject IDs, or direct identifiers like names, addresses or email addresses) is removed or pseudonymised, the indirectly identifying information is typically preserved as much as possible without compromising privacy through re-identification. Indirectly identifying information includes demographics, medical histories, event dates, diagnoses, treatments, and other information that may be used in combination to identify an individual person. Many transformation techniques, such as generalisation, randomisation, date shifting, and targeted suppression, can be used in a flexible manner that preserves as much utility (and transparency) as possible.

A recent article in the journal *Trials* (February 2020) included results from a commissioned re-identification attack on a clinical study that had been anonymised using statistical methods according to EMA Policy 0070 guidance.[14] The study results suggest that anonymisation provides adequate privacy protection for trial participants, with very low confidence match scores achieved with over 24 hours of effort per attempted match during the commissioned attack.

The same statistical anonymisation methodology can be applied to discretionary data sharing in support of transparency, transforming individual participant data to the degree necessary to safely support secondary research. Similarly, internal reuse of the data can be achieved through functional anonymisation. Contextual factors, such as security, privacy, and contractual controls, should be considered in the anonymisation approach, with less controlled releases needing more data transformation. By developing statistical anonymisation capabilities, organisations can safely share and reuse data for a variety of beneficial purposes, transforming data in a manner commensurate with the risk to protect privacy and achieve transparency.

Acknowledgements

The authors would like to thank Christine Lee Haggard, Associate Medical Writing Director at IQVIA, Isy Goodwin, Senior Director Medical Writing at IQVIA, Luk Arbuckle, Chief Methodologist at Privacy Analytics, and Niamh McGuinness, Senior Clinical Trial Transparency Analyst at Privacy Analytics for helpful comments and insight.

Disclaimers

The opinions expressed in this article are the authors' own and not necessarily shared by their employer or EMWA.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Cochrane, CRIT, Transparency International Pharmaceuticals and Healthcare Programme and TranspariMED. Clinical Trial Transparency: A Guide for Policy Makers. Transparency International UK. 2017. AllTrials Transparency Index. 2015 [cited 2020 Aug 29]. Available from: http://policyaudit.alltrials.net/.
2. Health Canada. Public Release of Clinical Information (PRCI). 29 Mar 2019 [cited 2020 Sep 19]. Available from: https://www.canada.ca/en/health-canada/services/drug-health-product-review-approval/profile-public-release-clinical-information-guidance.html.
3. EMA. Policy 0070. 2018 [cited 2020 Sep 19]. Available from: https://www.ema.europa.eu/en/documents/report/clinical-data-publication-policy-0070-report-oct-2016-oct-2017_en.pdf.
4. Yale University. The YODA Project. 2020 [cited 2020 Sep 19]. Available from: https://yoda.yale.edu/data-holders#:~:text=%20The%20YODA%20Project%20approach%20to%20data%20sharing,ensure%20that%20all%20required% 20information%20is . . . %20More%20.
5. Vivli. A global clinical research data sharing platform. 2020 [cited 2020 Sep 19]. Available from: https://vivli.org/.
6. U. S. Food and Drug Administration. Food and Drug Administration Amendments Act (FDAAA) of 2007, Pub. L. No. 110–85, 121 Stat. 823. Section 801. Expanded clinical trial registry data bank. 2007 [cited 2020 Aug 7]. Available from: https://www.govinfo.gov/content/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf.
7. ClinicalTrials.gov. Checklist for Evaluating Whether a Clinical Trial or Study is an Applicable Clinical Trial (ACT) Under 42 CFR 11.22(b) for Clinical Trials Initiated on or After January 18, 2017. [cited 2020 Aug 7]. Available from: https://prsinfo.clinicaltrials.gov/ACT_Checklist.pdf.
8. U. S. National Library of Medicine. ClinicalTrials.gov. History, Policies, and Laws [cited 2020 Aug 7]. Available from: https://clinicaltrials.gov/ct2/about-site/history#ReleasesResultsDatabase.
9. Federal Register. Department of Health and Human Services. Clinical Trials Registration and Results Information Submission; Final Rule, 42 CFR Part 11. 2016 [cited 2020 Aug 07]. Available from: https://www.federalregister.gov/documents/2016/09/21/2016–22129/clinical-trialsregistration-and-results-informationsubmission.
10. U. S. National Library of Medicine. ClinicalTrials.gov. ClinicalTrials.gov Modernization [cited 2020 Aug 11]. Available from: https://clinicaltrials.gov/ct2/about-site/modernization#events.
11. EMA. European Medicines Agency policy on publication of clinical data for medicinal products for human use. 2019 [cited 2020 Sep 19]. Available from: https://www.ema.europa.eu/en/documents/other/european-medicines-agency-policy-publication-clinical-data-medicinal-products-human-use_en.pdf.
12. EMA Minutes of the 108th meeting of the Management Board. EMA/MB/38294/2020. 2020 [cited 2020 Sep 24]. Available from: https://www.ema.europa.eu/en/documents/minutes/minutes-108th-meeting-management-board_en.pdf.
13. Health Canada. Public Release of Clinical Information (PCRI). 2019 [cited 2020 Sep 19]. Available from: https://www.canada.ca/en/health-canada/services/drug-health-product-review-approval/profile-public-release- clinical-information-guidance/ document.html.
14. PHUSE. Data Transparency Workstream: A Global View of the Clinical Transparency Landscape – Best Practices Guide. 2020 [cited 2020 Sep 19]. Available from: https://www.phusewiki. org/docs/WorkingGroups/Deliverables/ PHUSE%20DT%20Clinical%20Trial%20Landscape-%20Best%20Practices%20Guide.pdf.
15. Branson J, Good N, Chen J, Monge W, Probst C, El Emam K. Evaluating the re-identification risk of a clinical study report anonymized under EMA Policy 0070 and Health Canada Regulations. Trials. 2020;21:200.

Author information

Sarah Lyons is the General Manager of Privacy Analytics, an IQVIA company, where she helps companies deliver insight-rich health data while protecting patient privacy. As a recognised pioneer in anonymisation methodologies and wholly owned subsidiary of IQVIA, Privacy Analytics provides anonymisation software and services. Before joining Privacy Analytics as Chief Operating Officer in 2017, Sarah was a Partner at MNP LLP and holds professional designations in privacy engineering, risk management, information system control, and audit.

Vivien Fagan has over 24 years of experience in an international Contract Research Organisation. In her current position as Director, Medical Writing, Vivien manages IQVIA's medical writing group based in Livingston, Scotland, whilst heading up the global clinical trial disclosure group. Vivien is one of the co-authors of CORE Reference (www.core-reference.org), is an EMWA workshop leader and has been a DIA panellist.