

# A primer on anonymisation

**Jackie Raskind**

KPS Life, LLC, Malvern, PA, USA

## Correspondence to:

Jackie Raskind, PharmD  
KPS Life, LLC  
57 HaGallil Street  
Ra'anana, Israel, 4325143  
jackieraskind@gmail.com

## Abstract

Canadian and European regulators finalised guidelines that allow for sharing of clinical trial data. To maintain the utility of clinical information, risk-based anonymisation techniques are recommended. It behoves applicants to ensure organisational readiness to deliver anonymised dossiers. Key steps include understanding the regulatory requirements and anonymisation techniques, assessing internal competencies and/or commercially available software, and establishing infrastructure to deliver anonymised dossiers in a timely manner.

Anonymisation is defined as the process of turning data into a form that does not identify individuals and where identification is not likely to take place.<sup>1,2</sup> Regulations, policies, and guidance have been promulgated in the EU to allow clinical-trial data (EU Regulation EU No. 536/2014) and dossier-level data (EMA Policy 0070) to be shared with the scientific community, participants, and the public.<sup>3</sup> Recently, Health Canada finalised parallel guidance on the Public Release of Clinical Information for clinical trial and dossier-level data and expanded the scope of sharing to include not only drugs but also medical devices.<sup>4</sup> In sharing clinical trial and dossier-level data, applicants have the challenge of protecting company confidential information (CCI) and safeguarding the privacy of personal private data (PPD) whilst retaining the utility of the data after it has been anonymised.

Regulators do not prescribe a specific anonymisation method, although several data transformation techniques can be used to

anonymise the direct identifiers (e.g. names, initials, signatures, job titles/positions, addresses, fax numbers, and email addresses),<sup>5,6</sup> indirect identifiers (e.g., sex, age, dates, and socio-economic information),<sup>5,6</sup> and CCI in clinical study documents and datasets. Anonymisation methods include redaction, pseudonymisation, randomisation, offsetting, and generalisation (Box 1).<sup>3,4</sup>

## Anonymisation methods

In its guidance, the EMA recognises that, in an initial phase, applicants will anonymise dossiers using the redaction method. Given that this method decreases data utility,<sup>6</sup> the EMA recommends that other anonymisation techniques be used as soon as possible, whilst ensuring that data anonymisation is achieved.<sup>3</sup> Health Canada's guidance recommends that anonymisation favour methods that retain analytical value (e.g. generalisation, randomisation, and offsetting) instead of redaction.<sup>4</sup>

Preparing an anonymised dossier is a labour-

intensive, iterative process for the applicant that is constricted by strict regulatory timelines. The main steps include:

- Applicant submission of a redaction proposal document package
- Consultation between the applicant and the regulatory agencies
- Submission of the final redacted document package
- Publication

## Anonymisation reports

Both the EMA and Health Canada require that an anonymisation report be submitted with the anonymised submission in the proposal package. This anonymisation report contains the methods and justification for the processes used. The purpose of this report is to:

- demonstrate that changes included within the anonymised documents are adequate to protect study participants' privacy;
- provide the rationale for those changes; and
- demonstrate that after anonymisation, the

Preparing an anonymised dossier is a labour-intensive, iterative process for the applicant that is constricted by strict regulatory timelines.

### Box 1. Redaction methods

**Redaction:** This involves removing or masking values. Redaction may best be applied to direct identifiers. When a directly identifying variable is critical to understanding the clinical information, other anonymisation methods should be selected. Redaction may be useful for documents such as protocols and statistical analysis plans.

**Pseudonymisation:** Personal information (e.g., subject identification number) is re-coded to disassociate the variable from the participant.

**Randomisation:** This involves making small changes to variables to reduce the possibility that the data are used to identify a participant.

**Offsetting:** This involves replacing numerical data by adding or subtracting a fixed quantity.

**Generalisation:** This technique uses re-categorisation within a range to enlarge the number of "like" individuals. Examples of generalisation techniques include:

- Aggregation: Replacing a value by a range, for example, replacing a trial participant's age by an age range (e.g. 56 replaced by 50–60).
- K-anonymity: Trial participant data are grouped with at least  $k$  other trial participants in that range, preventing the participant from being singled out and identified.<sup>7</sup>



risk of re-identification when released in the public domain is at an acceptable level and that the impact on data utility has been considered.

Regulators review the applicant’s stated rules and the anonymised clinical documents to assess whether the applicant has executed the planned

data transformations systematically and consistently.

The risk of re-identification takes into account the number of direct and indirect identifiers (“quasi-identifiers”), size and nature of the disease studied (e.g. rare or common disease, paediatric population), number of

participants in the study, and the number and distribution of study centres across countries.<sup>6</sup> The acceptable maximum quantitative risk level for re-identification per the EMA and Health Canada is 9%,<sup>3,4</sup> or qualitatively at a risk level of high, medium, or low based on the characteristics of the source data (e.g., disease prevalence, trial sample size, number of sites). The quantitative risk threshold of 9% is equivalent to a group size of 11. This means that a single participant, when grouped by similar variables, cannot be re-identified from 10 other participants ( $k-1$ ) within the aggregated group. Health Canada also recommends that the applicants select a reference population for indirect variables to help estimate the risk of re-identification.

### Software solutions for anonymisation

To complete these tasks efficiently over the breadth of a dossier, applicants require software solutions.

#### Redaction-only commercial software

Multiple software platforms enable manual redaction of PDF documents by users (Table 1). They often offer free trials. Below are desktop software options that are used by pharmaceutical companies for redaction or that are capable of handling the breadth of redaction required in dossiers. All of the software packages listed require licences for use.

#### Acrobat® Pro® DC (Adobe)

Adobe Acrobat® Pro® DC enables a user to redact PDF documents by using a redaction toolbar. Redaction is conducted in two steps. First, the text is selected and marked for redaction. The text will then appear within a red border. Second, the redactions are applied, which result in permanent removal of the redacted content including its metadata. Text, images, or multiple pages may be marked for redaction. A PDF marked for redaction may be downloaded and reviewed by another reviewer before a redaction is applied. The tool also includes a search feature where all text meeting the search requirements may be marked for redaction at once.

#### Objective Redact (Objective Corp.)

Objective Redact enables users to manually

Table 1: Commercial redaction software

Software	For further information
Adobe Acrobat Pro DC	<a href="https://acrobat.adobe.com/us/en/">https://acrobat.adobe.com/us/en/</a>
Nitro Pro 10	<a href="https://www.gonitro.com/nps/pro/pdf-software">https://www.gonitro.com/nps/pro/pdf-software</a>
Objective Redact	<a href="https://www.objectiveredact.com/">https://www.objectiveredact.com/</a>
PDFelement	<a href="https://pdf.wondershare.com/pdfelement.html">https://pdf.wondershare.com/pdfelement.html</a>
PDFzorro.com	<a href="https://PDFzorro.com">https://PDFzorro.com</a>
PDFescape	<a href="https://pdfescape.com">https://pdfescape.com</a>
PDFfiller	<a href="https://.PDFfiller.com">https://.PDFfiller.com</a>
PDF Studio Pro	<a href="https://www.qoppa.com/pdfstudioviewer/">https://www.qoppa.com/pdfstudioviewer/</a>

redact PDF, Word, or Excel documents. Users can add annotations in comments indicating why the information was redacted, information that can be included in the anonymisation report. The program can search for phrases or structured data (e.g. birthdates or names). The software also removes any metadata or hidden code. Users can create an audit copy, which displays the redactions as translucent markings, allowing the underlying information to remain visible for review. A working copy can also be created that can be saved or emailed to reviewers. The working copy can then be opened within Objective Redact to complete the redaction process by accepting the redactions or can be further redacted and reviewed.

**PleaseReview version 6.1 (Ideagen)**

PleaseReview Version 6.1 enables multiple users to simultaneously and collaboratively redact PDF documents using pre-configured redaction categories (PPD-Policy 70 or CCI-Policy-70). Users have the option of further configuring appearance (font, font size, font colour, alignment), overlay text, and comment categories for PPD or CCI at a system-, workgroup-, or review-level. Redaction is conducted by selecting options from a dropdown menu. Users can apply redactions as a rectangle, highlighted text, or complete pages. Additionally, PDF document annotations (including redactions) can be imported into PleaseReview. The review owner may accept or reject redactions. Users can add justifications for redaction as comments when redacted. The information collected in the reconciliation report can serve as the basis for the justifications for de-identification of data reported to the EMA or Health Canada.

Automated software is recommended for handling anonymisation of dossiers, given their scope and complexity and the need to maintain data utility.

**Commercial anonymisation software**

Automated software is recommended for handling anonymisation of dossiers, given their scope and complexity and the need to maintain data utility. Applicants that need to anonymise a substantial number of clinical dossiers may opt to create and leverage in-house biostatistics and data management, who can customise SAS-based macros. This is a resource-intensive exercise and should be supported by a transparency team and governed by standard operating procedures. The process requires ongoing refinement to align with changing regulations and practices.<sup>8</sup> Therefore, applicants who have either restricted in-house resources or who anticipate having only a few dossiers to be anonymised may opt to outsource the work. Alternatively, applicants may wish to develop in-house capabilities by partially outsourcing anonymisation services, for example using vendor-provided software platforms in-house with vendor support for consulting services. Several vendors provide anonymisation software and services tailored to comply with EMA’s Policy 0070 and Health Canada’s regulatory requirements. In general, software solutions are aligned with established anonymisation rules and standards (e.g., PhUSE<sup>1,2</sup>) and are updated to comply with advances in data standards, artificial intelligence, and the evolving regulations. Given the wider scope of redacted submissions supported by these vendors and experience in redacting datasets for secondary use by researchers before these regulations were implemented, applicants can also leverage vendor expertise in relying on additional services provided. This may include:

- assisting with establishing standard operating procedures for anonymisation;

- providing software platforms and technical support and training;
- providing strategic recommendations for method selection of anonymisation and risk mitigation contingent on the population and dossier provided;
- reviewing and conducting quality control of the software-generated anonymised clinical reports and data;
- completing the anonymisation report and justification table (CCI); and
- assisting applicants in justifying the anonymisation method.

Vendor support may also extend beyond anonymisation and include protocol and results posting to registries, record maintenance, workflow planning, and writing of lay summaries.

In general, when approached by a client to anonymise a dossier, vendors first assign a project manager. Based on an analysis of the contents of the dossier, including the direct and indirect identifiers and CCI, they will recommend anonymisation rules for PPD and CCI based on the acceptable risk threshold for re-identification. Once they obtain client agreement on the risk threshold and methods used, they will generate the anonymised proposal package ensuring consistent replication of rules across data sets and documents. The package will then be reviewed by the client, and, once it has been agreed on, the vendor will finalise document proposal package. Vendors will also provide client support in responding to questions from health authorities and will prepare the final document package.

Key anonymisation software vendors are described below and listed in Table 2.

**ARARA (Real Life Sciences)**

ARARA is a software platform that enables automated anonymisation, quantitative risk, and data utility assessments of clinical documents

**Table 2: Anonymisation software vendors**

Platform	Company	For further information
ARARA	Real Life Sciences	<a href="https://rlsciences.com">https://rlsciences.com</a>
Blur	d-Wise	<a href="https://cdn2.hubspot.net/hubfs/188561/Blur%20datasheet-2018.pdf">https://cdn2.hubspot.net/hubfs/188561/Blur%20datasheet-2018.pdf</a>
ClinGenuity	Certara	<a href="https://www.synchrogenix.com/technology/transparency-disclosure/">https://www.synchrogenix.com/technology/transparency-disclosure/</a>
Eclipse	Privacy Analytics	<a href="https://privacy-analytics.com/software/privacy-analytics-eclipse/">https://privacy-analytics.com/software/privacy-analytics-eclipse/</a>
Redact360	Kinapse	<a href="http://www.kinapse.com/redact360">http://www.kinapse.com/redact360</a>
Terminator	XOGENE	<a href="https://www.xogene.com/">https://www.xogene.com/</a>

and datasets. It is available as a desktop or cloud-hosted application. Sponsors can either outsource the anonymisation service or use the ARARA platform internally.

ARARA provides user options for de-identifying direct and indirect identifiers and CCI with user control over the automated modelling and anonymisation process. ARARA is pre-packaged with a range of anonymisation rules and templates that are preconfigured to meet the EMA and Health Canada’s 9% re-identification risk thresholds. The rules can be customised for different risk thresholds as needed for different populations. Ten different anonymisation models are also provided. Templates can be customised by data variable, model, or model thresholds and can be re-run in the user interface as needed, enabling users to rapidly test outputs and iterate through the process. The platform can process multiple datasets and clinical documents at once using the same anonymisation techniques. The platform provides reporting features including: risk and data utility visualisation dashboards, automated anonymisation reporting, traceability, and audit tracking tools.

**Blur (d-Wise)**

Blur is a software platform that supports automated data and clinical document anonymisation. Applicants may either outsource the anonymisation service or choose to use the Blur application internally. The Blur application provides menu-driven options for de-identification, risk reduction, and creation of auditable workflows and controls. The application enables users to apply different anonymisation techniques by data variables and allows viewers to compare the original source data and the data once it has been anonymised. The rule sets created within the data can then be applied consistently to documents using a template. D-Wise partners with ClaritiDox to deliver additional anonymisation services including advisory services,

assistance with process management, review of anonymised documents and datasets, writing of anonymisation reports, managing registry postings, and writing of lay summaries.

**ClinGenuity (Certara)**

Certara provides a complete redacted and anonymised Policy 0070 submission package applying its redaction management software, ClinGenuity. Certara has also developed advanced anonymisation solutions using quantitative risk assessment methodology using internal expertise and aligned with globally accepted regulations and industry standards. Certara provides an anonymised package and consulting services as a service model.

**Eclipse (Privacy Analytics)**


Privacy Analytics provides a complete anonymised Policy 0070/Health Canada submission package applying its risk-based de-identification software, Eclipse. The quantitative risk-based methodology is aligned with globally accepted regulations, standards, and guidelines for anonymisation. Privacy Analytics provides the anonymised package and consulting services as a service model.

**Redact360 (Kinapse)**

Redact360 service is Kinapse’s technology-enabled service for redaction and anonymisation of clinical data and documents. Kinapse provides a service comprising advisory and regulatory support for the method of anonymisation, programme management of the anonymisation process including providing the anonymised proposal and final packages, completion of the anonymisation report and justification tables, writing of clinical trial summaries, registry postings, and lay summaries. Although only Kinapse can use the technology-enabled platform to anonymise the clinical documents and datasets, the applicant may select amongst the suite of advisory and support services provided. Kinapse has established key performance indicators for their level services and incorporates them into their service-level agreements.

**Terminator (XOGENE)**

XOGENE provides a complete anonymised Policy 0070/Health Canada submission package, including the anonymisation report, justification table, and final anonymised datasets and documents. To perform the anonymisation, XOGENE applies the technology of its automated anonymisation and redaction platform (XOGENE Terminator) to PPD in datasets and clinical documents. To complete the package,



The regulatory landscape and technological advances to support disclosure and transparency requirements are evolving rapidly. Effective leveraging of internal and external resources is critical for compliance with regulatory requirements and timelines.



XOGENE also reviews clinical documents to identify CCI and reviews against regulatory and company policies. XOGENE can also support expedited (24- to 48-h) redaction for EMA Policy 0070 requests. Additional disclosure services provided by XOGENE include protocol and results postings to all registries, record maintenance on registry sites. XOGENE can also manage the writing of lay summaries, translation of lay summaries, site distribution, tracking, and oversight.

## Applicant readiness for anonymisation

The regulatory landscape and technological advances to support disclosure and transparency requirements are evolving rapidly. Effective leveraging of internal and external resources is critical for compliance with regulatory requirements and timelines.

For initial marketing authorisation applications and line extension applications submitted to the EMA under the centralised procedure, the redaction proposal document package must be submitted between day 181 and day 220 of the procedure ( $\leq 30$  days pre-opinion and  $\leq 10$  days post-opinion). A total of 84 calendar days are allocated from submission of the redaction proposal package to final publication. During the consultation process (total of 47 days) the EMA

will review the anonymisation report, justification table, and redaction proposal. The EMA can seek clarifications from the applicant, after which the applicant updates the justification table. As part of the EMA's review, a redaction conclusion notification is sent to the applicant and the applicant is expected to submit a redaction consultation agreement within 7 calendar days. Applicants then have 27 calendar days to prepare the final redacted proposal package. The final redacted version is published within 60 days of the commission decision on the approvability of the marketing authorisation application.<sup>3</sup>

For marketing authorisation applications submitted to Health Canada, applicants may request a process initiation meeting (PIM) between 120 calendar days before the final regulatory decision and 20 days after the final regulatory decision. Redaction proposal document packages must be submitted to Health Canada within 60 days after a positive decision. Health Canada reviews the proposal package (30 days) and provides rejected redactions to the applicant for revision. The process for revising and finalising any rejected redactions is allocated 25 days, comprising 15 days for applicant revision, 5 days for Health Canada reassessment, and 5 days for applicant submission of finalised documents. Health Canada will then publish the

data within 5 days, meeting the 120 day target from the initiation of the process. In the case of a negative opinion, the process may start 31 days after a negative decision. If a letter of intent for reconsideration is submitted, the process will start after the reconsideration process is complete (70 to 140 days), and the 120 day process for submitting the redaction proposal package will commence.<sup>4</sup>

These timelines may require applicants to prepare for the redaction in advance of a health authority decision. Therefore, it is incumbent on applicants to establish a company-wide readiness strategy towards engaging vendor services and global transparency. These steps may include:

- understanding and educating key stakeholders of the global transparency regulatory landscape and anonymisation methods;
- identifying members of a transparency committee;
- defining the scope of the transparency committee in a charter;
- establishing standards for the acceptable risk thresholds of re-identification for anonymised proposal packages submitted in the public domain;
- conducting a feasibility analysis and selecting an anonymisation framework (e.g. entirely managed internally, partial outsourcing, complete outsourcing);

- vetting and selecting an anonymisation vendor, software, and solution;
- establishing standard operating process and procedures for managing anonymisation requests including the process for managing requests, expectations of team members reviewing anonymised clinical documents and data sets, standard timelines for internal review, the process for adjudicating comments, and the point person to interact with the vendor providing services;
- completing vendor procurement steps to ensure immediate preparedness for dossier submission or upon a regulatory request;
- investigating information technology capabilities to support software platforms and compliant document and data transfer procedures;
- developing document and data checklists;
- ensuring accessibility of documents and data sets to ensure rapid transfer of clinical dossier components and data sets; and
- engaging in a continuous quality improvement process with the submission of each anonymised package.

### Acknowledgements

The author would like to thank Laura Sheppard and Alexander Pratt at Endo Pharmaceuticals, Inc. for a review of a draft of this manuscript.

### Disclaimers

The opinions expressed in this article are the author's own and not necessarily shared by her employer or EMWA.

### Conflicts of interest

The author declares no conflicts of interest.

### References

1. PhUSE. Protection of Personal Data in Clinical Documents – A Model Approach. Available from: <http://www.transceleratebiopharmainc.com/wp-content/uploads/2017/02/Protection-of-Personal-Data-in-Clinical-Documents.pdf>.
2. PhUSE. De-identification and anonymization of individual patient data in clinical studies. A Model Approach. Available from: <http://www.transceleratebiopharmainc.com/wp-content/uploads/2015/04/TransCelerate-De-identification-and-Anonymization-of-Individual-Patient-Data-in-Clinical-Studies-V2.0.pdf>.
3. EMA. External Guidance on the Implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use. 2018 [cited 30 Oct 2019]. Available from: [https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data\\_en-3.pdf](https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en-3.pdf).
4. Health Canada. Guidance Document on Public Release of Clinical Information. 2019 [cited 30 Oct 2019]. Available from <https://www.canada.ca/en/health-canada/services/drug-health-product-review-approval/profile-public-release-clinical-information-guidance/document.html>.
5. Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. *BMJ*. 2010;340:c181.
6. EMA. Data anonymization – A key enabler for clinical data sharing. Workshop report. 2017 [cited 30 Oct 2019]. Available from: [https://www.ema.europa.eu/en/documents/report/report-data-anonymisation-key-enabler-clinical-data-sharing\\_en.pdf](https://www.ema.europa.eu/en/documents/report/report-data-anonymisation-key-enabler-clinical-data-sharing_en.pdf).
7. El Emam K, Dankar FK. Protecting privacy using k-anonymity. *J Am Med Inform Assoc*. 2008; -15(5):627–7.
8. Keerie C, Tuck C, Milne G, Eldridge S, Wright N, Lewis SC. Data sharing in clinical trials - practical guidance on anonymising trial datasets. *Trials*. 2018;19(1):25.

### Author information

**Jackie Raskind**, PharmD, is a principal medical writer for KPS Life, LLC, working remotely from Israel. She transitioned into medical writing 6 years ago after a 17-year career as a clinical pharmacist in ambulatory, inpatient, and pharmacy benefit managed care settings.



**Save the date: EMWA Conference in the Czech Republic**

# PRAGUE

**May 6 to 9, 2020** <https://www.emwa.org/conferences/future-conferences/>