21 May 2024
EMA/482639/2023
Transparency Department
Stakeholders and Communication Division

Anonymisation Report Form Instructions

Introduction

Anonymisation is the process of rendering data into a form that does not identify individuals and that makes identification unlikely. Applicants and marketing authorisation holders (MAHs) must ensure that the documents submitted for publication in the context of the Health Canada Public Release of Clinical Information (PRCI) and EMA Policy 0070 (Clinical Data Publication) guidance documents are properly anonymised and pose a low or very low risk of re-identification.

This document contains instructions and a set of definitions to guide applicants on how to complete the anonymisation report form for the clinical document package.

The anonymisation report form contains the following sections:
- Application Information
1. Anonymisation Methodology
2. Identification of Data Variables
    - 2.1. Direct Identifiers
    - 2.2. Indirect Identifiers
3. Risk Assessment
4. Data Utility
5. Deviations
6. Attestation

Application Information

- Date Prepared: Use the drop-down calendar to select the date on which the final version of the anonymisation report was prepared.
- Product Name: Indicate the product name (e.g., invented or brand name) in this text box. If the same product has different names across multiple jurisdictions, please ensure that the product name listed is the same as the one used in the jurisdiction of the anonymisation report. If this is a joint EMA/HC document package and the names are different in the EU and Canada, please include both names separated by a forward slash (/).
- Active Ingredient/INN(s): Indicate the active substance(s)/international non-proprietary name(s).
- EMA Procedure Number: For EMA or joint EMA/Health Canada publications, indicate the specific EMA regulatory procedure number to which the document package belongs. For Health Canada publications only, please indicate N/A.
- Applicant/MAH: Indicate the name of the applicant/manufacturer/marketing authorisation holder.
- Health Canada Control Number: For Health Canada or joint EMA/Health Canada publications, indicate the specific Health Canada control number for the submission. For EMA publications only, please indicate N/A.

A) Are there any indirect identifiers present within the clinical information package?

Select "Yes" if the clinical information package contains indirect identifiers and proceed with completing the remaining sections of the anonymisation form.

Select "No" if the clinical information package contains no indirect identifiers (i.e., it only contains direct identifiers) Please note, if "No" is selected, only the additional "Section 6: Attestation" at the end of the form will need to be completed.

Depending on the option selected ("Yes" or "No"), the attestation text in Section 6 will be automatically tailored to reflect the appropriate context.

1. Anonymisation Methodology

B) What anonymisation method was used to measure risk?

Select the anonymisation method that was utilised for the clinical information package. Please note that a quantitative methodology is the preferred approach over a qualitative methodology.

N.B. The use of a qualitative methodology must be justified in "Section 5: Deviations".

Quantitative – A methodology that uses empirical measurement (e.g., k-anonymity) to present the measurement of risk as a numerical value.

Qualitative – A methodology that is subjective in nature and uses a qualitative scale (e.g., high, medium, or low risk) to present the measurement of risk.

Both – If both qualitative and quantitative methods were used in different documents, list the name of each study/document type along with the anonymisation approach used for each one. For example: 'Study A – quantitative, Study B – qualitative'. Please note that justification as to why different methodologies were used for the studies in the clinical information package should be provided in the answer to Question M).

C) Please select the overlay text employed for redaction

Select "PI" (Personal Information) if the overlay text follows the Health Canada's format. The redaction box would be a blue box with "PI" written in white or black overlay text.

Select "PPD" (Protected Personal Data) if the overlay text follows the EMA format. The redaction box would be a blue box with "PPD" written in black overlay text.

2. Identification of Data Variables

Only identifiers present in this clinical information package should be listed in this section.

N.B. Do not include a general overview of any possible identifiers or a predefined list of identifiers developed for risk assessment.

Use the plus symbol (+) on the left to add a new row.

N.B. When adding a row, the new row will always appear as the second row if you have already entered identifiers in the table. Make sure that all identifiers are listed in the desired order prior to completing the table.

Use the minus symbol (-) on the left to delete a row.

N.B. if different anonymisation techniques have been used for a single identifier throughout the clinical documents, the identifier should be listed separately in the table for each approach. Please use the Comments box to include additional information for clarification, as required (e.g., relevant study or document, and what the transformed/recoded/generalised term has been changed to).

2.1 Direct Identifiers

N.B. For all EMA packages and any EMA/HC joint packages the following direct identifiers must be preserved:
- the name(s) of the clinical report(s) signatory(ies)
- the name(s) of the principal investigator(s).

This approach is in accordance with EMA's Policy 0070 guidance.

D) In the table below, please list the direct identifiers present in the clinical information package. Only include identifiers that are present in the document package.

Category column
Use the dropdown box to select the category of the direct identifier present in the clinical information package.

Participant/Personnel column
Use the dropdown box to select the relevant individual category (Participant/ Personnel) to whom the identifier pertains to in any of the clinical documents.

Anonymisation column
Use the dropdown box to select the anonymisation technique(s) employed for the respective identifier.

If multiple anonymisation techniques have been applied to the same identifier, use the plus symbol (+) in the top left corner to list the identifier separately within the Category section for each technique.

Use the Comments box to include any additional information (see instructions for Comments column below).

Please ensure that no empty rows are left in the final report.

Comments column
If applicable, use the Comments box to include any additional information relating to the listed identifier that would provide more context regarding the anonymisation techniques. For example, indicate to which study/studies the technique in question applies, or any other information that would facilitate the reader's understanding.

Example:

Category	Participant/ Personnel	Anonymisation	Comments
All Other IDs Unique to Participant	Participant	Yes (Redacted)	Manufacturer control number
All Other IDs Unique to Participant	Participant	Yes (Redacted)	Case number

2.2 Indirect Identifiers

E) In the table below, please list the indirect identifiers present in the clinical information package. Only include identifiers that are present in the document package.

Adverse Events Terms

Anonymisation column
Use the dropdown box to select the anonymisation technique applied to Adverse Event terms. If there are no adverse events present anywhere in the clinical information package, please indicate "N/A" in the comment column for this category.

It is expected that all Adverse Event terms are **retained** in both summary-level (i.e., tables, descriptive summaries) and participant-level data (i.e., narratives). In exceptional cases where the (Serious)/Adverse Event term poses a serious risk of re-identification or is highly unique, the term should be **Generalised** to the HLT (high level term), HLGT (high level group term) or SOC (system organ class) as classified by MedDRA.

Redacted: text is obscured masking the safety event from public access. Adverse Events should not be redacted. If redactions are present, a justification is required for HC/EMA review in the deviation section.

Comments column
If applicable, use the Comments box to include any additional information relating to the listed identifier that would provide more context regarding the anonymisation techniques. For example, indicate which study/studies the technique in question applies, or any other information that would facilitate the reader's understanding.

Other indirect identifier categories

Category column
Use the dropdown box to select the indirect identifier type present in the clinical information package.

<u>Participant/Personnel column</u>
Use the dropdown box to select the relevant type of individual (Participant/Personnel) to whom the indirect identifier pertains in the clinical information package.

<u>Anonymisation column</u>
Use the dropdown box to select the anonymisation technique(s) employed for the selected identifier.

If multiple anonymisation techniques have been applied to the same identifier, select the plus symbol (+) in the top left corner, add the same identifier and each additional anonymisation technique in a new row.

Use the Comments box to include any additional information (please refer to the subsection "Comments column" below for further details).

Please ensure that no empty rows are left in the final report.

<u>Comments column</u>
If applicable, you may complete the Comments box to include any additional information relating to the listed identifier that would provide more context regarding the anonymisation techniques. For example, indicate to which study/studies the technique in question applies, what the transformed, recoded, or generalized term has been changed to, etc.

N.B. If multiple anonymisation techniques apply to one identifier category (or subcategories), list the identifier category separately in the table for each technique (e.g., two rows should be added for the category "Geographical Location – Site Details" if the site identification number and the site address were treated differently). Use the Comments box to include additional information for clarification, such as the subcategories and their associated techniques.

See example below for medical history information and concomitant medications.

	Category	Participant/ Personnel	Anonymisation	Comments
+/-	Records - Medical History	Participant	No (Retained)	For terms/conditions part of the inclusion criteria during studies A and B
+/-	Records - Medical History	Participant	Yes (Generalized)	For Study A: all medical history terms have been generalized to the High Level Group Term (HLGT)
+/-	Records - Medical History	Participant	Yes (Suppressed)	For Study B: all medical history terms have been suppressed
+/-	Records - Concomitant Medications	Participant	Yes (Redacted)	For all concomitant medications used to treat previous ailments, as well as ongoing during studies A and B
+/-	Records - Concomitant Medications	Participant	No (Retained)	For all concomitant medications used to treat adverse events during studies A and B

3. Risk Assessment

F) Please input the selected reference population.

Use the dropdown box to select the appropriate reference population used for the risk assessment in the clinical information package.

CTs: shorthand for clinical trials within the selection options

The selection of the appropriate reference population determines the total participants group size and the level of anonymisation needed to reduce the risk of patient re-identification. The reference population can be informed from patients in the single trial in question (smallest population), all patients in similar trials by a specific study sponsor, all patients in similar trials (e.g., by disease or therapeutic intervention category), or all patients in a geographic area (largest population).

If a reference population other than those listed in the dropdown menu is used, please select "Other" and use the free text box to describe the reference population used.

If more than one reference population is used, please select "Other" and use the free-text field to list and/or describe the reference populations used.

G) Is this product indicated in the treatment of a rare disease/condition?

Select "Yes" to indicate that the product is intended to treat rare disease populations or conditions.

Select "No" otherwise.

__Rare disease__: a life-threatening, seriously debilitating or serious and chronic condition affecting a small number of patients. The definitions of a rare disease in Canada and the EU both indicate a prevalence of fewer than 5 in 10,000 persons.

H) Were special populations involved in the trials?

Select "Yes" to indicate that special populations were involved in the clinical trial(s). If "so, please indicate the characteristics of the special populations and the relevant studies.

Select "No" if there were no special populations involved in the clinical package.

__Special population__: individuals representing some groups in the general reference population who may require dedicated studies (e.g., paediatric, geriatric, pregnant or breastfeeding women). In the context of anonymisation, special populations might warrant special attention as they often represent a specific subset of participants and may also not necessarily share all key characteristics of the general reference population.

I) Please input initial risk of re-identification.

Use the free text box to enter the initial risk of re-identification. For qualitative methods, a value of high, moderate or low should be provided. For quantitative methods, maximum risk observed prior to anonymisation should be provided as either a numerical value (when calculated) or quantitative value (in the absence of a numerical value).

J) Please input target risk threshold.

__Target risk threshold__: A predefined maximum re-identification risk threshold that should not be exceeded after anonymisation has been applied to the set of documents.

Use the free text box to enter the target risk threshold. This risk measurement provides justification for any data transformation performed by the manufacturer based on the anonymisation methodology selected.

For __quantitative__ anonymisation, a numerical value should be entered. Health Canada PRCI and EMA Policy 0070 guidance encourages a 9% re-identification risk threshold (risk=0.09).

For **qualitative** anonymisation, the target risk measurement should be set to a level that ensures there is no serious possibility of re-identification of any study participant or personnel.

K) Please input residual risk.

Residual risk: The risk as assessed after the anonymisation has been performed.

Use the free text box to enter the actual value of the residual risk. A value equal to or below the target value is expected (e.g., under 0.09 for quantitative methods).

If multiple risk calculations have been performed (e.g., for each clinical study), please use the free text box to enter the values of the residual risk obtained for each study.

L) Did some of the above indirect identifiers require consideration due to the sensitivity of the information?

Select "Yes"" if any of the above listed indirect identifiers have required any considerations due to the sensitivity of the information. If "Yes" was selected, use the free text box to list the categories (refer to the "Category" column of the table in "Section 2: Identification of Data Variables", using the full name of the category) of identifiers which were deemed sensitive by the applicant.

Select "No" if there were no sensitivity considerations.

Sensitive information: any information related to the identifiable individual that, if disclosed, could result in significant emotional, social, or financial harm. This information could be medical (e.g., psychiatric disorders, substance use) or non-medical (e.g., road traffic accident, suicide, sexually explicit behaviour) and must be re-identifying in and of itself.

M) In the space below, provide a clear and concise explanation for why the selected methodology (qualitative or quantitative) was used. Please also provide an explanation regarding the limitations of the approach.

Use the free text box to describe why either a qualitative or quantitative approach was used and any limitations associated with this approach.

Please include rationale for the chosen anonymisation techniques (i.e., redaction, transformation, recoding, etc.), with reference to the different sections of the clinical documents (i.e., demographic tables, summaries, narratives, etc.) and explain why the specific technique was deemed the most suitable.

If a quantitative approach is used, please indicated which identifiers were included and/or excluded from the risk calculation.

4. Data Utility

N) List the variables with the highest data utility (up to five).

Use the free text box to list up to 5 identifiers with the highest degree of data utility within the clinical information package. These are typically expected to be known risk factors, confounders or other relevant elements warranting capture in the clinical documents themselves.

Data utility: in the context of a clinical information package, data utility refers to the ability of identifiers (e.g., direct and indirect identifiers denoted in "Section 2: Identification of Data Variables") to provide scientifically useful information with respect to the study population, indication and/or clinical findings described in the clinical information package.

O) How was data utility loss mitigated for these variables?

This section must clearly explain how the applied anonymisation methods have preserved clinical data utility throughout the clinical information package. This section is designed to help the applicant evaluate if the techniques used to anonymise certain identifiers should be reconsidered.

Use the free text box to provide a detailed explanation of how the anonymisation techniques applied to the identifiers listed under "N)" may impact data utility for the reader.

P) Have aggregated tables been appropriately retained?

Select "Yes" if aggregated tables have been retained throughout the documents submitted as part of the clinical information package. If any redactions have been applied to aggregated tables, the question should be answered with "No".

If "No" was selected, use the free text box to clarify what the specific circumstances were that required anonymisation of information within aggregated tables and how this will impact the clinical data utility.

Q) Has a differential approach been taken for the narratives?

Select "Yes" if a differential approach was performed on study narratives. Select "No" if identifiers within narratives were treated the same as identifiers outside of narratives. If the clinical information package does not contain any narratives, please select the option "N/A (No narratives present)".

Of note, narratives **must not** be redacted in full.

If "Yes" is selected, use the free text box to clearly detail the rationale behind this differential approach and why the applicant or MAH considered this different method necessary in terms of the risk assessment and how this impacts the clinical data utility.

5. Deviations

R) In the space below, provide a clear and concise explanation for why the manufacturer has deviated from the anonymisation methodology as set out in the HC/EMA Guidance documents. This may include for example, the use of a qualitative approach (and its impacts on the ability to calculate risk)

Please complete the free text box in this section to clearly state all elements that deviate from the HC/EMA guidance documents. Manufacturers must also provide a detailed justification as to why the deviation had to be chosen over the preferred approaches.

6. Attestation

The "Approve here" check box indicates that the applicant has completed this form, ensured its accuracy, performed all the due diligence associated with the assessment of re-identification and mitigated the risk of re-identification through adequate use of anonymisation.

Please note the box should not be checked until all elements of the form are **finalised,** as checking this box will trigger a "print to pdf" pop-up prompt. Follow the prompt and select "print to pdf" in the print options to create and save the finalised pdf version of the form.

Pop-up prompt once the approval checkbox is checked:
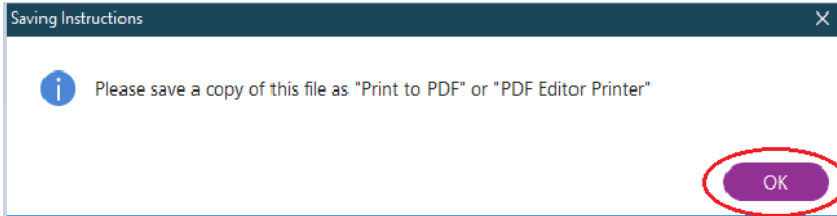
Print preview using "Print to PDF" function:
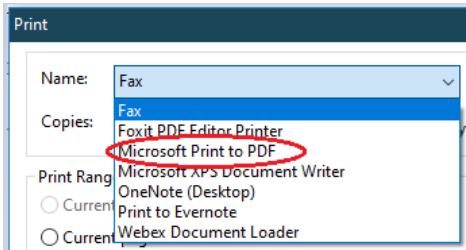
N.B. Whenever the anonymisation report is submitted as part of eCTD sequences (i.e., submission of the final redacted package), it should be submitted as a PDF and not as an active dynamic form as this will invalidate the package.

Definitions of key terms

Academic credentials: an academic rank conferred to an identifiable individual by a college, university, or other post-secondary education institution as official recognition of the successful completion of a study programme

Adverse event: any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have a causal relationship with this treatment. An adverse event (AE) can therefore be any unfavourable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal (investigational) product, whether or not related to the (investigational) medicinal product.

Company name/address: the official and legally recognised name by which a business is known and the street name and building number where the company is located. As an identifier, it needs to be associated with an identifiable individual.

Concomitant medications: any medicine or dietary supplement used by a study participant in addition to the investigational product. Concomitant medication can also include rescue medication used to treat adverse events that occurred during a clinical trial. As with adverse events, medications or procedures used to treat them should be retained by default.

Date offsetting: participant-related calendar dates are consistently replaced with new dates generated using a random offset (dates moved by a specific number of days after or before the original date) and the same offset time period is applied to all dates in the study for each participant to reduce the likelihood of re-identification.

Direct identifiers: identifiers which, on their own, can identify an individual (e.g., name, home address). Direct identifiers are not useful for data analysis purposes (except for the study participant identification number). As a general rule, all direct identifiers must be anonymised.

Generalization: anonymisation technique where the text is replaced by a broader term to dilute the attributes of the data into a larger scale of magnitude and make the initial text less identifiable.

Geographical location: the physical location associated with an identifiable individual, such as the country, city, and site details.

Identifier: any variable part of the transparency package which allows re-identification of an identifiable individual based on three criteria: (1) replicability (the variable values must be sufficiently stable over time so that the values will occur consistently in relation to the individual); (2) distinguishability (the variable must have sufficient variability to distinguish among individuals in the data); and (3) knowability (an adversary must know the identifiers relating to the individual to be able to re-identify them).

Indirect identifiers (also known as *quasi-identifiers*): can be used in combination with other variables to identify an individual (e.g., a combination of age, race, height, and weight might re-identify an individual). For an indirect identifier to require anonymisation, its disclosure must present a serious and reasonable possibility of re-identifying an individual when combined with other available information.

Job title/positions: the name or designation given to a specific role within an organisation or company which may also describe the roles and responsibilities of the identifiable individual (e.g., CEO, Director).

Laboratory values: the measurements collected through laboratory tests which provide diagnostic information regarding an identifiable individual's overall health. This also includes vital sign measurements.

Medical history: collection of an identifiable individual's past or present medical information such as prior diagnoses, prior and concurrent conditions, prior treatments or procedures. This category also includes familial medical history. Specific terms of an identifiable individual's medical history might be used as inclusion criteria in any given clinical trial.

Non-participants: an identifiable individual not enrolled in a clinical study for whom information was captured in the clinical information package, potentially allowing for re-identification of a participant. This may include family members, friends or colleagues.

Organisational department: a distinct functional division or unit within a larger company associated with an identifiable individual and which has its own specific duties, responsibilities, and areas of expertise (e.g., Biostatistics, Clinical Operations, Regulatory Affairs).

Other indirect identifiers: any other indirect identifier not mentioned in the drop-down list, which allows for the indirect re-identification of an identifiable individual. Details about such an identifier should be provided in the comments box.

Recoding: anonymisation technique where the element's original value is consistently replaced while preserving the format of the underlying text.

Redaction: anonymisation technique whereby the text or picture is obscured (masked) preventing access to the underlying information.

Relative day: time in days since a specific reference point for any event or procedure associated with an identifiable individual (e.g., Day 10). Study day is a subtype of relative day, in reference to the protocol-defined study start point, and on which an intervention, procedure, assessment, and/or collection of other study data occurs.

Retention: the original text is unchanged in the public (released) document.

Sex/gender: sex is assigned at birth based on a person's reproductive system and physical characteristics (e.g., female/male). Gender refers to an individual's personal and social identity as a man, woman, or non-binary person (e.g., female/male/other). This also includes gender-related pronouns used by the individual (e.g., she/her, he/him).

Sensitive information: any information related to the identifiable individual that, if disclosed, could result in significant emotional, social, or financial harm. This information could be medical (e.g., psychiatric disorders, substance use) or non-medical (e.g., road traffic accident, suicide, sexually explicit behaviours) and must be re-identifying in and of itself.

Site details: information about a specific clinical trial site location that relates to an identifiable individual. This may include site ID/number, site address, site name, etc.

Study participant: an individual enrolled in a clinical trial to receive specific interventions.

Study personnel: a delegated individual with a study-specific role in a clinical trial. Study personnel roles may include (but are not limited to): principal investigator, site investigator, sponsor medical officer, study coordinator, research personnel, Data Safety Monitoring Board member, etc.

Study role: the specific position or designation of responsibilities of an identifiable individual within the context of a clinical study (e.g., principal investigator, co-investigator).

Suppression: anonymisation technique where the value of the identifier is removed from the documents.

Verbatim text: the exact word-for-word report from the investigator which was recorded without any changes or alterations and which provides granular details about an identifiable individual.