

非特定化／匿名化データ共有時の 透明性のチェックリストについて

TransCelerate 作成のチェックリストより

令和6年10月

日本製薬工業協会

医薬品評価委員会 データサイエンス部会

2023年度 継続タスクフォース2

目次

1. はじめに	3
2. チェックリスト及びその解説	3
2.1 UNIQUE IDENTIFIERS	3
2.1.1 解説	4
2.1.2 アプローチ	4
2.1.3 留意点	5
2.2 DATES	5
2.2.1 解説	5
2.2.2 アプローチ	5
2.2.3 留意点	6
2.3 VERBATIM/FREE TEXT	7
2.3.1 解説	7
2.3.2 アプローチ	7
2.3.3 留意点	8
2.4 BANDING OF VARIABLES	8
2.4.1 解説	9
2.4.2 アプローチ	9
2.4.3 留意点	10
2.5 PATIENT DEMOGRAPHICS (SEX, RACE, ETHNICITY)	11
2.5.1 解説	11
2.5.2 アプローチ	11
2.5.3 留意点	12
2.6 DATA WITH LOW FREQUENCIES	12
2.6.1 解説	12
2.6.2 アプローチ	12
2.6.3 留意点	13
2.7 SENSITIVE INFORMATION	13
2.7.1 解説	14
2.7.2 アプローチ	14
2.7.3 留意点	15
2.8 ADVERSE EVENTS & MEDICAL HISTORY	15
2.8.1 解説	15
2.8.2 アプローチ	16
2.8.3 留意点	16
2.9 CONCOMITANT MEDICATIONS	17
2.9.1 解説	17

2.9.2 アプローチ	17
2.9.3 留意点	17
2.10 GEOGRAPHIC LOCATION	18
2.10.1 解説	18
2.10.2 アプローチ	18
2.10.3 留意点	19
2.11 RECORDS OF PARTICIPANTS WHO HAVE DIED.....	19
2.11.1 解説.....	19
2.11.2 アプローチ.....	19
2.11.3 留意点.....	20
2.12 その他の項目	20
2.12.1 Information Collected Under Copyright Licenses	20
2.12.2 Data Derived from Genomic Data	21
2.12.3 Seasonality	21
2.13 共有するデータについて	22
2.13.1 解説	22
2.14 その他の情報.....	23
2.14.1 解説	23
3. おわりに	24
4. 謝辞	24
5. 資料作成者	24
APPENDIX.....	25
参考文献	27

1. はじめに

臨床試験データをデータ利用者に共有する際、共有するデータ（変数）に対してどのように非特定化を行ったのかを示すことが推奨されている¹²³。2023年8月 TransCelerate BioPharma Inc.

（以後、TransCelerate）よりデータ共有する際の非特定化方法に関するチェックリストの Example⁴ が公表された。この透明性のチェックリスト⁴は、共有するデータの作成者が各データの非特定化方法の点検に活用可能なだけでなく、チェックリストのカラムを埋めることによりどのような非特定化手法を適用したのかを示す資料にもなり、サポート情報の一つとしてデータ利用者へ提供可能である。ただし、個人情報の復元につながる情報を含めてはならない。

TransCelerate の公表物⁵は、一般に公表されているものであり、共有するデータの非特定化手法の例もいくつかわかりやすく解説されている。この度、日本語による報告書を公表することについて TransCelerate の合意を得られたため本報告書の作成に至った。2章において、チェックリストの項目とそれらの非特定化手法の解説及びカラムに記載すべき例を提示し、日本語でも理解可能・利用可能となるような報告書とした。また、Appendix として日本語に対応したチェックリストも用意した。なお、企業ポリシーにより本内容と同等の文書やチェックリスト等をすでに作成しているのであれば、追加で作成する必要はない。また、適切であれば、現在、使っている非特定化の方法をそのまま利用することもできる。臨床試験データを非特定化することで、データ提供者だけでなくデータ利用者にとっても、参加者のプライバシーを保護しつつ、データの2次利用に有用であるという理解の促進の一助になれば幸いである。

2. チェックリスト及びその解説

本章では TransCelerate の公表物⁵を参考に、透明性のチェックリスト⁴の内容を解説し、各カラムをどう記載するのかについて継続タスクフォース2で追記した。

非特定化手法はいくつかあり、TransCelerate では推奨アプローチ（Recommended approach）、代替アプローチ（Compatible approach）として紹介されている。データ提供者は、データ共有後に臨床試験の参加者が再特定されるリスクを勘案し、どのアプローチを利用するか決める必要がある。これらのアプローチは、単独ではデータを匿名化するのに十分ではないため、既に利用されている非特定化の手法と併せて使用されるものであることに注意する。また、関連するすべての地域／国の法律に準拠する必要がある。まず、チェックリストの例示をそのまま表示し、解説・アプローチ・留意点を記載した。

2.1 Unique Identifiers

（例示）

	選択したアプローチ	適用したアプローチの詳細
<u>Unique Identifiers</u> <u>（固有識別子）</u> <i>NOTE: Where identifiers have been removed, the rationale should be</i>	Recommended approach	Relevant unique identifiers have been scrambled

	選択したアプローチ	適用したアプローチの詳細
<i>described</i>		

2.1.1 解説

臨床試験データの中には、そのデータ単独でもしくは複数のデータを併せることで、参加者を一意に特定できる可能性のある大量のデータやイベントが収集される可能性があり、再特定化リスク（参加者が特定されるリスク）が生じる。したがって、データ内のすべての変数において、再特定化の可能性を臨床試験ごとに評価して、その変数が参加者に一意に紐づく情報を提供する可能性があるかどうかを判断する必要がある。Unique Identifiers は、2つのカテゴリに分けられる。

- Direct Identifiers（直接識別子）：それ自体が個人を特定するもの
例：症例番号、シリアル番号、外れ値など
- Indirect Identifiers (Quasi-identifiers)（準識別子）：他の情報と合わせることで個人を特定可能とするもの
例：患者背景情報、施設番号、ベンダー番号、バッチ／ロット番号、試験薬の投与履歴

2.1.2 アプローチ

■推奨アプローチ（Recommended Approach）：

直接識別子に対しては、データセットとサポート文書（データセットの内容や非特定化情報を記載した文書）間の一貫性を持たせて、元の値と同じ長さ（Length）、タイプ（Type）、形式（Format）でスクランブル（他の記述等へ非可逆的に置き換え、Scramble）することが推奨される。

準識別子に対しては、臨床試験ごとに非特定化が必要かどうかを評価することが推奨される。非特定化の必要がない場合はそのまま残し、必要がある場合は、可能であれば編集（Redact）や隠す（Suppress）のではなく、スクランブル（Scramble）する。

<適用したアプローチの詳細の記載例>

直接識別子および（必要に応じて）準識別子は、同じ Length, Type, Format でスクランブルに変換した。

■代替アプローチ（Compatible Approach）：

代替アプローチは、参加者の識別番号を除くすべての一意の識別子を編集、削除（Remove）、または上書き（Overwrite）することである。参加者の識別番号は主たるキー変数とみなされ、スクランブル形式で維持する必要がある。ただし、編集した場合、その根拠はサポート文書に記載される必要がある。

<適用したアプローチの詳細の記載例>

参加者の識別番号（USUBJID）のみ同じ Length, Type, Format でスクランブルに変換し、それ以外の直接識別子および（必要に応じて）準識別子は削除した。

2.1.3 留意点

Unique identifiers は、個々の参加者に固有であることや、他の情報源（例えば、試験情報登録）を介して臨床試験に関する追加情報へのリンクを提供する可能性があることにより、参加者の再特定化リスクを増加させる。PhUSE のガイダンス¹では、直接識別子を削除することを提案している。しかし、直接識別子は、データ分析のキー（例えば、データの値と参加者の紐づけ情報となる変数）としてしばしば必要となるため、削除するのではなく、スクランブルすることが推奨される。

準識別子についても参加者の再特定化リスクを評価する必要がある。特定の準識別子を保持することで生じる再特定化リスクとその準識別子を保持することでもたらされるデータの有用性とのバランスを、他の準識別子を非特定化处理した後に吟味すべきである。また、科学的または医学的な観点から、準識別子の保持を必要とする特定の要件があるかどうかを考慮することも重要である。可能な限り準識別子は保持し、再特定化リスクが高すぎる場合は、準識別子を削除するのではなくスクランブルすべきである。

2.2 Dates

(例示)

	選択したアプローチ	適用したアプローチの詳細
<u>Dates</u> (日付)	Recommended approach	Dates offset

2.2.1 解説

臨床試験データの日付変数は有用な情報を提供するが、例えば、別のデータソースとの組み合わせによって日付情報のパターンから個々の参加者が再特定される可能性もある。そのため、データ提供者は、個人に関連する日付を変換する必要がある。臨床試験データには、Relative Day（相対日付）と実日付の両方が存在する。相対日付からは一般的に個人が特定されないため、そのままにしておくことができるが、実日付が存在する場合は、変換する必要がある。

2.2.2 アプローチ

■推奨アプローチ（Recommended Approach）：

日付変数に対しては、元の日付に定義された日数を加算または減算することによって、元の日付を変換する Date Offset（オフセット法）が推奨される。この方法はさまざまな日付範囲で使用でき、参加者ごとに1つのランダムな日数をオフセットして使用することができる。オフセットした日数は、非公開とする。

<適用したアプローチの詳細の記載例>

実日付に対して参加者ごとに一意のオフセット法を使用した。

■代替アプローチ（Compatible Approach）：

代替アプローチは相対日付を用いる。この手法は、特定の日付（例えば、各参加者のランダム化日、試験薬投与開始日など）を基準日（Day1）として使用し、その特定の参加者の他のすべての日を基準日からの相対的な日数に変換する方法である。この場合、元のすべての日付変数は相対日付に編集（Redact）される。

表 1 Example of Redaction of Dates

	BEFORE	AFTER
SUBJID	DMDAT	DMDAT
1234-5678-10001	2024-01-01 (基準日)	1
1234-5678-10001	2024-01-05	5
1234-5678-10001	2024-01-10	10

<適用したアプローチの詳細の記載例>

その参加者の試験薬投与日を Day1 として、実日付を相対日付に変換し、実日付は削除した。

2.2.3 留意点

日付データはさまざまな目的で収集されるが、そのすべてがセンシティブ情報とみなされるわけでも、再特定化リスクを引き起こすわけでもない。例えば、薬剤バッチ日や臨床試験の開始日・終了日などは、再特定化リスクと関連する可能性が低いため、通常は変換する必要はない。

オフセット法を使用する場合、適切なオフセット範囲（例：±30 日、±90 日）の選択は臨床試験に依存する可能性がある。例えば、狭い範囲内の日付の割合が高い臨床試験の場合は、小さなオフセット範囲で十分かもしれないが、登録期間が長期にわたる臨床試験では、参加者のプライバシーを保護するためにより広いオフセット範囲が必要になる場合がある。また、部分日付または補完日付をオフセットする場合は、オフセット日付の作成に使用された日数が明らかになり、元の日付が特定される可能性があるため、特別な注意が必要である。例えば、臨床試験における補完は、多くの場合標準ルールに従う（つまり、欠測日は 01（1 日）によって補完され、欠測月は 01 または 06（1 月または 6 月）によって補完される）。これらのルールは、多くの場合、統計解析計画またはその他の仕様書等で説明されており、公開される可能性がある。したがって、どのデータで日付が補完されたかが知られてしまった場合に、補完された値を非特定化された値と比較することによって、オフセット範囲が明らかになる可能性があることに注意する。

リスクを最小限に抑え、高いデータ有用性を維持するには、補完日付に関連するすべての情報、つまり補完日付フラグ（CDISC ADAM *DTF 変数）および不完全な日付変数を削除することが推奨される。

季節性が重要な研究の場合、データ提供者は季節情報を維持するために狭い帯域（夏季冬季や雨期乾季などが変わらない期間内）を適用できるかどうかを判断する必要がある、オフセットの最小値と最大値を共有データの非特定化サポート文書に記載することも考慮する。

さらに、データ共有により、複数の臨床試験データを統合することが考えられることから、データを加工する場合は、CDISC 標準（例：YYYY-MM-DD の形式）にしておくことが望ましい。

2.3 Verbatim/Free Text

（例示）

	選択したアプローチ	適用したアプローチの詳細
<u>Verbatim/Free Text</u> (逐語的なテキスト／フリーテキスト)	Recommended approach	All fields have been marked as redacted

2.3.1 解説

逐語的なテキスト及びフリーテキストには文字列として個人データが含まれる可能性がある。

逐語的なテキストはコーディング目的で 사용되는場合があり、例えば有害事象の医師記載名は MedDRA コーディングのためのテキストとして収集され、度数表などの解析にコード化された変数が使用される。

その他のテキスト変数は、収集された他のデータフィールドに貴重な洞察や情報を提供する可能性がある。しかし、テキスト変数は非構造化データであるため、データをどう非特定化処理すべきかの分析が困難となり、テキスト中の個人の特定につながる情報部分を編集（Redact）するには、各データの値ごとに吟味が必要となる可能性がある。

2.3.2 アプローチ

■推奨アプローチ（Recommended Approach）：

フリーテキストフィールドは、変数や列が削除されたのか元々ブランクなのかを区別できるように、ブランク以外の値で編集する。

編集が必要な逐語的なテキスト及びフリーテキストフィールドでは、編集後のフィールドまたは特定の文言を文字列「-redacted-」に置換する（表 2）。特にセンシティブ用語（[2.7 Sensitive Information](#) 参照）は、データ提供者のポリシーに基づいて削除し、「-redacted-」等の用語に置換すべきである。

表 2 Example of Redaction of Verbatim/Free Text

		BEFORE	AFTER
STUDYID	SUBJID	COVAL	COVAL
1234-5678	1234-5678-10001	Comment 1	-redacted-

		BEFORE	AFTER
STUDYID	SUBJID	COVAL	COVAL
1234-5678	1234-5678-10002	Comment 2	-redacted-
1234-5678	1234-5678-10003	He got Huntington disease on January 1, 2024.	He got -redacted-

<適用したアプローチの詳細の記載例>

全てのフリーテキストフィールドを「-redacted-」に置換した。

■代替アプローチ（Compatible Approach）：

フィールドを「-redacted-」等の用語置換でなくブランクにして個人データを保護する。その際、データの有用性を確保するために、次の①かつ／又は②の追加情報が必要である。

- ① 研究チームに提供するためのサポート文書内に、ブランクにされた変数／列の詳細を記載した包括的な情報
- ② 削除された情報の代用として提供される派生情報／変数

<適用したアプローチの詳細の記載例>

個人データ保護のため COVAL のテキストデータをブランクに変換したが、重篤な有害事象又は中止に至った有害事象の医師コメントは CTD 本文内に叙述した。

2.3.3 留意点

フリーテキストには、プライバシーリスクに影響を及ぼす可能性のある、稀な名称、疾患名、治療名、または稀な有害事象などの情報があり、これらは再特定化リスクを伴う可能性がある。一方で、研究や組織によってフリーテキストの使用方法は大きく異なる可能性がある。そのためデータ変換プロセスでは、変換の労力、データの有用性、プライバシーリスクの間の効果的なバランスを考慮する必要がある。

組織によっては、センシティブ用語（[2.7 Sensitive Information](#) 参照）のフリーテキストを電子的にスキャンするツールを使用している場合がある。研究データベース内にある逐語的なテキスト及びフリーテキストフィールドの数によっては、手動で全フリーテキストフィールドの個人情報を確認すると、リソースを大量消費する可能性がある。

2.4 Banding of Variables

(例示)

	選択したアプローチ	適用したアプローチの詳細
<u>Banding of Variables</u> (変数のバンディング)	Compatible approach	Semifixed banding applied to Height

2.4.1 解説

バンディングは、年齢、身長、体重、BMI 等の連続変数をバンド（例：20～29、30～39 等のカテゴリ値）に集約する手法であり、特定の数値を範囲に変換することで参加者の再特定化リスクを低くすることができる。

バンドサイズ（値の範囲）は、バンドサイズを狭くすることでデータの有用性を最大化することと、バンドサイズを広げることで再特定化リスクを下げることとのバランスを考慮する。

【バンディング手法】

		説明	例
Static bands		各臨床試験データで一律に全ての値を 10 歳刻み等の同じ範囲で区切る	20～29, 30～39
Semi-fixed bands		該当人数が極端に少ないバンドが発生しないように、Static bands を適宜組み合わせることで再特定化リスクを軽減する	20～29 の範囲が一人しかない場合に 30～39 の範囲と合わせて 20～39 にする
Flexible band 再特定化リスクを抑えながらデータの有用性を極力維持するべく、データセット毎に各バンド内の人数が適切（例：各バンドの人数が最低でも 3 人以上）となるように個別のバンドを設定する（各バンドの範囲の大小は不問）。 一次元と多次元の 2 種類がある。	一次元バンディング（single-dimensional banding）	データセット内の関連する各変数で独立にバンドを設定する	年齢：20～29, 30～39, ... 身長：150～159, 160～169, ...
	多次元バンディング（Multi-dimensional banding）	2 つ以上の変数を組み合わせてバンドを設定する（例えば、変数 A のバンドと変数 B のバンド間のどの組み合わせでも組み合わせ内の人数が x 人以上になるようにする）	20～49 コレラ感染 20～49 水俣病 50～79 コレラ感染 50～79 水俣病

2.4.2 アプローチ

■推奨アプローチ（Recommended Approach）：

Single-dimensional flexible banding を用いて変換する。

例えば年齢のバンドで、範囲が 1 年（例：50）のバンドがあった場合、変数が意図的に変更されたことを明示するため、バンドを「年～年」（例：50～50）と表記する。（表 3）

表 3 Examples of Single-Dimensional Flexible Banding of Age and Height

		BEFORE	AFTER	BEFORE	AFTER
STUDYID	USUBJID	AGE	AGE	HEIGHT	HEIGHT
1234-5678	1234-5678-10001	46	46～48	153	150～159
1234-5678	1234-5678-10002	47	46～48	161	160～169
1234-5678	1234-5678-10003	48	46～48	175	170～179
1234-5678	1234-5678-10004	50	50～50	168	160～169
1234-5678	1234-5678-10005	50	50～50	163	160～169
1234-5678	1234-5678-10006	50	50～50	180	180～189
1234-5678	1234-5678-10007	50	50～50	210	≥190

<適用したアプローチの詳細の記載例>

DM ドメインの年齢データと VS ドメインの身長データに Single-dimensional flexible banding を適用した。

■代替アプローチ（Compatible Approach）：

Semi-fixed bands を用いて変換する。しかし、全体的なデータ有用性が低下する可能性がある。

<適用したアプローチの詳細の記載例>

VS ドメインの身長データに Semi-fixed bands を適用した。

2.4.3 留意点

ほとんどの場合、数値の準識別子にバンディングなどの変換を適用しないと、許容できない再特定化リスクが生じることがある。

Static bands は推奨しない。推奨しない理由は、例えば各バンドの該当人数が3人以上となるように全試験の値を同じ範囲で区切ろうとすると、全て 20 歳区切りの様に十分に広い範囲幅が必要となってしまい、データの有用性が低下してしまう可能性があるためである。

Semi-fixed bands と Flexible bands では、これらが適用された複数の臨床試験データを結合する際、全試験のうち最大のバンドを適用する必要がある。したがって、Semi-fixed bands を適用した方が複数の臨床試験間でバンドのパターン数が少なくなり、変数間の一貫性が向上する可能性があるため、複数データの結合を考えると Semi-fixed bands が望ましい。

Multi-dimensional banding は推奨しない。推奨しない理由は、データ有用性の著しい上昇が期待される一方で、その技術はまだ広く採用されておらず、迅速に採用することができない可能性があり、データ提供者側の統計担当者の業務に影響を及ぼす可能性があるためである。

バンドの表現として、範囲（例：20～29）ではなく単一の数値（例：25）で表現することが望ましい場合、特定の数値を“random number in band”または“mean within band”等の用語に置換することができる。その際、透明性のためにこの手法が使用されたことが明確に分かるよう文書

化する。

2.5 Patient Demographics (sex, race, ethnicity)

(例示)

	選択したアプローチ	適用したアプローチの詳細
<u>Patient Demographics (sex, race, ethnicity)</u> (患者背景：性別，人種，民族)	Recommended approach	Patient ethnicity removed due to the low frequency in the dataset of some groups and there is no statistical evidence that indicates different ethnicities have an impact on the findings of the study

2.5.1 解説

性別，人種，民族，年齢などの情報は，データ共有や再利用のために保持する価値があるが，他の準識別子と組み合わせられることにより，全体的な再特定化リスクが高まる可能性がある。

一方で，人口統計学的データを集約して解析することで，多様な集団における病状や治療効果をより深く理解でき，その解析から得られた洞察を将来の臨床試験に前向きに組み込んで包括的な研究を進めることができる。

データ共有の際は，人口統計学的データの一部を編集したりグループ化したりすることで，プライバシー保護とデータ有用性のバランスをとる必要がある。

2.5.2 アプローチ

■推奨アプローチ (Recommended Approach) :

性別，人種，民族はできるだけ多くの情報を保持する(件数の少ないデータ (low frequency events) は [2.6 Data With Low Frequencies](#), [2.7 Sensitive Information](#) で考慮)。

再特定化リスクは，他の準識別子（年齢，身長，体重等）と組み合わせて評価する。

各組織で設定した再特定化リスクの閾値を超える場合，件数の少ないデータ (low frequency event(s)) は削除する。

人口統計学的属性が保持できない場合，性別関連情報（妊娠検査，勃起不全等）や人種・民族関連情報（地域関連性等）も削除し，研究チームに提供するサポート文書内で，削除した旨とその理由について説明する。

<適用したアプローチの詳細の記載例>

患者背景の民族データに件数の少ないデータ (group) があるため，民族データは「-redacted-」に置換し，サポート文書に説明を記載した。なお，異なる民族性が研究結果に影響を与えることを示す統計的証拠はない。

■代替アプローチ (Compatible Approach) : なし

2.5.3 留意点

件数の少ないデータ（low frequency groups）が存在しない限り、性別、人種、民族の情報を保持する。

外れ値は分析において重要な要素であるが、特定の参加者の再特定化リスクが高まる可能性があるため、特別な考慮が必要である。

参加者のノンバイナリーな性別情報（“BOTH”，“Unknown”，“Undifferentiated”）は、相対的に希少なため考慮が必須である。

PhUSE¹は、一部の国で人種データを収集しておらず、かつ、国情報を（大陸レベルなどに）一般化する場合は、人種データを収集できない国が（人種データがないことにより）明らかにならないよう人種データを削除する必要があると記載している。

2.6 Data With Low Frequencies

（例示）

	選択したアプローチ	適用したアプローチの詳細
<u>Data With Low Frequencies</u> (件数の少ないデータ)	Recommended approach	A minimal frequency of 1 out of 11 has been used for race

2.6.1 解説

件数の少ないデータを含む変数は、再特定化リスクを高める可能性がある（例：いくつかの変数を各水準でグループ化した際に、セル又はグループのサイズが極めて小さくなるデータ等）。その変数が複数になると、再特定化リスクがさらに悪化する可能性がある。

2.6.2 アプローチ

■推奨アプローチ（Recommended Approach）：

件数の少ないデータを含む変数を評価し、再特定化に重大なリスクがある場合は、件数の少ないデータの編集（redact）を推奨する（表 4、表 5）。

企業が定めた最小頻度値（Minimum Frequency Value）^aの適用を推奨する（例：5 件）。

また、研究チームに提供するサポート文書内で、件数の少ないデータの削除または編集について説明する。

表 4 Example of Redacting Low Frequency Sex and Race

^a PhUSE は成果物 ”The De-identification Standard for Study Data Tabulation Model (SDTM) 3.2” の中で、「特定条件下の安全なポータルでのデータ共有であれば、最小頻度 5（又は確率 0.2）が許容される可能性がある」と述べている。

		BEFORE	AFTER	BEFORE	AFTER
USUBJID	DOMAIN	SEX	SEX	RACE	RACE
1234-5678-USA003-10001	DM	F	-redacted-	WHITE	WHITE
1234-5678-DNK001-10002	DM	F	-redacted-	WHITE	WHITE
1234-5678-POL002-10003	DM	M	-redacted-	WHITE	WHITE
1234-5678-GER002-10004	DM	F	-redacted-	UNKNOWN	-redacted-

表 5 Example of Redacting Additional Information Revealing Trial Participants' Sex Through a Rare Event (e.g., Event of Oligospermia^b)

		BEFORE	AFTER
USUBJID	DOMAIN	AEDECOD	AEDECOD
1234-5678-POL002-10003	AE	Oligospermia	-redacted-
1234-5678-POL002-10003	AE	Headache	Headache
1234-5678-POL002-10003	AE	Diarrhoea	Diarrhoea
1234-5678-POL002-10003	AE	Rhinitis	Rhinitis

<適用したアプローチの詳細の記載例>

DM ドメインの SEX と RACE に件数の少ないデータが存在したため、SEX の全データ及び RACE の一部データを「-redacted-」に置換し、サポート文書に説明を記載した。

■代替アプローチ (Compatible Approach) : なし

2.6.3 留意点

特定の件数の少ない臨床試験データから重要な知見が得られる可能性があるため、通常、臨床試験では稀な事象や外れ値を解析することに特別な関心を持たれる。

そのため、参加者のデータが非常に件数の少ないグループに分類された場合、再特定化リスクが著しく増加してしまうため、これらの臨床試験データを保持すべきか慎重に検討する必要がある。

2.7 Sensitive Information

(例示)

	選択したアプローチ	適用したアプローチの詳細
<u>Sensitive Information</u>	Recommended approach	All relevant values have been

^b 精子減少症

	選択したアプローチ	適用したアプローチの詳細
(センシティブ情報)		marked as redacted

2.7.1 解説

センシティブ情報^cは極めて個人的なものであり、情報の開示は個々の参加者に危害を及ぼす可能性がある。

(例)・アルコール乱用

- ・薬物使用
- ・HIV/AIDS などの疾患
- ・メンタルヘルス情報に関するデータ

臨床データセット内で何がセンシティブ情報とみなされるかは、データセットが関連するコンテキストと治療領域によって異なる。例えば HIV に言及するデータは、HIV/AIDS 研究のデータセットではセンシティブ情報とみなされないが、肥満研究のデータセットではセンシティブ情報とみなされる可能性がある。

2.7.2 アプローチ

■推奨アプローチ (Recommended Approach) :

全変数でセンシティブ情報は編集 (Redact) し、研究チームに提供するサポート文書内で説明する。治療分野によって異なるため、どの変数をセンシティブ情報に分類するか方針を決めておく (表 6)。

表 6 Example of Removing Sensitive Substance Usage Records

			BEFORE	AFTER
USUBJID	DOMAIN	SUCAT	SUOCCUR	SUOCCUR
1234-5678-10001	SU (Substance Usage)	ALCOHOL HISTORY	N	-redacted-
1234-5678-10001	SU	TOBACCO HISTORY	Y	-redacted-
1234-5678-10004	SU	ALCOHOL HISTORY	Y	-redacted-
1234-5678-10004	SU	TOBACCO HISTORY	Y	-redacted-

^c 個人の思想・信条や国家機密など、きわめて慎重に扱われるべき情報のこと。機微情報ともいう。定義として、OECD の個人情報保護ガイドラインでは、広く、『(情報漏えいによって) 社会的差別を受けうる情報』と規定されている。

また、個人情報保護に関するコンプライアンスプログラムの要求事項 (JIS Q 15001: 2006) では、センシティブ情報の具体的項目に関して、以下の 5 項目を挙げている。

(1) 思想及び信条に関する事項、(2) 政治的権利の行使に関する事項、(3) 労働者の団体交渉に関する事項、(4) 医療、性に関する事項、(5) 犯罪の経歴、人種、民族、社会的身分、門地並びに出生地及び本籍地など社会的差別の原因となる事項。

<適用したアプローチの詳細の記載例>

SU.SUOCCUR の全データを「-redacted-」に置換し、サポート文書に説明を記載した。

■代替アプローチ (Compatible Approach) : なし

2.7.3 留意点

特定の臨床研究では、アウトカムを決定するために、センシティブ情報の収集及び解析が必要となることもある。このセンシティブ情報を共有することは、再特定された場合に臨床試験参加者に危害を及ぼすリスクが高い。潜在的な危害の重大性を考慮し、センシティブ情報を共有する場合は、再特定化リスクの評価において追加の措置（例：情報セキュリティにおける保護措置、利用者のアクセス管理、特定の使用事例の制限）を考慮しなければならない。

センシティブ情報は、本書で提案された方法論で既に言及されているいくつかの変数（例：2.5 Patient Demographics (sex, race, ethnicity)）にも関連する可能性がある。臨床試験の領域に応じて、これらの変数も本項の推奨に従って評価すべきである。

データフィールドが変換／変更されたのか、元のデータセットで欠測していたのかをデータ利用者が区別できるように、データ削除よりも編集が望ましい。

2.8 Adverse Events & Medical History

(例示)

	選択したアプローチ	適用したアプローチの詳細
<u>Adverse Events & Medical History</u> (有害事象及び病歴) NOTE: If any MedDRA levels are removed, please describe the reasons behind the removal.	Recommended approach	LLT has been removed due to limited trial size and rarity of disease. MedDRA version 27

2.8.1 解説

臨床試験中の有害事象の収集は、試験薬の安全性を評価する上で重要な役割を果たす。同様に、参加者の病歴と併存疾患を理解することは、安全性と有効性を評価するのに役立つ。MedDRA 辞書には、有害事象用語を 5 つのレベル^dで記述するためのコード^eがある。病歴の日付または併用薬の日付の値は、日付に関する推奨事項に従う必要がある (2.2 Dates 参照)。

^d 5 つのレベルは階層構造となっており、上位から順に、器官別大分類 (System Organ Class ; SOC), 高位グループ語 (High Level Group Terms ; HLG), 高位語 (High Level Terms ; HLT), 基本語 (Preferred Terms ; PT), 下層語 (Lowest Level Terms ; LLT) と呼ばれる。上位語はその上位語に属する各々の下位語を包括する広範囲のグループ用語となっている。

^e 用語へ固有に付与される 8 桁の数字コードと、用語のテキスト文字列としてのコードがある。

2.8.2 アプローチ

■推奨アプローチ (Recommended Approach) :

使用する MedDRA 辞書のバージョンと共に、MedDRA 辞書コードの 5 つのレベル全てを共有する。この情報は研究チームに提供するサポート文書内に記載する。逐語的な AE 医師記載名 (AETERM) は編集し、対応する MedDRA 用語に置き換える (2.3 Verbatim/Free Text 参照)。参加者の再特定化リスクを引き起こす稀な病歴または有害事象をデータ提供者が特定した場合、参加者のプライバシーを保護するためにこれらの MedDRA 用語も編集する。また、共有データが二次研究目的に適しているかどうかをデータ利用者が評価できるように、適用した編集手順を研究チームに提供するサポート文書内に記載する。

<適用したアプローチの詳細の記載例>

AE ドメインの AETERM の全データを「-redacted-」に置換し、サポート文書に説明を記載した。また稀な疾患名を特定したため、参加者のプライバシー保護を目的として該当疾患の AELLT, AELLTCD, AEDECOD, AEPTCD を「-redacted-」に置換し、サポート文書に説明を記載した。

■代替アプローチ (Compatible Approach) :

2.3 Verbatim/Free Text, 2.6 Data With Low Frequencies, 2.7 Sensitive Information, 2.9 Concomitant Medications 等を総合的に考慮して、再特定されるリスクが非常に高く、5 つの MedDRA レベルすべてを共有することができないと評価された場合は、共有するレベルを減らしてもよい (例: LLT を共有しない)。これにより、個人のプライバシーリスクが低下するが、臨床試験データの有用性が低下する可能性もある。しかし、有害事象や病歴は一般的に SOC 及び PT レベルで安全性の解析が実施されるため、LLT を共有しても付加的有用性は最小限である。

<適用したアプローチの詳細の記載例>

罹患者数が少数又は治療内容が特殊で稀、又はセンシティブ情報が含まれる疾患名を特定したため、参加者のプライバシー保護を目的とし、該当疾患に対する SUPPMH 及び AE ドメイン内の LLT データを削除した。

2.8.3 留意点

有害事象や病歴に関して、MedDRA コードの 5 レベルのうち PT や LLT のように詳細なレベルがより多く共有されるほど、その情報から確認できる有用性のレベルも高くなる。しかし、有害事象の記述が緻密であればあるほど、これらのデータを用いて、又はこれらに関する他のデータと組み合わせて、参加者が再特定されてしまうリスクが大きくなる。

有害事象の性質及び他の情報を組み合わせることによって特定の参加者の再特定化リスクが高まるかどうかを検討する必要がある。

5 つの MedDRA コードレベルすべてを共有できるかどうかを検討する際は、再特定化リスクに影響を与える可能性があるため、有害事象や病歴がどの程度一般的であるかを考慮すべきであり、

非常に一般的な有害事象はリスクが低いと考えられ、より稀な有害事象は再特定されるリスクが高いと考えられる。

2.9 Concomitant Medications

(例示)

	選択したアプローチ	適用したアプローチの詳細
<u>Concomitant Medications</u> (併用薬剤) <i>NOTE: The version information provided here unless it is provided in a variable in the dataset.</i>	Recommended approach	no changes, version included in dataset

2.9.1 解説

研究者が薬剤情報を使って交絡要因の分析などをする上で、正確な薬剤情報は重要である。WHO の ATC 分類^fのデータであれば、統一的な取り扱いができる。

2.9.2 アプローチ

■推奨アプローチ (Recommended approach) :

5 つの Level の ATC コードとその Version を共有する。また、医師記載 (CMTRT/CMMODIFY) は削除する。再特定化リスクを考慮し、ATC コードのレベル 1 からレベル 4 もしくは、レベル 1 からレベル 3 に限定しても良い。その場合、チェックリストの中で研究者に共有する ATC コードのレベルを減らした理由を記載する。

<適用したアプローチの詳細の記載例>

センシティブな薬剤名を考慮してレベル 1 からレベル 3 に限定した。薬剤辞書の Version は、〇〇〇である。

■代替アプローチ (Compatible Approach) : なし

2.9.3 留意点

薬剤辞書は定期的に更新されるので、その Version 情報を共有することは重要である。また、

^f 薬の消費量に関する国際的な統計を取るために考案された医薬品の分類法。DDD : Defined daily dosage (主な適応症に対する成人の想定平均 1 日用量) と併せた ATC/DDD システムとして WHO が推奨し、WHO の医薬品統計方協働センターで統括管理されている。薬効、作用部位・器官および科学的特徴によって 5 段階のレベルで分類され、原則的に一成分の主薬効に対して 1 つのコードが付与される。ただし、複数の薬効が治療上の異なる用途を示すことが明確な場合には、その薬効別に複数のコードが与えられる場合もある。

WHO の ATC 分類を使う場合は、データ提供者とデータ利用者の双方が、有効なライセンスを持っている必要があるだろう。ATC 分類をすべて共有する場合、再特定化リスクが増加するかもしれないことに留意する必要がある。例えば、医薬品の商品名から地理的な情報の特定につながる可能性があったり、薬剤の使用情報自体がセンシティブであったり、薬剤情報から間接的にセンシティブな情報（HIV、精神疾患、希少疾患など）がわかってしまったりすることがある^g。

2.10 Geographic Location

(例示)

	選択したアプローチ	適用したアプローチの詳細
<u>Geographic Location</u> (地理的な位置)	Recommended approach	Generalisation to “Middle East” region due to low number of participants and countries in region, other left at country level

2.10.1 解説

地理的な位置は準識別子であり、他のデータと紐づけることにより参加者の特定につながる可能性があることに留意する必要がある。

2.10.2 アプローチ

■推奨アプローチ (Recommended approach) :

可能な場合は元の国情報を保持するか、収集されたデータにできるだけ近づける。大陸レベルに一般化することは、単一施設の国や参加者が非常に少ない国の場合にのみ推奨される。国別情報を地域または大陸に一般化することで、地域別または国別の分析のためのデータの有用性を維持しながら、参加者の保護もできる。地理情報の一般化のために、ISO-alpha3 および M49 (UNSD – 方法論) Standard^hが利用できる。地理的な場所のカテゴリとして「その他」を使用しないことが推奨される。臨床試験実施施設の識別子や関連情報（住所、電話番号など）等により、地理的位置を間接的に示す可能性があることに注意が必要である。これらに該当する識別子も考慮（変更）する必要がある。

^g 例えば、抗 HIV 薬である zidovudine and lamivudine のコード (5 レベル) は J05AR01 である。5 レベルだと薬剤名より、HIV に対する治療がわかってしまう。4 レベル (J05AR) にしても Antivirals for treatment of HIV infections, combinations となり、依然として HIV に対する治療薬が使われている情報が残ってしまう。3 レベル (J05A) : DIRECT ACTING ANTIVIRALS にすることにより、HIV に対する治療という情報が削除することが可能となる。

^h ISO-alpha3 は、国や地域を英語 3 文字で表示したもので、日本は JPN となる。アジアは ABB、ヨーロッパは EEE と表示でき、人数が十分多い国はそのまま表示、人数が少ない国は大陸のコードを当てはめるということがありうる。同様に、M49 standard は、国や地域を 3 桁の数値で表したもので、国連事務局の統計部門によって作成されている。アジアが 142、東アジアが 030、日本が 392 と設定されており、人数が少ない国や地域の情報は上位の地域にまとめることにより、有用性と再特定化リスクの両方に対処することができる。

<適用したアプローチの詳細の記載例>

国ごとの参加者が少ないので大陸レベルに一般化した。

■代替アプローチ (Compatible Approach) : なし

2.10.3 留意点

一般に、地理的な位置情報を保持すると、特に複数の研究を統合する場合に、データの有用性を最大化するのに役立つ。ただし、国ごとの施設や参加者が非常に少ない場合、地理情報は、個々の参加者の再特定につながる。このリスクを軽減するために、大陸レベルに一般化することで、参加者の再特定が難しくなる。ただ、大陸レベルでも一定の地理情報を保持することができる。地理的位置の代わりに「その他」または同様の非記述的な表現を使用することは好ましくない。医療制度・医療環境や患者背景集団が大きく異なる国が「その他」に入っている場合は、分析時に問題になる。季節が分析の重要な因子であるならば、北半球、南半球、赤道を示す変数を追加する。データ共有者は、季節を分析する変数の追加による再特定化リスクを確認する（増加させない）必要がある。

2.11 Records of Participants Who Have Died

(例示)

	選択したアプローチ	適用したアプローチの詳細
<u>Records of Participants Who Have Died</u> (死亡した患者の記録)	Recommended approach	N/A

2.11.1 解説

臨床試験の参加者が、臨床試験の間もしくはその後に亡くなること (during or post-treatment) がある。また、一部の国では、死後（死後 10 年など）でもプライバシー規制が適用される場合があり、データ共有者は、その規制に準拠する責任がある。

2.11.2 アプローチ

■推奨アプローチ (Recommended Approach) :

死亡した参加者の記録は、生きている人の記録と同様に扱う。参加者が亡くなったとしても、その個人に対して必要な非特定化／匿名化のレベルは変わらない。つまり、生きている参加者のデータに適用されるのと同じデータプライバシー対策を、死亡した参加者にも適用する必要がある。

<適用したアプローチの詳細の記載例>

死亡した参加者の記録を生きている人と同様に扱う限り、記載は不要となるため、N/A と記載

した。

■代替アプローチ (Compatible Approach) : なし

2.11.3 留意点

子供や青年を失うことは、近親者にとって特に苦痛である。したがって、彼らのプライバシーを保護することは最優先事項である。GDPR や英国データ保護法 (UK Data Protection Act) などの一部のプライバシー規制は、死亡した個人に関連する (識別可能な) データには適用されなくなった。ただし、死亡の前に結ばれた守秘義務は死後も続くため、サービスや機関への信頼が損なわれないように、プライバシーと機密性を維持することが重要である。

2.12 その他の項目

1b としては、[2.12.1 Information Collected Under Copyright Licenses](#), [2.12.2 Data Derived from Genomic Data](#), [2.12.3 Seasonality](#) に関連して、特別な対応をした場合、詳細にその内容を記載する。

<u>1b. [MANDATORY] If applicable, please elaborate on the approach applied to the following variables.</u>	
---	--

2.12.1 Information Collected Under Copyright Licenses

(例示)

	詳細
<u>Information Collected Under Copyright Licenses</u> (著作権ライセンスに基づいて収集される情報)	N/A

2.12.1.1 解説

いくつかの臨床試験のデータは、第三者が作成したツール、質問票 (questionnaires)、データ辞書によって収集されているかもしれない。データの有用性を最大化するために、すべての適用されるライセンスの許諾により、多くのデータがそのまま共有されるべきである。我々が普段より使用している有害事象や病歴、薬剤のコーディング辞書は、ライセンス契約でデータの共有について基準を設けていることがある。データ提供者は、関連するデータプライバシー法への遵守を確保しながら、適用されるライセンスや著作権を評価する責任がある。ライセンスで許可されているのであれば、以下のことを推奨する。

- ライセンスの制限内でできるだけ多くの変数/データを共有すること

- 薬剤コードおよび有害事象コードという形で共有されるデータが、データ提供者が所有するライセンスの範囲内であることを確認すること
- ライセンスのバージョンと発行番号をサポート文書に含めること
- データ利用者に関連する可能性のあるライセンス制限をサポート文書に含めること
- 使用した質問票のバージョンと発行番号をサポート文書に含めること

2.12.2 Data Derived from Genomic Data

(例示)

	詳細
<u>Data Derived from Genomic Data</u> (ゲノムデータから得られたデータ)	N/A

2.12.2.1 解説

ある種の遺伝子／ゲノム (genetic/genomic) データは、再特定化リスクが高いため、削除が必要である。

遺伝子型の分類 (例: 野生型と変異型) や、染色体異常 (例: フィラデルフィア染色体陽性 [Ph+] または陰性 [Ph-]) に関する情報は残すことが望ましい。ただし、データプライバシーの観点から、変異解析 (mutation analysis)、一塩基多型 (SNPs)、全ゲノムおよび全エクソームシーケンスに関連する情報は (可能な限り) 削除が推奨される。このデータをデータ利用者の解析のために提供する場合は、遺伝学／オミックス (omics) の専門家に相談して、定性的リスク評価における適切な保護レベルを決定する必要がある。

<詳細の記載例>

遺伝情報は削除した。

2.12.3 Seasonality

(例示)

	詳細
<u>Seasonality</u> (季節性)	N/A

2.12.3.1 解説

季節性情報は、喘息や慢性閉塞性肺疾患などの特定の治療領域における疾患パターンの理解に役立つ重要な情報である。季節の情報が、関係する疾患領域や主解析の要素となっている場合は、半球に正規化した月の情報を含めることで、多くの季節性に関連する解析が可能となる。

半球に正規化した月の情報とは、南半球の「月」を北半球の「月」に相当するものにマッピングすることであり、このような季節情報の導出には、「大陸（または半球）」と元の「月」情報をデータに保持が必要となる。

日付（オフセット、[2.2 Dates](#)）および地理情報（大陸への一般化、[2.10 Geographic Location](#)）の非特定化の推奨アプローチを適用すると季節性を導出できない可能性がある。季節性を導き出すために必要な情報が、すべての標準的な非特定化データセットに存在するとは限らない。そのため、データ中の「月」と「半球」を維持するためには、他の変数の非特定化手法を調整する必要がある。例えば、年齢のバンドをより大きくすることにより、全体としての再特定化リスクを調整することが考えられる。

次の 2 点を検討する必要がある。1) そのデータにとって季節性が重要かどうか、2) 再特定化リスクを高めることなく、元の「月」と「半球」の情報をデータセットに含めるにはどうすればよいか。

季節性を追加のカテゴリ変数（例：冬、春）として含めることは、季節性分析に必要な特異性を提供しない可能性があるため、推奨されない。カテゴリ別の季節性変数は、非特定化された地理的および日付情報の組み合わせと互換性がない可能性がある。

2.13 共有するデータについて

（例示）

PART 2. Data Participants in Dataset	
<u>2a. [MANDATORY] Has any individual participant's data been removed from the dataset due to anonymization requirements? Indicate Yes/No</u>	No
<u>2b. [MANDATORY] If yes in 2a, provide the current number of participants included in the dataset.</u>	
<u>2c. [OPTIONAL] If yes in 2a, provide the rationale for removal.</u>	

2.13.1 解説

Part2 では、共有するデータについて特別な対応があったかの記載が必要になる。2a では、非特定化／匿名化のために個別被験者／参加者のデータを削除したかについて回答する必要がある。2b では、2a で個別被験者／参加者が削除された場合に、現在データに含まれている、被験者／参加者数を求めている。2c では、オプションとのことであるが、個別被験者／参加者のデータを削除する理由について記載することが求められている。

2.14 その他の情報

(例示)

PART 3. Other Information	
3. [OPTIONAL] Indicate if your anonymization report has been provided in the study upload package or if one can be publicly accessed. If available, it is strongly recommended that you share your anonymization report or equivalent document. Please redact any information identifying the vendor before sharing. <i>If the anonymization report covers any of the other elements in this Transparency Checklist, there is no need to duplicate the information.</i>	No
4. [OPTIONAL] Please describe the data format of the provided dataset, e.g., SDTM, ADaM and/or Other. In DataCelerate®, indicate the data format specifications in the Transparency Checklist.	SDTM
5. [OPTIONAL] Please indicate if the study is for an indication where seasonality is a key factor, any adaptations that were required to the methodology (e.g., how variables related to regions and dates have been protected).	No
6. [OPTIONAL] Please provide any other information that is considered helpful to a future research team.	13 participant records have been removed due to national EC requirements

2.14.1 解説

Part3 は、その他の情報として、回答については、すべてオプションとして設定されている。Part3-3 として非特定化／匿名化レポートについての共有について問われている。非特定化／匿名化レポートは共有されたデータを扱う上での非常に重要な情報である。透明性のチェックリストの提供もしくは、それに準拠する情報が提供されるのであれば、さらなる対応は必要ない。Part3-4 では、共有するデータフォーマットを記述する。企業が実施した臨床試験であれば通常、SDTM や ADaM が用いられる。Part3-5 として季節性が重要な要素である場合、地域や日付に関連した変数をどのように保護されているのかについての記載が求められている。Part3-6 として将来のデータ利用者にとって役に立つ情報を記載することができる。

3. おわりに

日本においては、次世代医療基盤法の改訂により、特異な値や希少疾患名等の削除等が不要となる「仮名加工医療情報」の作成・利用が整備され、医療データの2次利用の促進が図られている。匿名加工のデータを扱う場合、どのような匿名化（非特定化）が当てはめられたのかがわかりやすくまとまっていると、データ利用者が扱い易い。TransCelerate が作成した、透明性のチェックリスト⁴に基づくと、その非特定化の内容を端的に示すことができ、データに適応した一連の非特定化をわかりやすく提示し、データの使いやすさにつなげることができる。

本報告書は、そのチェックリストに含まれる項目に対しての非特定化の手法および推奨されるアプローチ、何をデータ利用者に伝えるべきかについて日本語で解説を行った。実際の情報提供は、多くの場合、英語により実施されると想定しているが、日本語による理解を促進することが、日本国内でのデータの利活用の促進になることを考え、本報告書の作成に至った。日本における今後のデータ共有の促進の一助になれば幸いである。

4. 謝辞

貴重な時間を割いて専門的な観点から本報告書をレビュー頂いた TransCelerate Privacy Methodology team に感謝の意を表します。

5. 資料作成者

医薬品評価委員会 データサイエンス部会 2023 年度 継続タスクフォース 2

加藤 智子	サノフィ株式会社（担当副部長）
小山 暢之	第一三共株式会社（TF リーダー）
青木 真	アステラス製薬株式会社（TF リーダー）
田嶋 幸聖	中外製薬株式会社（TF リーダー）
佐土原 和宏	鳥居薬品株式会社
大塚 晶仁	ゼリア新薬工業株式会社
吉岡 彬生	富士フイルム富山化学株式会社（～2024 年 3 月）

Appendix

データの透明性チェックリスト（日本語版）

TRANSPAARENCY CHECKLIST (透明性のチェックリスト)		
Part 1. 適用したプライバシー保護のためのアプローチ		
*TransCelerate の Clinical data sharing: a proposed methodology to enable data privacy while improving secondary use に基づく		
1a. [必須]それぞれの変数タイプに対して適用したアプローチを特定すること	以下から 1 つ選択する* ・推奨アプローチ ・代替アプローチ ・その他	適用したアプローチの詳細 (例：変数が削除されたか変更されたか, 特定の変数が削除／変更された理由の根拠)
<u>Unique Identifiers</u> (固有識別子) ※Identifiers を削除した場合はその根拠を記載する		
<u>Dates</u> (日付)		
<u>Verbatim/Free Text</u> (逐語的なテキスト／フリーテキスト)		
<u>Banding of Variables</u> (変数のバンディング)		
<u>Patient Demographics (sex, race, ethnicity)</u> (患者背景：性別，人種，民族)		
<u>Data With Low Frequencies</u> (件数の少ないデータ)		
<u>Sensitive Information</u> (センシティブ情報)		
<u>Adverse Events & Medical History</u> (有害事象及び病歴) ※MedDRA 情報を削除した場合はその理由を記載する		
<u>Concomitant Medications</u> (併用薬剤) ※データセット中にバージョン情報が無い場合は記載する		
<u>Geographic Location</u> (地理的な位置)		
<u>Records of Participants Who Have Died</u> (死亡した患者の記録)		

参考文献

- ¹ PhUSE De-Identification Working Group: Providing De-Identification Standards to CDISC Data Models
<https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Data+Transparency/De-identification+Standard+for+SDTM+3.2+Version+1.0.xls>
- ² De-identification Guidelines for Structured Data
<https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf>
- ³ TransCelerate BioPharma Inc., Data de-identification and anonymization of individual patient data in clinical studies– A model approach, 2013
<https://www.transceleratebiopharmainc.com/wp-content/uploads/2015/04/CDT-Data-Anonymization-Paper-FINAL.pdf>
- ⁴ Transparency Checklist
<https://www.transceleratebiopharmainc.com/wp-content/uploads/2024/03/Standalone-Transparency-Checklist-August-2023.docx>
- ⁵ CLINICAL DATA SHARING: A PROPOSED METHODOLOGY TO ENABLE DATA PRIVACY WHILE IMPROVING SECONDARY USE AUGUST 2023
<https://www.transceleratebiopharmainc.com/wp-content/uploads/2023/08/FINAL-Privacy-Methodology-Revision-August-25.pdf>