![CRDSA - Clinical Research Data Sharing Alliance]

Standard for Secondary Analysis of Clinical Study Data (v. 1.0)

Draft for Public Comment

Version Date: 27 February 2024

Review Notes:

- This document is a draft for public comment.
- Please make comments using the comment submission form: https://members.crdsalliance.org/document/dl/422. Comments that do not use the form will not be considered.
- Email your completed form to the Clinical Research Data Sharing Alliance (CRDSA) secondary use standards work group at su@members.crdsalliance.org
- The closing date for comments is 31 May 2024.
  .

CRDSA is a multi-stakeholder alliance that serves the clinical data-sharing ecosystem. Our mission is to accelerate the discovery and delivery of life-saving and life-changing therapies to patients by expanding the research value of secondary use data. Broad access to these data has the power to transform the research process, improve trial design and delivery, and benefit the patients who donate their time and their data as part of the clinical development process.

19 **PREFACE**

20 The Clinical Research Data Sharing Alliance has created two draft documents outlining
21 standards for both the sharing and secondary analysis use of clinical study data. Both standards
22 aim to facilitate the responsible sharing and use of anonymized individual patient data (IPD)
23 from clinical studies to enable further research and scientific understanding while protecting
24 patient privacy and innovation.

25 Each document applies to a different audience - broadly, contributors of clinical trial data and
26 researchers using that data. However, it is important to recognize that the standards are
27 complementary and intended to work together to facilitate good science. For example, the
28 standard for secondary analysis is predicated on adherence to the data sharing standard, as the
29 former relies on the proper sharing of data, metadata, and documents outlined in the latter.

30
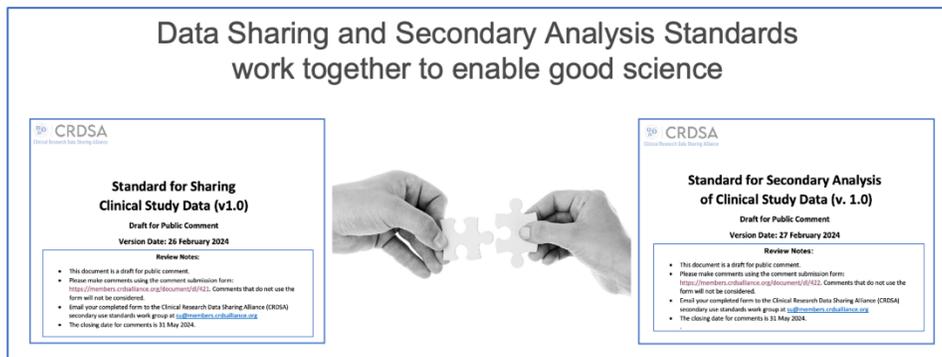
31

32

33 Each standard provides principles, supporting criteria, and best practices for clinical study data
34 sharing policies and procedures. CRDSA considers the principles and supporting criteria to be
35 mandatory. However, it is recognized that adherence is not possible or applicable in some
36 cases. Each document provides a checklist so that implementing organizations can allow for
37 deviations.

38 The standards can be adopted by data sharing platforms, funders, research institutions, and
39 scientific journals. Implementation may vary depending on the organization and its use case(s)
40 and include adopting the standards as written or modifying them to suit organizational needs
41 (provided alterations are clearly outlined).

42 When public comments are incorporated and final standards are published, the two standards
43 establish consistent guidelines for responsibly sharing clinical study data and conducting robust
44 secondary analyses of that data to advance scientific knowledge while safeguarding key
45 considerations.

46

47

CONTENTS

70

71

72 ## 1. INTRODUCTION

73

74 **Why do we need a standard for secondary analysis of clinical study data?**

75 Clinical study investigators, sponsors, and funders are increasingly providing researchers with
76 access to anonymized [1] individual patient data (IPD) from clinical studies. The availability of
77 IPD to test hypotheses and find new insights about diseases and treatments is an important and
78 valuable resource for research. To realize this value, data must be shared and used
79 responsibly. To this end, data contributors must adhere to standards relating to the data,
80 metadata, and documents that are shared [2], and researchers who use this data for secondary
81 analysis must similarly adhere to standards that promote good science and transparent
82 research.

83 There is currently no globally accepted standard for the secondary analysis of IPD from clinical
84 studies that can be applied across different types of secondary research (e.g., reanalyses,
85 meta-analyses, supplemental analyses — see Definitions). Such standards are needed to
86 complement data sharing standards [2] and are predicated on data sharing standards being
87 followed. Taken together, these standards will reduce the risk that inadvertent errors are made
88 that could lead to conclusions and interpretations that are not robust and therefore potentially
89 detrimental to scientific understanding and patient care.

90 This standard for the secondary analysis of clinical study data aims to help researchers conduct
91 robust analyses and objectively interpret the findings. The standard can be required by data
92 providers; cited by researchers as standards they will follow; applied by journal editors and peer
93 reviewers to assess submitted papers; and used by other researchers, prescribers, and patients
94 to evaluate the validity of analyses and conclusions that are drawn.

95

96 **Scope of this standard**

97 This initial version of the standard addresses secondary analyses of interventional clinical trials
98 conducted in patients and non-interventional clinical studies using patient data.

99 Further analyses of datasets generated by other researchers for secondary analyses (e.g.,
100 analysis-ready datasets for meta-analyses) are currently out of scope.

101 **Organization of this standard**

102 The analysis process can be divided into three phases:

103 - **Plan**:
104   o Define the research question and hypothesis.
105   o Identify the studies needed.
106   o Assemble a team with the experience and expertise required.
107   o Determine the appropriate statistical methods to be used.
108   o Obtain required approvals.
109   o Publicly disclose a summary of the planned analysis.
110   o

- **Conduct**:
  - Ensure that the IT system has controls and processes in place to protect the integrity and security of data.
  - Prepare to conduct the analysis by extracting data and transforming the data into a suitable format for analysis.
  - Test the code and reproduce selected analyses from the original study or studies.
  - Perform the statistical analysis according to the statistical analysis plan; justify and document any changes; and implement relevant quality control measures.
  - Interpret the results of the analysis, considering the research question and the study design.
  - Draw objective conclusions based on the findings and study limitations.
- **Report:**
  - Communicate findings through presentations and publications.
  - Share data, documents, and code used for the analysis for transparency and reproducibility.

This standard is organized around these phases in the analysis process. It provides principles that CRDSA considers to be mandatory. Where needed, principles are supplemented with criteria to be followed to meet the principle. Non-mandatory guidance is provided as best practices, which are described with "should" terminology.

These principles, criteria, and best practices are not intended to provide step-by-step instructions; rather, they are intended as a framework for the secondary analysis of clinical study data that can be adapted to different circumstances so that robust analyses can be conducted and interpreted appropriately.

The principles and criteria may not be applicable in every circumstance, and a checklist is provided where any deviations from the principles can be explained.

2.   PRINCIPLES AND PRACTICES

PLAN

2.1 RESEARCH QUESTION
PRINCIPLE: THERE IS TO BE A DOCUMENTED AND WELL-DEFINED RESEARCH QUESTION OR HYPOTHESIS THAT IS TESTABLE BY ANALYSIS OF CLINICAL STUDY DATA

Having a documented and well-defined research question and/or hypothesis helps ensure that the analysis is focused and meaningful. A clear research question is also important when using data mining techniques to identify patterns and relationships in the data so that valid inferences can be made from the results. In addition to having a research question/hypothesis to show the validity of the research, how this question/hypotheses can be tested using clinical study data that may be available is to be shown.

151

Best Practices

- Published literature and information available on data sharing platforms/study registries about ongoing studies and analyses should be reviewed to identify gaps in knowledge or areas where further investigation is needed. The research questions that have been addressed in previous studies and the limitations of these studies should be considered. Areas where there may be conflicting or inconclusive findings, or where further investigation is needed to confirm or extend previous findings, should be identified.
- Key variables should be identified to help clarify and refine the research question. By considering which variables are most relevant to the research question, a more focused and specific research question can be developed.
- For meta-analyses, the value of IPD analysis compared with the traditional aggregate approach should be considered [3].

2.2 STUDIES
PRINCIPLE: CLINICAL STUDIES THAT INCLUDE THE DATA FOR THE ANALYSIS ARE TO BE OBJECTIVELY IDENTIFIED AND ASSESSED USING PREDEFINED CRITERIA

To help ensure the validity of the analysis and to avoid requesting studies that are not needed, the quality and suitability of the studies are to be assessed using predefined criteria. Using inclusion and exclusion criteria helps ensure that the study selection process is guided by the research question or hypothesis, rather than by other factors such as the availability of data. Doing so also helps minimize the risk of bias and increase the validity of results. It is to be confirmed that selected studies include the data required for the analysis and that any transformations of data — e.g., to protect privacy — will not impact the analysis.

The data provider usually provides documentation to help researchers understand the original study design, conduct, and analysis methods (including the protocol, statistical analysis plan, and core clinical study report), as well as the data structures (such as annotated case report forms and dataset specifications; see Sudlow et al [4] for further information). Some data providers may provide some or all of this documentation to support planning, while some provide documentation only once a research proposal has been agreed. It may therefore be necessary to seek and gain access to the documentation before finalizing plans such as the clinical studies to be included in the analysis and the statistical analysis plan (SAP).

Best Practices

- Study eligibility criteria should be used to inform a literature search to identify relevant studies; a comprehensive search strategy should be used to identify relevant clinical studies and minimize potential sources of bias.
- The studies should be screened to determine which ones meet the inclusion criteria.
- Tools such as the Cochrane Risk of Bias Tool [5] can be used to assess the quality of the studies.
- To assess whether studies include the data for the analysis and are suitable, publications, study protocols (if available), and methods data contributors use to

192   anonymize data (which may be available on data sharing platforms) should be checked.
193   Alternatively, if possible, data contributors should be asked to confirm the data required
194   for the analysis are included and that data transformations will not remove or alter data
195   needed for the analysis.

196 • Where data cannot be obtained (e.g., the study predates data contributors' data sharing
197   policies), the possibility that this absence introduces bias, and if so, how the bias can be
198   managed, should be considered.

199

200 **2.3 TEAM**
201 THE RESEARCH TEAM IS TO HAVE THE EXPERIENCE AND EXPERTISE TO CONDUCT
202 THE ANALYSIS

203 A research team is to be assembled with relevant experience, skills, and capabilities, because
204 these qualities are critical for conducting a robust analysis of clinical study data and objectively
205 reporting the results.

206 **CRITERIA**

207 **2.3.1** The team is to include statistical expertise and experience in clinical study data analysis,
208 as shown by statistical qualifications and previous analyses of clinical study data.

209 There is to be expertise and experience of a variety of statistical analysis methods, including
210 techniques to estimate sample sizes and conduct power calculations, so that appropriate
211 methods can be selected based on the research question and data being analyzed. The team is
212 to have experience with clinical study data specifically, because this type of data has unique
213 characteristics and requirements for analysis. Expertise in clinical trial design may be needed,
214 including knowledge of different types of relevant study designs (such as randomized controlled
215 trials, adaptive designs, and crossover trials).

216 **2.3.2** The team is to include the expertise and skill sets needed to navigate clinical study
217 documents and fully understand the relationship of study designs to the intended analysis, as
218 shown by formal training and/or previous research experience.

219 It is important to have expertise on the team so that the original study designs and documents
220 that come with the data are understood. These documents include the clinical study protocol,
221 the annotated case report form (aCRF), dataset specifications, and the statistical analysis plan.
222 Understanding these documents is critical for ensuring that the secondary analysis is conducted
223 in accordance with the original study design and that the data is being used appropriately. For
224 example, the study protocol will provide information on the inclusion and exclusion criteria and
225 the primary and secondary outcomes, and the statistical analysis plan will provide statistical
226 methods that were used in the original study. Further, it is important that the team has the
227 experience to understand how complex study designs (e.g., parallel or crossover studies) affect
228 analyses.

229 **2.3.3** The team is to include expertise in managing the types of datasets being accessed and
230 using the relevant software, as shown by formal training and/or previous research experience.

231 It is important to include expertise in managing the types of datasets that will be used in the
232 analysis and the software that will be used. For example, SAS datasets — data files created

using the SAS statistical software — are widely used in clinical trial research. To work with SAS
datasets, the team is to include some level of expertise in using SAS software.

2.3.4 The team is to include specific expertise relevant for the analysis, as shown by formal
training and/or previous research experience. Depending on the specific research question or
hypothesis being tested and the methods used, additional expertise is to be included in areas
such as the disease area, safety, and artificial intelligence/machine learning.

Safety Expertise

MedDRA (Medical Dictionary for Regulatory Activities) [6] expertise is important for
analyzing safety data in clinical studies. MedDRA is a standardized medical
terminology used to classify and code adverse events and medical conditions related to
drug safety, and it is commonly used in the pharmaceutical industry and by regulatory
agencies.

For many aspects of safety data analysis, such as coding and classifying adverse
events and identifying trends or patterns in safety data, it is critical to have a team
member with MedDRA expertise obtained through formal training. This team member
can use knowledge of MedDRA coding and classification rules to ensure accurate and
consistent coding of adverse events across the clinical study data and can help to
identify potential safety concerns or signals that may require further investigation.

Disease Area Expertise

It is crucial to have a team member with in-depth knowledge of and experience in the
specific disease or medical condition being studied, obtained through previous
research or clinical practice. This can include knowledge of the underlying biology,
pathophysiology, and treatment options, as well as familiarity with relevant clinical
guidelines and standards of care. This expertise is important for designing and
conducting the analysis, interpreting the results, and making informed conclusions
about the implications of the findings for patient care and future research.

Artificial Intelligence (AI)/Machine Learning (ML) Expertise

AI/ML expertise gained through formal training and/or previous research can be
important for secondary analysis of clinical study data if techniques like clustering,
classification, or predictive modeling are planned. AI techniques, such as natural
language processing (NLP) and deep learning, can also be useful for analyzing clinical
study data. For example, NLP can be used to extract and analyze unstructured data
from clinical study reports, while deep learning can be used for image analysis or to
model complex relationships in data.

Best Practices

- Team members who have the needed expertise should be identified. They may include
  colleagues within the organization or external consultants and collaborators.
- Where there is a skills or expertise gap, a plan should be developed to address the gap
  — for example, with training.

272   • Team member roles and responsibilities should be clarified. This process may include
273   assigning specific tasks or responsibilities to each team member.
274

275 **2.4 STATISTICAL ANALYSIS PLAN**
276 PRINCIPLE: A PRESPECIFIED STATISTICAL ANALYSIS PLAN (SAP) IS TO BE IN PLACE

277 Because clinical study IPD offers the potential to analyze data in many different ways, the
278 statistical methods relating to the analysis must be prespecified in detail (see Tierney et al [3] for
279 references). Finalizing and dating a statistical analysis plan (SAP) in advance demonstrates that
280 the plan was developed prior to the analysis being conducted and helps to avoid concerns of
281 further post-hoc data exploration or cherry-picking of results.

282 Developing and documenting the SAP also enables careful consideration of the appropriate
283 statistical methods for the analysis to be appropriate and valid. It also helps identify potential
284 biases that may arise in the analysis, such as selection bias or confounding variables.
285 Identifying these potential sources of bias in advance also enables identification of steps to
286 address them. For further information about analysis considerations for specific types of
287 analysis, see Hollis et al [7].

288 **CRITERIA**

289 **2.4.1** The SAP is to include: [7] [8]:

290   • The questions and hypotheses being addressed
291   • Effect measure of interest (e.g., for inferential studies: odds ratio, risk or rate ratio, risk or
292   rate difference, absolute difference)
293   • The populations and variables to be analyzed, including details of any subjects and data
294   that will be included and excluded
295   • Statistical analysis methods (e.g., logistic regression, Kaplan-Meier curves, log-rank test,
296   multiplicity adjustments)
297   • Any planned adjustment for covariates
298   • Meta-analysis methods, if applicable (e.g., random effects meta-analysis, stratified meta-
299   analysis, meta-regression)
300   • Power to detect a clinically important effect, or the precision of the effect estimate given
301   the sample size available
302   • Any data transformations to be used, and how any missing data or outliers will be
303   handled
304   • Any planned sensitivity analyses to explore the robustness of the results
305   • Any planned investigation of subgroups; for example, by age, disease status, ethnicity,
306   socio-economic status, presence or absence of comorbidities, different types of
307   interventions (e.g., drug dose)

308 **Best Practices**

309   • Study documents should be reviewed or (when possible) discussions held with the data
310   provider to ensure that the data being requested can support the intended analyses.
311   • Biostatisticians and other experts should be consulted as needed to develop the SAP.

312

2.5 REPORTING PLAN

PRINCIPLE: A PRESPECIFIED REPORTING PLAN IS TO BE IN PLACE

A reporting plan is to be in place before conducting the analysis to avoid potential bias in the reporting of results. If the reporting plan is not developed beforehand, there is a risk of selecting only those results that support the hypothesis and neglecting other important findings.

Best Practices

- The target audience for publication of the findings should be determined.
- How the findings will be disseminated (e.g., publication in a peer-reviewed journal, posting on a public website, congress presentation) and any timelines or requirements associated with dissemination should be determined.
- An outline of the content of the publication should be drafted, including key sections such as the introduction, methods, results, discussion, and conclusion. Using standard reporting guidelines such as the CONSORT statement or the PRISMA and PRISMA-IPD statements should be considered [9] [10] [11].
- Specific individuals should be assigned responsibility for preparing the publication and timelines for completion of each section should be established.

2.6 APPROVALS

PRINCIPLE: REQUIRED APPROVALS AND AGREEMENTS ARE TO BE OBTAINED AND DOCUMENTED

All the necessary approvals and agreements governing data access and use are to be obtained before the data is accessed. This is crucial to ensure compliance with relevant laws, regulations, and guidelines, as well as the policies of data providers. Documenting these approvals and agreements provides a comprehensive record that is valuable for future reference.

2.7 TRANSPARENCY OF THE PLAN

PRINCIPLE: A SUMMARY OF THE SAP IS TO BE PUBLICLY DISCLOSED BEFORE CONDUCTING THE ANALYSIS

A summary of the SAP is to be publicly disclosed before the analysis is conducted to ensure transparency, help avoid publication bias, and help demonstrate that significant changes to the plan were not made during the analysis (or help identify where justification and explanation of changes are required in reports). Publicly disclosing the SAP summary can also help to prevent unnecessary duplication of the research.

Best Practices

- The summary of the SAP should include:
  - The research question
  - Analysis approach (e.g., meta-analysis)
  - Study inclusion/exclusion criteria

351   o   Population inclusion/exclusion criteria
352   o   Primary and secondary outcomes
353   o   Statistical analysis methods
354   o   Planned sensitivity analyses
355   o   Analysis complete date
356   o   Contact information
357 •   Whenever possible, the SAP summary should be posted on a study registry.
358

359 CONDUCT

360 **2.8 IT SYSTEM**
361 PRINCIPLE: THE IT SYSTEM, POLICIES, AND PROCEDURES USED FOR DATA HANDLING
362 AND ANALYSIS ARE TO HAVE SUFFICIENT CONTROLS TO PROTECT THE INTEGRITY
363 AND SECURITY OF DATA

364 The IT system used by the research team plays a crucial role in the analysis process and can
365 affect the accuracy, reliability, and security of the data. To prevent unauthorized access,
366 modification, or deletion of data, the IT system is to have controls such as password protection,
367 encryption, and firewalls, and policies and procedures that govern data access, security, and
368 user management are to be in place. There may also be contractual requirements in data
369 sharing agreements that must be followed.

370 **CRITERIA**

371 **2.8.1**   Team members are to be trained on the proper use of the IT system(s) being used to
372   conduct the analysis, including IT security requirements and any contractual
373   requirements in relevant governing Data Sharing or Data Access agreements.

374 **2.8.2**   The research team is to have controls and processes in place to prevent unauthorized
375   access, modification, or deletion of data.

376

377 **2.9 COMPUTER CODE TO RUN THE ANALYSIS**
378 PRINCIPLE: THE CODE IS TO BE TESTED AND THE TESTING IS TO BE DOCUMENTED

379 The code is to be tested to ensure the accuracy, reliability, and reproducibility of the analysis.
380 Testing the code helps identify errors and ensures that the code works as intended. The testing
381 is to be documented, including any assumptions, limitations, and/or dependencies to show that
382 the code works as expected.

383 **Best Practices**

384 •   The code developed for data preparation and analysis should be clear and concise. It
385   should also adhere to referenceable and commonly accepted coding practices.
386 •   Version control software should be used to track changes to the code over time and
387   ensure that any updates are thoroughly tested and documented.
388 •   The code should be reviewed by at least one other team member to ensure its quality
389   and reliability.

390

2.10 DATA MANAGEMENT

PRINCIPLE: PREDETERMINED METHODS FOR DATA MANAGEMENT AND ANY READADJUDICATION ARE TO BE FOLLOWED AND ANY DEVIATIONS ARE TO BE DOCUMENTED

Where the secondary analysis involves extracting data from the datasets provided and/or creating an analysis-ready dataset, predetermined methods are to be followed to ensure the validity of the analysis and to minimize bias due to subjective decisions during data preparation. Any deviations from these methods in data preparation are to be justified and documented. Readjudicating adverse events or efficacy outcomes requires careful consideration because readjudication in a secondary analysis can be challenging and can negatively affect the validity of the analysis.

CRITERIA FOR READJUDICATION

2.10.1 There is to be a documented justification for any analysis involving readjudication of a source study.

2.10.2 Multiple independent adjudicators are to be involved in the readjudication process. This helps to reduce the impact of individual biases and increases the reliability of the results.

2.10.3 The analyst or reviewers are to be blinded to the treatment group assignments and other relevant information to reduce bias.

2.10.4 MedDRA is to be used for the readjudication of adverse events. MedDRA is a standardized medical terminology developed by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). It provides a common language for the classification and coding of adverse events. The use of MedDRA helps to ensure that the process is consistent and reduces the potential for bias.

2.10.5 For efficacy outcomes, the same standards as used in the original analysis or other standardized and referenceable criteria (e.g., Hamilton Depression Rating Scale [HAM-D] for depression) are to be used. Where different standards or scales are used, this is to be documented and reported.

Best Practices for Data Management

- The accuracy and completeness of the data should be verified by running basic descriptive statistics and comparing them to the original study report or publication.
- Data management methods that should be considered include imputing missing values, recoding variables, creating derived variables, and removing outliers.
- Where different studies have collected and defined data in different ways, methods for recoding data items into a common format should be developed and definitions harmonized [3].

429   •   Quality control measures should be established to ensure the accuracy and
430   completeness of the extracted data and ensure that any data transformations are
431   accurate and complete.
432   •   The analysis-ready dataset should be validated to ensure that it is accurate, complete,
433   and fit for purpose (for example, checking the data against the source dataset and
434   looking for outliers and missing data).
435   •   The impact of the preparation methods on the results of the analysis should be
436   considered and sensitivity analyses should be conducted to assess the robustness of
437   the findings.

438

439 **Best Practices for Readjudication**

440   •   The methods used in the original study should be reviewed to ensure that readjudication
441   methods are using the appropriate criteria and standards.
442   •   Relevant information used by the original researcher, such as original ECG recordings,
443   should be sought in adjudicating adverse events or outcomes (although in some cases
444   this information may not be available).
445   •   The use of external experts to help with the readjudication process should be considered
446   to ensure that the process is unbiased and reliable.
447   •   The results of the readjudication should be validated using sensitivity analyses and other
448   methods.

449

450 **2.11 STUDY AND DATA UNDERSTANDING**
451 PRINCIPLE: TO DEMONSTRATE UNDERSTANDING OF THE STUDIES AND STUDY DATA,
452 SELECTED ANALYSES IN THE ORIGINAL STUDIES ARE TO BE REPRODUCED AND ANY
453 DIFFERENCES ARE TO BE EXPLAINED AND DOCUMENTED

454 Selected analyses in the original studies are to be reproduced using the same methods and
455 assumptions as used in the original research, to provide confidence that the studies and the
456 data from the studies are well understood and that researchers can navigate the datasets
457 correctly. Any major differences from the original study (or studies) and the reasons for those
458 differences are to be documented.

459 **Best Practices**

460   •   Where there are major differences that cannot be explained, the original researchers
461   should be contacted (if possible) to discuss the discrepancies.
462   •   If the original analysis used a method or library that the researchers conducting the
463   secondary analysis are not familiar with, they should learn the method or library, seek
464   assistance (e.g., from the data contributor), or use alternative methods.
465   •   If a method or library used in the original research has been updated since the time the
466   original study was conducted, the researchers conducting the secondary analysis should
467   seek to use the original methods and libraries. If this is not possible, it is important for
468   the researcher to document any changes made and to be transparent about the impact
469   of these changes on the results.

470 • Study documentation such as the original SAP, dataset specifications and code (when
471 available) should be reviewed to understand the original analysis approach.
472
473

474 **2.12 ANALYSIS**
475 PRINCIPLE: THE SAP IS TO BE FOLLOWED AND ANY DEVIATIONS ARE TO BE JUSTIFIED
476 AND DOCUMENTED

477 The SAP is to be followed to ensure the reliability and validity of the analysis and enable others
478 to reproduce the analysis. Any deviations are to be justified and documented; unplanned
479 analyses can still play an important role in explaining or adding to the results, but such
480 exploratory analyses are to be justified and clearly reported as such [3].

481

482 **2.13 QUALITY CONTROL**
483 PRINCIPLE: QUALITY CONTROL MEASURES ARE TO BE IMPLEMENTED

484 Quality control measures are to be implemented to ensure that the results of the analysis are
485 accurate and valid. These measures can include various techniques, such as a review of
486 statistical output and checks for outliers or inconsistencies. By implementing these measures,
487 potential errors or issues with the analysis can be identified and corrected.

488

489 **CRITERIA**

490 **2.13.1** Peer review is to be carried out by having a second statistician or researcher review
491 the analysis to ensure that methods and results are appropriate and accurate.

492 **Best Practices**

493 • Any data entry should be double-checked by having a second person independently
494 check the accuracy of data entry to ensure that errors are identified and corrected.
495 • The analysis should be repeated on a subset of the data to ensure that results are
496 consistent and robust.
497 • Outliers or extreme values in the data should be identified and investigated to ensure
498 that they are not driving the results.
499

500 **2.14 INTERPRETATION**
501 PRINCIPLE: THE RESULTS ARE TO BE INTERPRETED IN A SCIENTIFIC AND OBJECTIVE
502 MANNER

503 Interpretation of study results is to be based on scientific evidence and rigorous analysis, rather
504 than personal biases or subjective opinions. It is not generally possible to limit bias and control
505 multiplicity to the extent that it is possible in the original study design, and this is to be
506 acknowledged and taken into account when interpreting the results of analyses [7]. Secondary

507 analyses of study data are essentially post hoc analyses, and therefore of exploratory rather
508 than confirmatory value [7].

509 **CRITERIA**

510 **2.14.1** The objectives and research question for the secondary analysis are to be reviewed to
511 ensure that the interpretation of the findings is consistent with the research question
512 and objectives.

513 **2.14.2** The design and methodology used are to be carefully reviewed to ensure that the
514 interpretation of the results is appropriate given the strengths and limitations of the
515 analysis. Doing so may involve considering potential sources of bias or confounding
516 factors that may have influenced the results.

517 **2.14.3** Alternative explanations for the results are to be considered, such as factors that may
518 have influenced the findings.

519 **2.14.4** Clinical relevance in the context of any limitations, as well as statistical significance, is
520 to be considered.

521

522 REPORT

523 **2.15 PUBLICATION**
524 PRINCIPLE: THE RESULTS ARE TO BE PUBLICLY DISCLOSED FOLLOWING THE
525 REPORTING PLAN

526 The predetermined reporting plan (see 2.5) is to be followed to reduce the risk of selective
527 reporting or reporting bias. An explanation of the results, their significance, and the conclusions
528 that can be drawn from them is to be included.

529 **CRITERIA**

530 **2.15.1** It is to be stated that the results were generated after the completion of the original
531 clinical studies, and references to publications containing the original results are to be
532 included.

533 **2.15.2** The statistical methods used to analyze the data and any assumptions or limitations of
534 these methods are to be described.

535 **2.15.3** Any deviations from the SAP are to be clearly explained and a rationale for the
536 changes is to be provided.

537 **2.15.4** Any potential biases or confounding factors that may have influenced the results are to
538 be discussed.

539 **2.15.5** A thorough explanation of any unexpected or conflicting results and possible reasons
540 for these findings is to be provided.

541 **2.15.6** Clinical relevance in the context of any limitations, as well as statistical significance, is
542 to be discussed.

543 **2.15.7** Publication-specific guidelines for reporting the results of the analysis are to be
544 followed, such as the CONSORT statement for clinical trials or Preferred Reporting
545 Items for Systematic Reviews and Meta-Analyses/IPD (PRISMA/PRISMA-IPD).

546 **2.15.8** The comparability of the results to the population of interest and any limitations of the
547 analysis or population are to be discussed.

548 **2.15.9** The statistical analysis plan and reporting plan, as well as the analysis replicating the
549 original analysis, is to be included in supplementary information with the publication
550 (with any personal information redacted if necessary).

551 **Best Practices**

552 • The results should be presented with tables, figures, and statistical analyses to support
553 the conclusions.
554 • Plain language should be used and technical jargon avoided to ensure that the research
555 is accessible to a wide audience.
556 • Feedback from colleagues and peer reviewers should be sought to ensure that the
557 publication or presentation is accurate, complete, and relevant.
558

559 **2.16 RESULTS TRANSPARENCY**
560 PRINCIPLE: DATA, DOCUMENTS, AND CODE USED FOR THE ANALYSIS ARE TO BE
561 SHARED OR MADE AVAILABLE ON REQUEST

562 Data, documents, and code used for the analysis are to be shared or made available on
563 request to allow for transparency in the research process, thus enabling other researchers to
564 examine the data and methods used in the analysis. This promotes the reproducibility of the
565 findings, which is crucial for scientific advancement and validation of research.

566 **CRITERIA**

567 **2.16.1** The code is to be documented and commented to enable others to understand the
568 approach. Any necessary libraries and dependencies are to be included in the code.

569 **2.16.2** Documentation used in the analysis (e.g., SAP) is to be prepared for sharing by
570 ensuring that appropriate steps are taken to protect the privacy of any individuals (e.g.,
571 patients, investigators, etc.).

572 **Best Practices**

573 • The data should be organized in a clear and consistent manner, with appropriate labels
574 and annotations. Doing so will make it easier for others to understand and use the data.
575 • A suitable repository should be chosen to share the data and code. Popular options
576 include Clinical Study Data Request (CSDR) and Vivli.
577 • If the clinical study dataset accessed and used in the analysis is only available in a
578 secure repository, it may not be possible to share the data directly with other
579 researchers. In this case, a link to the secure repository should be provided where the
580 data can be accessed by others who want to replicate the analysis.

581   •   If the code uses libraries that are not freely available, instructions on how to obtain them
582   or alternative methods for running the code without the libraries should be provided.

583

584

585

586 ## 3.0 CHECKLIST

587

588 Any deviations from the principles can be explained in the checklist below.

589

590 **NOTE:** For a principle to be deemed satisfied, there must be documented and verifiable support
591 for compliance.

592

593 **PLAN**

		Yes/No/NA	NA Explanation
RESEARCH QUESTION	Is there a documented and well-defined research question and/or hypothesis that is testable using clinical study data?		
STUDIES	Have studies that include the required data been objectively identified and assessed using predefined criteria?		
TEAM	Does the team have statistical expertise and experience in clinical study data analysis, as shown by statistical qualifications and previous analyses of clinical study data?		
	Does the team include the expertise and skill sets needed to navigate clinical study documents and fully understand the relationship of study designs to the intended analysis, as shown by formal training and/or previous research?		
	Does the team include expertise to manage the types of datasets being accessed and use the relevant software, as shown by formal training and/or previous research?		
	Does the team include specific expertise relevant for the analysis (e.g., MedDRA expertise for analysis of safety), as evidenced by formal training and/or previous research?		
SAP	Is there a prespecified statistical analysis plan (SAP) dated before the analysis was conducted?		
	Does the SAP include all of the following? - The questions and hypotheses being addressed - Effect measure of interest (e.g., for inferential studies: odds ratio, risk or rate ratio, risk or rate difference, absolute difference) - The populations and variables to be analyzed, including details of		

	any subjects and data that will be included and excluded - Statistical analysis methods (e.g., logistic regression, Kaplan-Meier curves, log-rank test, multiplicity adjustments) - Any planned adjustments for covariates - Meta-analysis methods, if applicable (e.g., random effects meta-analysis, stratified meta-analysis, meta-regression) - Power to detect a clinically important effect, or the precision of the effect estimate given the sample size available - Any data transformations to be used, and how any missing data or outliers will be handled - Any planned sensitivity analyses to explore the robustness of the results - Any planned investigation of subgroups (e.g., by age, disease status, ethnicity, socioeconomic status, presence or absence of co-morbidities, different types of intervention [e.g., drug dose])		
REPORTING PLAN	Is there a reporting plan dated before the analysis was conducted?		
APPROVALS	Are all required approvals and agreements obtained and documented?		
TRANSPARENCY OF THE PLAN	Has a summary of the SAP been publicly disclosed?		

594

595 **CONDUCT**

		Yes/No/NA	NA Explanation
IT SYSTEM	Does the IT system for the analysis have controls (system, policies, and procedures) in place to protect the integrity and security of data?		
	Have team members been trained on the proper use of the IT system, including IT system security and other contractual requirements in relevant governing Data Sharing or Data Access agreements?		

	Does the research team must have controls and processes in place to prevent unauthorized access, modification, or deletion of data?		
COMPUTER CODE TO RUN THE ANALYSIS	Has the code been tested and has this testing been documented?		
DATA MANAGEMENT	Have predefined methods for data management been developed and followed, and have any deviations been documented?		
	If the analysis involves readjudication: - Is there a documented justification for readjudication? - Are multiple independent adjudications involved in the process? - Are the analyst and reviewers blinded to treatment group assignments and other relevant information to reduce bias? - Is MedDRA used for the readjudication of adverse events?		
STUDY AND DATA UNDERSTANDING	Has it been possible to reproduce selected analyses in the original studies and have any major differences been explained and documented?		
ANALYSIS	Has the secondary analysis followed the SAP and have any deviations been justified and documented?		
QUALITY CONTROL	Have measures been implemented to assure the quality of the analysis?		
	Has a second statistician or researcher reviewed the analysis to ensure that methods and results are appropriate and accurate?		
	Is the interpretation of study findings consistent with the original research question and study objectives?		
	Is the interpretation of results appropriate given the strengths and limitations of the study design?		
	Have alternative explanations for the study results been considered?		
	Has the clinical relevance in the context of any limitations, as well as statistical significance, been considered?		

596

597

598

599   **REPORT**

		Yes/No/NA	NA Explanation
PUBLICATION	Are the findings being publicly disclosed following the reporting plan?		
TRANSPARENCY OF RESULTS	Are the data, documents, and code for the analysis being shared or made available on request?		
	Has the code been documented and commented to enable others to understand the approach?		
	Has documentation used in the analysis (e.g., SAP) been prepared for sharing by ensuring that appropriate steps are taken to protect the privacy of any individuals (e.g., patients, investigators, etc.)?		

600

601

602   ## 4. DEFINITIONS

603

Analysis-ready datasets	Datasets (databases) where all the data derivations are generated. It is the "analysis ready" datasets that are then used as the input to any statistical analyses and data summaries.
Data provider (based on Sudlow et al [4])	Any organization that conducts clinical trials and has the rights to share the database of the electronic IPD for that study. This could include pharmaceutical companies, biotech companies, medical device companies, academic groups, and medical charities.
Meta-analysis (Hollis et al [7])	Further investigation of the efficacy and safety of one or more interventions using individual IPD from several clinical studies, such as: - Meta-analysis to learn more about an intervention by pooling several studies including the same comparison - Network meta-analysis to learn more about the relative effect of various interventions by making indirect comparisons across several studies with different comparators
Multiplicity	In clinical study research, multiplicity refers to the potential for multiple statistical comparisons to be made between study

	groups or different endpoints, which increases the risk of false-positive findings or Type I error.
Raw datasets	Datasets (databases) of the data as it was recorded on the case report form. These are usually split into a number of raw datasets reflecting the different types of data that have been collected, such as adverse events, laboratory assessments, disease-specific measurements.
Reanalysis (Hollis et al [7])	Further investigation of the efficacy and safety of the intervention, such as: - Using a new measure of benefit or risk that can be derived from the available data - Exploring the impact of analysis assumptions made, such as the handling of missing data - Verification of the results in the original study report or publication
Researcher (Sudlow et al [4])	Any individual or group who seeks access to IPD to address a specific research question. Researchers are external to the Data Holder's organization and could be academics employed in the public or private sector.
Statistical analysis plan (SAP)	A statistical analysis plan is a document that outlines the statistical methods to be used for a particular research study or clinical trial.
Supplemental analysis (Hollis et al [7])	Research question that is not directly assessing the randomized intervention, such as: - Exploring prognostic factors and characterizing disease evolution over time - Evaluating new statistical methods - Understanding relationships between endpoints, gaining information to inform the design of a future study

604

605
606

5. REFERENCES

[1] L. Arbuckle, "A new standard for anonymization," International Association of Privacy Professionals, 14 March 2023. [Online]. Available: https://iapp.org/news/a/a-new-standard-for-anonymization/.

[2] CRDSA Technology and Innovation Work Group, "CRDSA Data Contribution Standard," 2024. [Online]. Available: https://crdsalliance.org/resources/#tiwg.

[3] J.F. Tierney, L.A. Stewart, M. Clarke; on behalf of the Cochrane Individual Participant Data Meta-analysis Methods Group, "Cochrane Handbook Chapter 26: Individual participant data," [Online]. Available: https://training.cochrane.org/handbook/current/chapter-26.

[4] R. Sudlow, J. Branson, T. Friede, D. Morgan and C. Whately-Smith, "EFSPI/PSI working group on data sharing: accessing and working with pharmaceutical clinical trial patient level datasets — a primer for academic researchers," *BMC Medical Research Methodology,* vol. 16, no. 73, pp. DOI 10.1186/s12874-016-0171-x, 8 July 2016.

[5] J. P. T. Higgins, D. G. Altman, P. C. Gøtzsche, P. Jüni, D. Moher, A. D. Oxman, J. Savović, K. F. Schulz, L. Weeks, J. A. C. Sterne, Cochrane Bias Methods Group, Cochrane Statistical Methods Group, "The Cochrane Collaboration's tool for assessing risk of bias in randomised trials," *BMJ,* vol. 343, p. d5928 doi: https://doi.org/10.1136/bmj.d5928, 18 October 2011.

[6] MedDRA, [Online]. Available: www.meddra.org.

[7] S. Hollis, C. Fletcher, F. Lynn, H.-J. Urban, J. Branson, H.-U. Burger, C. Tudur Smith, M. R. Sydes and C. Gerlinger, "Best practice for analysis of shared clinical trial data," *BMC Medical Research Methodology,* vol. 16, no. 76, pp. DOI 10.1186/s12874-016-0170-y, 8 July 2016.

[8] CSDR, [Online]. Available: ClinicalStudyDataRequest.com.

[9] D. Moher, S. Hopewell, K. F. Schulz, V. Montori, P. C. Gøtzsche, P. Devereaux, D. Elbourne, M. Egger and D. G. Altman, "CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials," *BMJ,* vol. 340, p. c869 doi: https://doi.org/10.1136/bmj.c869, 24 March 2010.

[10] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman; PRISMA Group, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *PLoS Med,* vol. 6, no. 7, p. e1000097. doi: 10.1371/journal.pmed.1000097., 21 July 2009.

[11] L.A. Stewart, M. Clarke, M. Rovers, R.D. Riley, M. Simmonds , G. Stewart, J. F. Tierney; PRISMA-IPD Development Group., "Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data: The PRISMA-IPD Statement," *JAMA,* vol. 313, no. 16, pp. 1657-1665. doi:10.1001/jama.2015.3656, 28 April 2015.

[12] T. P. Debray, K.G. Moons, G. van Valkenhoef, O. Efthimiou, N. Hummel, R. H. Groenwold, J.B. Reitsma; GetReal Methods Review Group., "Get real in individual participant data (IPD) meta-analysis: a review of the methodology.," *Res Synth Methods,* vol. 6, no. 4, pp. 293-309. doi: 10.1002/jrsm.1160. Epub 2015 Aug 19. PMID: 26287812; PMCID: PMC5042043., December 2015.

[13] International Conference on Harmonisation E9 Expert Working Group, "ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials.," *Stat Med,* vol. 18, no. 15, pp. 1905-42, 15 August 1999.

609

610