

Biostatistics, data management, and medical writing:

A multidisciplinary approach to the development of the CTD integrated summaries

Shiri Diskin,¹ Tanya du Plessis,¹
Ofra Barnett-Griness²
Bioforum Group

¹ Bioforum Group, Ness Ziona, Israel,
and Bloemfontein, South Africa

² Statistical consultant to Bioforum Group



Correspondence to:

Shiri Diskin, PhD
VP, Medical Writing
Bioforum Group
3 Golda Meir St
Ness Ziona, Israel
Shiri.Diskin@Bioforumgroup.com

Abstract

Analyses of integrated databases of efficacy and safety are a Food and Drug Administration (FDA) requirement. They are very useful in evaluating the safety and efficacy data gathered in multiple clinical studies. However, their utility is dependent upon the quality of the

studies and the data gathering methods, which affect the quality of the data. It also depends on a scientifically sound strategy for pooling and analysis of the data, and finally, on the adequate reporting of results. Early involvement of professionals from the data management and biostatistics fields can facilitate the development of valuable integrated summary of safety (ISS) and integrated summary of efficacy (ISE) through implementation of study design and data management strategies that are geared toward pooling of data from multiple studies. Medical writers should also join the process early to acquire the knowledge and understanding required for reporting the data in an accurate and meaningful way.

Introduction

A clinical development programme of a pharmaceutical product is designed to collect information that is pertinent to the evaluation of its benefit-to-risk ratio. At later stages of the programme, when preparing for submission, the task at hand is to understand the picture that arises from all available data. In this article, we focus on the rigorous analysis and presentation of integrated clinical safety and efficacy data gathered from prospective, interventional, sponsor-initiated studies from the perspectives of the data management (DM), programming, biostatistics, and medical writing (MW) functions.

Clinical safety and efficacy data from a full programme can be presented in two main ways:

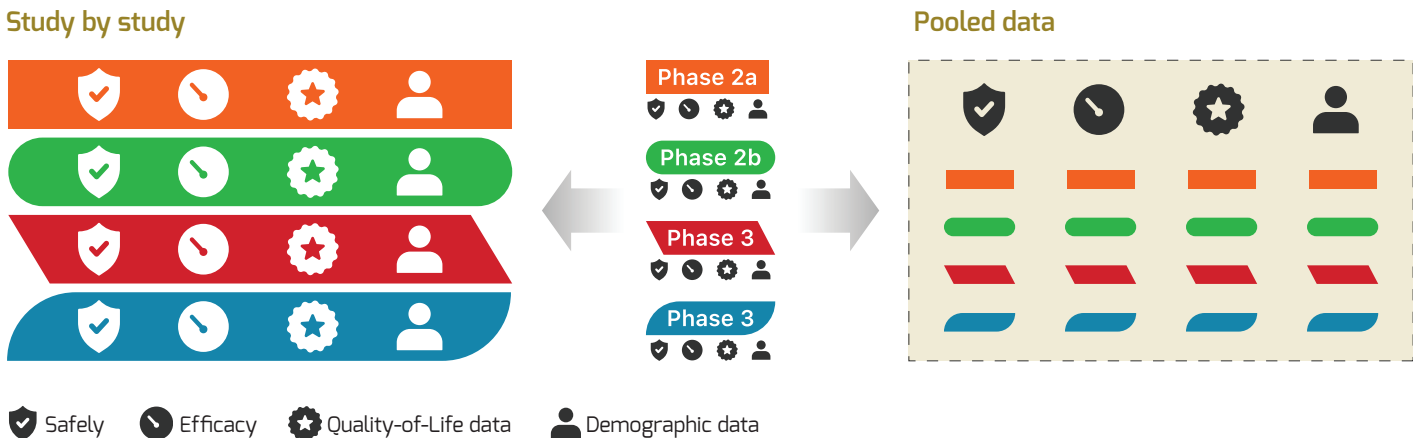


Figure 1. A schematic illustration of data analyses in summary documents

Data can be presented as separate studies side by side, allowing the comparison of specific types of data between studies. Data can also be presented in a pooled manner, referring to the group of studies as a single dataset, providing the benefit of a large sample size.

by study, or by data module. The latter can be done following pooling of data from a few studies into an integrated database that is used as one large study (see Figure 1). This approach can provide valuable tools for understanding the “big picture” as well as addressing specific clinical issues with the product, examples are provided later in this article.

Regulatory requirements

The International Council for Harmonisation

(ICH) guideline M4E (R2)¹ *Common technical document for the registration of pharmaceuticals for human use – Efficacy*, refers to three levels of detail of clinical efficacy and safety data presentation (See Figure 2):

- The clinical overview that includes the overviews of efficacy and safety (Modules 2.5.4 and 2.5.5). These are required across ICH countries and are intended as concise and critical analyses of clinical data pertinent to the evaluation of efficacy and safety of the

medicinal product in the intended population, focusing on interpretation and discussion.

- The summary of clinical efficacy (SCE) and summary of clinical safety (SCS) (Modules 2.7.3 and 2.7.4). They are required across ICH countries and provide detailed factual summarisation of all data relevant to efficacy and safety in the intended patient population and may include a summary of the results from integrated databases.

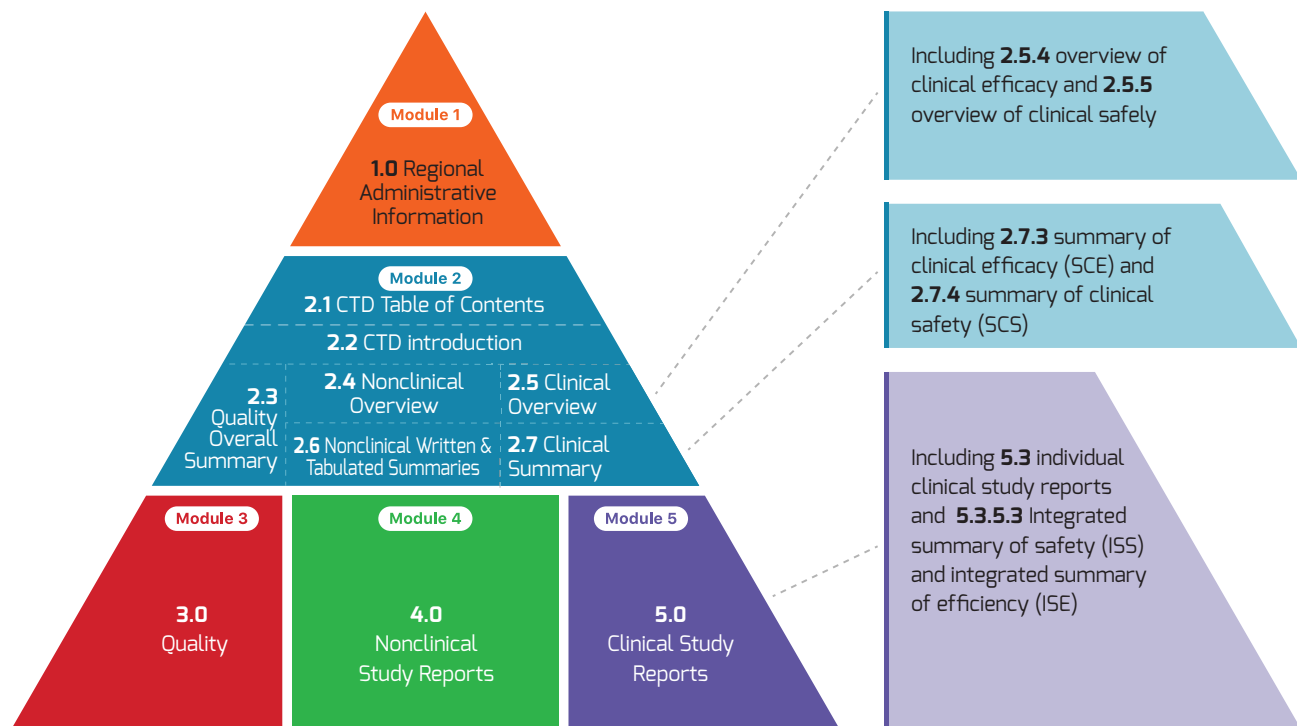


Figure 2. A schematic representation of the hierarchical structure of the Electronic Common Technical Document

- The detailed reports of individual clinical studies written in accordance with the ICH E3 guideline² (Module 5.3.5). In addition, FDA requires reports of analyses of data from more than one study, the integrated summary of efficacy (ISE) and the integrated summary of safety (ISS) (Module 5.3.5.3) that provide a detailed description and presentation of the results obtained from integrated safety and efficacy databases.³

The data management perspective

There are a few methodologies for the gathering of data in clinical studies. A solid data strategy plan across all studies involved with as much consistency as possible regarding data capture and cleaning allows for a streamlined analysis of data, a reduced need for retrospective reviews and processing.

When initiating a Phase 2a study, many sponsors are uncertain whether their product would be eligible for marketing approval submissions, and at times, are not yet adequately funded. For these reasons, instead of forming a long-term programme-wide data strategy, they opt for “minimum essential” data capture and management plans by using cheap, less reliable, and inconsistent methods for data capture such as paper case report forms or excel sheets.

The FDA requires adherence to specific database design standards for database submissions⁴ (such as Clinical Data Interchange Standards Consortium [CDISC] Study Data Tabulation Model [SDTM]). When planning Phase 2b studies, companies are more likely to choose electronic data capture (EDC) systems, however, due to a lack of awareness or lack of resources, FDA-required standards are not always taken into account.

For the purpose of pooling and cross-study analyses, data have to be available in a consistent format (for example, all adverse event [AE] data should be coded in the same version of the MedDRA dictionary⁵). Phase 3 studies are usually designed in collaboration with DM experts with submission in mind. An adequate EDC system with CDISC-compliant data capture is likely to be chosen. Moreover, when a Phase 3 programme includes more than one study, the structure of the studies, the duration of treatment, visit schedules, and data collection of safety and efficacy variables are all planned in a consistent manner allowing for standardisation

of data capturing and cleaning. Legacy data should be processed to achieve the same standard. Legacy data that were captured on paper are manually inputted into an EDC system retrospectively with minimal resources for data cleaning and resolving queries. Electronic legacy data that were captured in a format that is not CDISC-compliant must be converted. Taken together, the retrospective processing incur additional costs and time that can be minimised if a data strategy for integration and submission is implemented early on in development.

The biostatistics perspective

It is essential to apply statistical considerations when forming the data integration strategy. The analytical strategy should include the following elements:

- The objectives for integration
- The regulatory guidelines
- Which studies to pool
- The outcomes and time points
- The statistical methods

Regulatory guidelines and integration objectives

The following relevant guidelines can be used in establishing a strategy for the integrated summaries or pooling of data across studies:

- “Summarising the Clinical Database” in the ICH E9⁶
- “Meta-Analyses of Randomised Controlled Clinical Trials to Evaluate the Safety of Human Drugs or Biological Products (2018)”, draft FDA guidance⁷
- “Integrated Summary of Effectiveness” (2015), an FDA guidance.⁸

The ISE aims to provide insights beyond those observed in individual clinical trials.⁸ Individual studies are presented to demonstrate the claimed effects, and if applicable and appropriate, a statistical combination (pooling) of results may then be considered.

Generally, studies are pooled for two main purposes: to achieve a greater power and increase

precision and to assess the drug effect in demographic or clinical subpopulations when there are too few subjects in each individual study to support meaningful conclusions.⁸ However, many of the pooled analyses are exploratory in nature and are designed to probe the data for trends across studies, e.g., in disease-specific subgroups.⁸

The draft FDA guidance⁷ primarily focuses on meta-analyses with predefined hypotheses designed to confirm a suspected safety risk associated with a drug rather than on exploratory meta-analyses. As such, it is a source for detailed and scientifically rigorous discussions on important considerations when pooling study data for regulatory purposes. It is a valuable resource even when there is no formal predefined safety hypothesis to check.

Which data to pool

When considering which data to pool, the first step is to list all clinical studies with their critical design characteristics, and their respective roles in the development programme.

Importantly, any study included in pooled data will be evaluated and discussed thoroughly as an individual study.

The aim of data integration is to provide a valid description of expected safety and efficacy in the target population, and its usefulness is dependent on individual trials with high-quality data. Thus, the first principle is **Quality over Quantity**.

The application of this principle to efficacy data integration, involves limiting the candidate list to individual studies that are considered “adequate and well-controlled” and serve as the basis for establishing efficacy claims.⁶ When applied to Safety integration, findings from a limited set of trials, selected with careful attention to trial and data quality and the intended use of the product, can yield a more informative view of product safety than a broader set of trials

that includes trials of poor quality.⁷

The candidate studies are likely non-homogenous, and thus, it should be assessed

A single integrated database can be planned if all candidate studies are deemed similar enough for integration. Otherwise, multiple data pools can be proposed for groups of similar studies. For example, pooling studies using a certain active control or studies with long-term follow-up.

At later stages of the programme, when preparing for submission, the task at hand is to understand the picture that arises from all available data.



whether they can be meaningfully pooled with respect to important elements such as:

- the population, including demographic or clinical factors (such as age range and disease severity);
- the exact indication being evaluated (per study objectives and inclusion/exclusion criteria);
- relevant trial design factors, such as the control group (if any), treatment and follow-up duration, allocation ratio, and collected endpoints.

The degree of variability tolerated with respect to each factor may differ between the integrated datasets for efficacy and safety. For instance, the analysis of AEs may be appropriate in an integrated dataset that includes subjects with different disease severities, but it may be harder to draw efficacy conclusions from such a varied dataset.

A single integrated database can be planned if all candidate studies are deemed similar enough for integration. Otherwise, multiple data pools can be proposed for groups of similar studies. For example, pooling studies using a certain active control or studies with long-term follow-up.

Healthy volunteer studies should not be included in integrated datasets for either efficacy

or safety because they assess a population distinctly different from the target patient population and are commonly much shorter. Those studies will be analysed separately and may be pooled as a distinct safety cohort.

Statistical methods can be applied to adjust for important differences between patients and studies that will form the integrated database (see below).

Endpoints and time points

When combining efficacy data, the focus should be on the prespecified primary endpoints (defined for the confirmatory Phase 3 studies). However, when important outcomes are common to all studies (even when the primary endpoints differ), analyses of such outcomes can provide an important assessment of consistency.⁸ An example provided in the ISE guidance⁸ is a series of studies in which an important variable was assessed at multiple time points, and an analysis of the results obtained at a common time point can be shown, even when the time point for the primary analysis differed among studies.

Unlike efficacy endpoints, safety endpoints are generally standard (usually AEs, laboratory data and vital signs), thus, pooling of all safety outcomes can be expected. Still, the collection timepoints of these measures may not be uniform

and studies may vary in treatment duration and follow-up. The use of common time points shared by all the studies and the statistical handling of differences in follow-up duration may facilitate integration of these measures across studies.

The statistical analysis methods

Efficacy analyses are mainly comprised of a pooled treatment effect estimation, by comparing treatment and control groups using appropriate statistical models. A regression model can be applied to the integrated database, the same as would be used for each endpoint in the individual confirmatory study, and further stratified for the study factor. The pooled treatment effects are accompanied by assessment of the homogeneity of the treatment effect across studies, by contrasting the study-specific treatment effects and testing study-by-treatment interaction. A forest plot presenting individual study results and pooled effect is often provided. Thus, the approach taken to derive the pooled effect is to treat the integrated database as a single large trial (see Figure 1), while accounting for study variability (and maintaining randomisation within each study) through stratification.

The statistical analyses for the common safety outcomes are descriptive in nature yet require



special attention. AEs for example, are generally analysed in individual clinical study reports using crude percentages (number of patients with events divided by number of treated patients).

However, naive pooling of the safety database by treating it as one large study and calculating crude percentages for each treatment group, may result in bias when trials employ different randomisation allocations.⁷ An illustration of this bias, termed Simpson's paradox is provided in the FDA draft guidance:⁷ in the example, the risk for a specific safety event was identical for the treatment and control groups in each of the trials. Thus, analysis of the individual studies would result in the safety event not being a concern for the product. However, the risk was not the same across the trials: in one of the trials, in both the treatment and the control groups, the risk was higher than in others. This study employed a 4:1 allocation ratio (treatment: control), thus simple pooling of trials enriched the treatment group with high-risk patients leading to a biased overall result of increased risk. Statistical solutions for this bias (employing

stratification by study and weighting approach) are provided in Chuang-Stein and Beltangady 2011.⁹ Contrasting the results from a pooled analysis with that of each specific study can help understand whether a bias has occurred – this substantiates the importance of presenting individual study results as well.

We have assumed throughout that subject-level data is available, as opposed to only trial-level summary measures, for which meta-analytic statistical methods are available.

The medical writers' perspective

MWs are responsible for taking the data and all the background materials and turning them into a coherent narrative that conveys the current

knowledge about the efficacy and safety of the product. Therefore, the MW should get involved early in the process of planning submission documents. The MW, whether in-house or outsourced, should be familiar with the clinical studies of the programme and understand their distinct characteristics as well as the

The presentation of both efficacy and safety in the documents should focus on the sought indication in the target patient population, taking care to include the types of information that the regulator is specifically interested in.

shared characteristics that make them eligible for pooling. It is also highly important to understand the evaluation methods used in each study and how each result contributes to the overall claim.

Preparation for writing the submission documents should be based on three classes of documents:

1. Sponsor-submitted documents, including Clinical Study Report (CSRs) from earlier studies in the programme and any submission documents from previous programmes of the same product
2. Documented communications with regulatory authorities
3. Templates and guidance provided to sponsors and reviewers^{10,11}

Together, these background materials can help the MW understand the company message, how the knowledge about the product has evolved over time and the agreements reached with the different regulators relevant to global submission. The presentation of both efficacy and safety in the documents should focus on the sought indication in the target patient population, taking care to include the types of information that the regulator is specifically interested in.

The documents in Module 2.7 and in 5.3.5.3 (see Figure 2) generally follow the same outline, however, the 2.7 summary documents are limited

in length (up to approximately 200 pages) and cannot delve into the same level of detail as the ones in 5.3.5.3 that can be thousands of pages long. Thus, it may be advisable to start by writing the ISS and ISE, and then summarise the most important information, taking care not to “cherry pick” favourable results. Reduction in volume can be achieved by fewer methodological details about the pooling and integration, as well as ample use of cross-references from the summary documents to corresponding sections of the 5.3.5.3 documents and to source tables and listings.

In terms of project management, the summary documents and the CSRs of the pivotal studies on which the marketing application relies, are likely to be written concurrently. Thus, a submission should be written by a team of MW, that should maintain very frequent communication to ensure that the messages and focus are consistent across documents. The MW team should plan for multiple cycles of cross-review of submission documents. It may be reasonable to assign a leader of efficacy documents (ISE, SCE) and leader of safety documents (SCS, ISS).

Summary

In conclusion, for valid and informative results of drug safety and efficacy, it is recommended to design studies and their data capture and management strategies with integration in mind. It is important to select trials for pooling with careful attention to trial design and data quality, and to combine selected studies using appropriate statistical methods while being careful with naive data pooling. Like the DM and biostatistics professionals, MW should get involved early in the planning for submission, allowing them to get acquainted with the pivotal points of product information and devise a project management plan.

Disclaimers

The opinions and suggestions presented in this article are based on the authors’ cumulative experience and do not refer directly to any specific global submission situation.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. M4E(R2) – Common technical document for the registration of pharmaceuticals for human use – Efficacy [Internet]. EMA/CPMP/ICH/2887/1999 Committee for Human Medicinal Products; 2016. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-m4e-r2-common-technical-document-registration-pharmaceuticals-human-use-efficacy-step-5_en.pdf
2. Structure and content of clinical study reports, July 1996 [Internet]. CPMP/ICH/137/95 European Medicines Agency; Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-3-structure-content-clinical-study-reports-step-5_en.pdf
3. Guidance for Industry Integrated Summaries of Effectiveness and Safety: Location Within the Common Technical Document. Food and Drug Administration; 2009.
4. Providing Regulatory Submissions In Electronic Format – Standardized Study Data Guidance for Industry [Internet]. Food and Drug Administration; 2014. Available from: <https://www.fda.gov/media/82716/download>
5. MedDRA | [Internet]. [cited 2020 Apr 29]. Available from: <https://www.meddra.org/>
6. ICH E9 statistical principles for clinical trials [Internet]. European Medicines Agency; 1998. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf
7. Meta-Analyses of Randomized Controlled Clinical Trials to Evaluate the Safety of Human Drugs or Biological Products Guidance for Industry – DRAFT GUIDANCE [Internet]. Food and Drug Administration; 2018. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/meta-analyses-randomized-controlled-clinical-trials-evaluate-safety-human-drugs-or-biological>
8. Integrated summary of effectiveness- Guidance for industry [Internet]. Food and Drug Administration; 2015. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/integrated-summary-effectiveness>
9. Chuang-Stein C, Beltangady M. Reporting cumulative proportion of subjects with an adverse event based on data from multiple studies. *Pharm Stat.* 2011;10(1):3–7.
10. Day 80 assessment report – Clinical template with guidance rev.02.20 (DOCX/116.3 KB) (updated) [Internet]. Available from: [https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/assessment-templates-guidance#day-80-and-day-120-assessment-report-templates-\(containing-guidance\)-section](https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/assessment-templates-guidance#day-80-and-day-120-assessment-report-templates-(containing-guidance)-section)
11. Good Review Practice: Clinical Review of Investigational New Drug Applications [Internet]. Food and Drug Administration. Available from: <https://www.fda.gov/media/87621/download>

Author information

Shiri Diskin, PhD, has been a medical writer since 2009, and is currently working as head of medical writing at Bioforum group, a global data-focused clinical research organisation. Shiri teaches medical writing courses to medical doctors and to biopharma companies.

Tanya du Plessis, M.Med.Sc, CCDM, and PMP, has over 16 years industry experience in data management. Tanya has led many strategies for customised data delivery solutions focusing on high-quality data delivered in a timely manner, while driving numerous innovation teams for optimal data delivery.

Ofra Barnett-Griness, PhD, has been a biostatistician for more than 20 years, working with pharmaceutical industry and academia in the design and analysis of randomised clinical trials and real-world data studies. Ofra is a statistical consultant to Bioforum group.