The "dynamic" integrated database for pre-marketing risk assessment – a paradigm shift

Michael J. Klepper*

Limited Liability Company, Cary, NC, USA
*Correspondence: mklepper@mjkmd.com

As Friedrich Hegel stated, "The only thing we learn from history is that we learn nothing from history." In my career of more than 25 years in the pharmaceutical industry, I continue to witness the same lack of planning, systematic sloppiness, and missteps that I saw at the beginning of my career in determining risk during the pre-marketing phase of a drug's lifecycle.

Determining risk during this period is inherently challenging due to limited exposure to an investigational drug. The recommended minimum size of the safety database for long-term use for non-threatening conditions is 1500 patients with 300–600 patients exposed to the investigational drug for at least 6 months and 100 patients exposed for at least 1 year (ICH Secretariat, 1994). This means that based on the "Rule of 3," if 1500 patients were given an investigational drug and an adverse reaction, for example hepatotoxicity, was not seen, one can be 95% confident that the true incidence of hepatotoxicity is less than 0.2% ($1500/3 = 500$; $1/500 = 0.2\%$; Rosner, 1995; US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, 2009). With these known limitations there are many proactive approaches that can be taken that will enhance risk assessment throughout the clinical development phase of a drug's lifecycle. To do this it is helpful to "begin at the end." The "end" is the manufacturer must show a favorable benefit–risk profile before market authorization (drug approval) is granted. To do this, because of the limited exposure to the investigational drug, safety data will in most cases have to be pooled (combined) across clinical studies in order to enhance the ability to identify and characterize the drug's risk profile. Many manufacturers do this data pooling (integration) at the *end* of a clinical

development program, where in reality it should be done at the *beginning* of the clinical development program.

There are many steps required to take the safety data collected from a clinical study site, i.e., raw data, and convert it into data that allows aggregate analyses, i.e., analyses of data from groups of patients, so changes from baseline (pre-treatment) values and comparisons of these changes from baseline can be compared between treatment groups (e.g., the investigational group vs. the placebo group) in order to identify any differences. Any identified differences between treatment groups can indicate a drug effect. These various steps include: data entry; coding of adverse event (AE) data, e.g., to ensure standard medical terms are used for similar medical concepts [e.g., the use of the Medical Dictionary for Regulatory Activities (MedDRA) so an accurate incidence of AEs can be determined]; data programming and data analysis with the creation of tables, listings, and figures (TLFs); and summarizing and interpreting the analyzed data so that a risk assessment can be done (Medical Dictionary for Regulatory Activities, 2010). Any misstep along this data pathway can result in an incorrect assessment of risk. Some of the common problems encountered include: collecting incorrect information; missing information; delays in data entry; data entry errors; programming errors; doing the wrong types of analyses; and errors in data interpretation. Many of these problems are not identified at the individual study level but become more evident when the data are pooled. Programming and data analyses also become more complicated when more studies are combined together. All of these potential problems can be minimized by good planning, ongoing review of the data, correcting data errors, and obtaining missing information in "real time." It is also important to have a close collabora-

tion between the medical writer, medical reviewer, programmer, and statistician to ensure any errors in programming or data analyses are identified early and corrected. This is all accomplished by creating and maintaining a "dynamic" integrated safety database early on in the clinical development program rather than doing this at the time of the preparation of the summary of clinical safety (SCS) and/or integrated summary of safety (ISS). The SCS and ISS are documents that are required for submission in order to obtain market authorization in the European Union and the United States (ICH Secretariat, 2002; US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, 1988). The word "dynamic" is used to convey that the database will not be static but will change and grow with the completion of each study. When data are displayed in TLFs, it is easier to identify missing information and errors such as data or programming errors early on. Fixing errors and obtaining missing information improves the quality of the data. Pooled data with more and more studies added to the database also increases the chances of identifying new safety signals sooner than later. The building blocks of the integrated database are the raw data referred to as data elements, and the statistical analysis plan (SAP) is the blue print of how the database will be built. The SAP defines how the data will be pooled, counted, analyzed, and displayed. The majority of safety analyses are analyzed the same way across studies as well as for different development programs evaluating different investigational drugs. Therefore in many cases a SAP created for one investigational drug can be recycled and used for other investigational drugs. Elements of the SAP for the integrated database can also be used for preparation of individual

clinical study reports. Another benefit is that early integration forces standardization of the data that is a requirement for data pooling. For example, if reasons for premature termination of treatment are not standardized, and different reasons are used across studies, e.g., one study has five reasons, another study has seven reasons, while another study might have six reasons, a standard category for reasons for discontinuation will have to be established for the SCS/ISS so this information can be pooled across studies. In order to do this the studies using different reasons for discontinuation will have to be mapped to the standard categories established for the SCS/ISS. Using standard data at the beginning of the clinical development program will obviate the need for data mapping. By doing this, transcription errors are avoided and time, money, and resources are saved. For all these reasons, the most important being enhanced ability to detect safety signals and better data quality, the creation, and maintenance of a "dynamic" integrated database, is highly recommended.

REFERENCES

ICH Secretariat. (1994). "The extent of population exposure to access clinical safety for drugs intended for long-term treatment of non-life threatening conditions E1," in *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use*, Geneva. Available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E1/Step4/E1_Guideline.pdf [accessed November 30, 2011].

ICH Secretariat. (2002). "The common technical document for the registration of pharmaceuticals for human use-efficacy-M4E(R1) clinical overview and clinical summary of module 2 module 5:clinica safety reports," in *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use*, Geneva. Available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/CTD/M4__R1__Efficacy/M4E__R1_.pdf [accessed November 30, 2011].

Medical Dictionary for Regulatory Activities (MedDRA). (2010). Available at: http://www.meddramsso.com [accessed November 30, 2011].

Rosner, B. (1995). "The binomial distribution," in *Fundamentals of Biostatistics*, ed. B. Rosner (Belmont, CA: Duxbury Press), 82–85.

US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). (1988). *Guideline for the Format and Content of the Clinical and Statistical Sections of an Application*, Washington, DC. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071665.pdf [accessed November 30, 2011].

US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). (2009). *Guidance for Industry Drug-Induced Liver Injury: Premarketing Clinical Evaluation*. Washington, DC. Available at: http://www.fda.gov./downloads/Drugs/GuidancesComplianceRegulatoryInformation/Guidances/UCM174090.pdf [accessed November 30, 2011].