

Data analysis - quality aspects

Jelena Acevska¹, Michel Ulmschneider², Katerina Brezovska¹, Aneta Dimitrovska¹,
Bruno Spieldenner³

¹*Institute of Applied Chemistry and Pharmaceutical Analysis, Faculty of Pharmacy, University "Ss. Cyril and Methodius" in Skopje, 1000 Skopje, R. North Macedonia*

²*F. Hoffmann - La Roche AG, Pharmaceutical Division Quality Control, Basel, Switzerland*

³*European Directorate for the Quality of Medicines and Healthcare, European Pharmacopoeia Department, Council of Europe, Strasbourg, France*

Introduction

The European Pharmacopoeia provides means for assuring the quality of active pharmaceutical ingredients, excipients, and finished dosage forms, covering all aspects of control of medicines in order to achieve their consistency, safety, efficacy, and compliance with specifications. All of these aspects depend on reliable analytical procedures and validity of the results, and consequently on reliable data.

However, it must be acknowledged that new technologies, combined with the ongoing digital transformation and the Industry 4.0 paradigm shift, affect the pharmaceutical industry (Cannavacciuolo et al., 2023). Progresses in data analysis and measurement techniques are speeding up. In particular, because of the multiplication of data sources, more data of all sorts becomes available. Control of manufacturing processes and analytical procedures solely based on direct determinations will be complemented by indirect data-driven and algorithm-based predictions of process and quality attributes. These new developments put many different challenges for the control laboratories.

From the point of analytical testing, data is usually seen through its influence on the results. Interpretation and reporting of results involve decision rules based on statistics (USP <1010> Analytical data- interpretation and treatment). Suitable guidance on the scientifically acceptable evaluation of analytical data is essential.

Over the past decade there has been an increased focus on decision-making processes based on data processing. As a result, regulatory agencies have launched initiatives

through the development of various systems that enable safe sharing of information, motivated by advances in technology, data analytics and artificial intelligence, successfully demonstrate the potential of regulatory decision-making based on data processing.

The European Pharmacopoeia has already advanced to support many different aspects for ensuring good use of modern tools for data analysis, and corresponding issues to manage the quality of the generated data, with a family of cross-linked general chapters entirely dedicated to advanced data analysis. Currently published are chapters: 5.21 - Chemometric methods applied to analytical data, 5.25 - Process analytical technology (PAT), 5.28 - Multivariate Statistical Process Control (MSPC), not to forget the upcoming Chapter 5.33 - Design of Experiments.

This presentation aims to reflect on the challenges related to the critical aspects of managing the data obtained from laboratory analysis, as well as the validity of the information generated by data-driven algorithms, during the decision-making process for compliance of the products, from a perspective of an Official Control Medicine Laboratory (OMCL).

Materials and methods

A review of scientific literature and the current regulatory status was done to point out the critical aspects for ensuring the quality of data and the validity of the data-driven decisions, thus enable future implementation of good practices in healthcare community.

Results and discussion

The pharmaceutical industry and regulatory agencies make fair use of 'big data' during many data-driven decision-making processes (EMA, 2022; Lv, Z. and Qiao, L., 2020). Processing of data from different types of sources (such as data from pharmaceutical development based on QbD, data from clinical trials, real world data, electronic health records, data generated by patients, data from social networks, etc.) enables comprehensive evaluation of the safety, efficacy and quality of medicines in quick manner (Liu, & Panagiotakos, 2022). By leveraging these data sources, regulatory agencies and stakeholders can make informed decisions related to the drug marketing approval process, safety labeling, and post-marketing surveillance (Dash et al. 2019; EMA/787647/2022). Additionally, the advanced statistical and chemometric tools for different areas of application allow for great advantages of analysis of large amounts of data, while enabling the extraction of meaningful information otherwise mined beyond human comprehensibility (Yu, Beam & Kohane, 2018).

However, the usual understanding of data for quality control of medicines is related mostly to the data obtained from the testing the quality requirements, and therefore the 'quality of data' may be interpreted as a way to obtain valid results as a means to reach conformity decision. The path of analytical data becoming result included in the certificate of analysis is the most delicate part of the decision-making process for the compliance of the sample tested. There are many considerations (statistical rather than analytical) to be taken into account in order to be certain that the data is managed based on scientific knowledge and good statistical practices.

Furthermore, laboratory work also encompasses information originated from different data-driven algorithms implemented in many laboratories, even by the most simply organized ones (such as sophisticated software solutions incorporated in laboratory equipment and commercially available applications and prediction tools), that the 'common' analysts still see as 'black boxes' (mainly due lack of fundamental IT and data science knowledge). The quality of the data here is mainly covered by the principals of data governance, covering a lot of different aspects - such as its collection, storage, sharing, update, quality, transfer, but also transformation, enhancement, traceability, etc.

Both aspects of the 'quality of data' are inter-related, as laboratory testing always covers many aspects (analytical, statistical, algorithm-based models, etc.).

Therefore, an unambiguous definition of the term data should make a clear understanding, to distinguish whether it is data derived from laboratory testing becoming an

analytical result, or data in a more general sense to feed algorithm-based techniques.

In any case, many concerns arise out of issues to (or rather the lack of) control the quality of the collected and/or generated data. Elements of good practice should accompany the deployment of any data-driven usage, to build confidence into modelling and prediction, leading to decision on pharmaceutical quality or other healthcare decisions.

Conclusion

Decision-making based on data processing became a new and significant aspect of pharmaceutical regulation, opening up many possibilities for improvement in the development phase of the medicine, its approval for placing on the market as well as its post-marketing follow-up. There is arising need for regulatory guidance on leveling the expectations and applicability for intended use of the information generated by algorithm-based tools. The European pharmacopeia is the least common denominator between the OMCLs and industry's way of thinking, and from the point of view of the quality control laboratories, it is the most crucial stakeholder to address the very crucial principles of the data analysis.

References

- Cannavacciuolo, L., Ferraro, G., Ponsiglione, C., Primario, S., Quinto, I., 2023. Technological innovation-enabling industry 4.0 paradigm: A systematic literature review. *Technovation*, 124, 102733. <https://doi.org/10.1016/j.technovation.2023.102733>
- Dash, S., Shakyawar, S.K., Sharma, M., Kaushik, S., 2019. Big data in healthcare: management, analysis and future prospects. *Journal of big data*, 6(1), 1-25. <https://doi.org/10.1186/s40537-019-0217-0>
- EMA's Infosheet Review: real-world data studies. 2022 <https://www.ema.europa.eu/en/about-us/how-we-work/big-data>
- Good practice guide for the use of the metadata catalogue of real-world data sources", EMA/787647/2022, https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/good-practice-guide-use-metadata-catalogue-real-world-data-sources_en.pdf
- Liu, F., Panagiotakos, D., 2022. Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Medical Research Methodology*, 22(1), 287. <https://doi.org/10.1186/s12874-023-01937-1>
- Lv, Z., Qiao, L., 2020. Analysis of healthcare big data. *Future Generation Computer Systems*, 109, 103-110. <https://doi.org/10.1016/j.future.2020.03.039>
- Yu, K.H., Beam, A.L. and Kohane, I.S., 2018. Artificial intelligence in healthcare. *Nature biomedical engineering*, 2(10), 719-731. <https://doi.org/10.1038/s41551-018-0305-z>