EDITORIAL

Meta-analysis in the era of big data

Lucía Silva-Fernández[1] ⓘ · **Loreto Carmona**[2]

Systematic reviews (SR) and meta-analyses (MA) are crucial to synthesize the enormous amount of information gathered to answer specific questions. This synthesis is extremely valuable in order to make decisions, as it allows to get rid of all the noise caused by individual studies. SR are essentially reviews of the published—and unpublished—evidence that follow well-developed guidelines to avoid bias while ensuring maximal reproducibility and transparency. MA statistically combine the results of multiple studies addressing the same question. Due to their relevance and their immediate applicability in decision-making, MA need to be conducted and interpreted with a strict methodology. Such methodology includes selecting in a transparent way the sources of evidence, critically appraising the potential biases of individual studies, and making explicit arguments in favour of combining the data before actually pooling them. The reader should be able to judge whether all relevant and valid sources of evidence were included. They must be transparent about how their conclusions are generated to avoid misrepresentation of the knowledge base.

In this issue, Kelley et al. describe the major components of a SR and MA, discuss the different types of MA and include a guide for the interpretation of results [1]. The paper focuses on aggregate data MA, which is the most common type of MA. The main advantages of MA are the increased statistical power, the potential to address uncertainty when different study findings are contrasting or opposing, the improved estimates of effect and the ability to answer questions not identified in individual trials. Although they clearly are a powerful research tool and play a fundamental role in evidence-based healthcare, they also have some disadvantages. The main controversy is found in the procedure the reviewers use. Even minor

deviations from protocol can produce biased and misleading results [2]. Publication bias is relatively common as positive studies are more likely to be published. The search for studies can also produce biased results if an incomplete set of keywords is used or there is a wide variation in strategies used to search in different databases. The selection bias can happen when researchers do not clearly define the criteria for choosing the studies to be included from the long list of potentially suitable. To find all the appropriate studies to be included can be difficult and time-consuming, and MA require complex statistical techniques. In this context, interpreting and correctly applying to practise the results of MA can be challenging for clinicians. Kelley et al. provide a list of five general questions to help clinicians applying the results of MA to their own clinical practice with the goal of improving patient care [1]. These questions fundamentally recommend to first check if, considering the population included and the questions addressed, the results are applicable to their routine clinical practice. It is also important to compare the different treatment alternatives taking into account the balance between benefits and harms. Finally, the strength of the evidence should be considered before any therapy is applied to patients.

Although MA are highly regarded for compiling data from bigger populations than individual studies, there are other methods, such as big data, that also offer results extracted from a huge number of individuals. Big data refers to datasets whose size or type is beyond the ability of traditional data processing software to capture, manage and process. Big data is characterized by volume, velocity and variety, including structured, semi-structured and unstructured data from different sources and in different sizes, but in all cases, processing implies advanced analytic techniques [3]. Particularly in medicine, big data enables the analysis of data from thousands of patients to identify clusters or correlations between variables obtained from different datasets. The analysis of big data may be behind improvements in personalized medicine, by assessing risks and predicting outcomes, avoiding waste and reducing unwanted variability, all of which may be accomplished by making reporting of patient data automated [4]. In this scenario, where big data have revolutionized

---

See related article, https://doi.org/10.1007/s10067-019-04590-6

✉ Lucía Silva-Fernández
    luciasilva@ser.es

1   Rheumatology Department, Complexo Hospitalario Universitario de
    A Coruña, A Coruña, Spain

2   Instituto de Salud Músculo-Esquelética (INMUSC), Madrid, Spain

biomedical research, the role of MA needs to be reassessed. In the next few years, we will witness a kind of two-speed research in medicine where some conclusions will be very quickly reached by analysing big data but small studies with classical methodology will still be conducted. Although the volume and velocity seem clear advantages of big data, the validity, variety and value of the data and of the correlations that can be obtained are still under scrutiny [4]. Human inspection at the big data scale is impossible, and there is a desperate need for intelligent tools for accuracy and believability control and handling of information missed [5]. Most information of big data comes from electronic healthcare administrative records, which makes it unstructured and difficult to use. The use of this information has also raised ethical challenges about individual rights, privacy, autonomy, transparency and trust [6].

The European League against Rheumatism (EULAR) has recently developed points to consider for the use of big data in rheumatic and musculoskeletal diseases [7]. They especially emphasize that big data should be findable, accessible, interoperable and reusable, and that privacy must always be applied to the collection, processing, storage, analysis and interpretation of big data. Given the variability on big data publications, they encourage to report explicitly and transparently the methods used to analyse big data (artificial neural networks, support vector machine, random forests, etc.). Finally, before implementation, it is crucial that conclusions or models drawn from big data are independently validated due to the risks of overfitting and the opacity problem of the black-box.

Although the traditional pyramid of evidence, where MA were on the top, may no longer be sustainable in the era of big data, given the current limitations of healthcare big data processing, MA still stand as key elements for clinical decision-making. It may be the time to seek a more integrated approach in research where single studies are incorporated into large registries and networks of large datasets. New approaches like sparse MA, in which variable selection for MA is based solely on summary statistics, have been proposed to handle big data into MA [8]. Hopefully, in a near future, we will be able to contemplate the integration of these two-speed research methodologies into a more developed and efficient approach.

Compliance with ethical standards

Disclosures   None.

References

1. Kelley GA, Kelley KS (2019) Systematic reviews and meta-analysis in rheumatology: a gentle introduction for clinicians. Clin Rheumatol. https://doi.org/10.1007/s10067-019-04590-6
2. Cheung MW, Vijayakumar R (2016) A guide to conducting a meta-analysis. Neurophychol Rev 26(2):121–128
3. Ristevski B, Chen M (2018) Big data analytics in medicine and healthcare. J Integr Bioinform 15(3):20170030. https://doi.org/10.1515/jib-2017-0030
4. Viceconti M, Hunter P, Hose R (2015) Big data, big knowledge: big data for personalized healthcare. IEEE J Biomed Health Inform 19(4):1209–1215
5. Mirkes EM, Coats TJ, Levesley J, Gorban AN (2016) Handling missing data in large healthcare dataset: a case study of unknown trauma outcomes. Comput Biol Med 75:203–216
6. Vayena E, Salathé M, Madoff LC, Browstein JS (2015) Ethical challenges of big data in public health. PLoS Comput Biol 11(2): e1003904
7. Gossec L, Kedra J, Servy H, Pandit A, Stones S, Berenbaum F et al (2019) EULAR points to consider for the use of big data in rheumatic and musculoskeletal diseases. Ann Rheum Dis. https://doi.org/10.1136/annrheumdis-2019-215694
8. He Q, Zhang HH, Avery CL, Lin DY (2016) Sparse meta-analysis with high-dimensional data. Biostatistics 17(2):205–220