

## EDITORIAL



## Are real world data real world data?

© The Author(s), under exclusive licence to Springer Nature Limited 2025

*Leukemia* (2025) 39:2311–2312; <https://doi.org/10.1038/s41375-025-02751-y>

Unless you have been sleeping under a rock you will have noticed an explosion of clinical articles with titles threatening to present *real world data*. This, of course raises the question of what *real word data* are and how they contrast with *unreal world data*.

Briefly put, *unreal world data* are data from clinical trials. Why unreal? This answer lies with strict entry-criteria of most clinical trials which, test interventions (i.e., dose and schedule) different from what most people are likely to receive and where compliance is measured. And there are other discordances over who gets an intervention and how it is given to most people. For example, results of several recent US Veterans Administration prostate cancer studies were found to apply to fewer than 10 per cent of men with prostate cancer.

Because of these limitations we need *real world data* to understand safety and efficacy of an intervention. But what are *real world data*?

The common definition is *big data* extracted from large datasets, usually electronic health records (EHRs) using computers assisted by artificial intelligence and machine learning. These datasets come from insurance claims, billings, disease or drug registries and patient-generated data such as those on mobile phones. Data of this type is available in some but not in most countries. We discuss these issues in detail elsewhere [1, 2]. Currently, most efforts to aggregate data from physician offices and EHRs are sponsored by pharma. Even when such data are available there are constraints: (1) subject-level data are needed; and (2) the sample needs to be very large, especially when dealing with new therapies, under-represented phenotypes, and genotypes and rare diseases; and (3) data in EHRs are collected for health care management, not research consequently important biological, clinical and therapy data may not be studied, recorded or may not be in a useable form.

There are other considerations. Generalizability of *real word data* is not always possible: results in one population may not apply to another. 2<sup>nd</sup>, data sharing requires universal or at least inter-operable technical standards. 3<sup>rd</sup>, data protection is critical. It is important to regulate personal data processing and sharing whilst pursuing the public interest to avoid conflicts between personal and research freedom.

So why has there been an explosion in the title with *real world data* in the medical literature? The causes are 3-fold: (1) misunderstanding of what *real word data* are; (2) confusion over the distinction between *real word data* and *real world evidence*; and (3)

the hope adding *real world data* to the title increases the likelihood of acceptance.

Most typescripts submitted to *LEUKEMIA* entitled *real world data* not so. For example, we recently received a typescript whose title included term *real world data*. It was a retrospective analysis of about 400 subjects treated at 10 centers of excellence. Study-entry criteria were defined and therapy dose and schedule were specified. The study was registered in clinical trials databases such as ClinicalTrials.gov and had Ethics Committee approval. This is decidedly not *real world data*. We also need to distinguish *real world data* from post-approval surveillance studies with voluntary reporting and from studies of observational databases. These studies, although valuable, are also not what is meant by *real world data*.

In this Editorial we review potential use of *real world data* to generate *real world evidence* of safety, efficacy and bases for clinical decision-making in haematology. Although *real world data* cannot replace clinical trials they are needed to support appropriate health care decisions. In the future technology advance in artificial intelligence will make it possible to increase meaningful *real world evidence* drawn from *real world data*.

When authors debate adding *real world data* to the title of their typescript they should consider whether their study is really what is meant by *real world data*. In our experience and that of other journal Editors few studies labeled *real world data* meet the criteria discussed above. An even further leap is suggesting a study provides *real world evidence*.

We encourage potential authors to carefully consider these definitions before submitting a typescript to *LEUKEMIA* and to other journals. We will lose the important potential contribution of *real word data* and *real world evidence* if we misuse these terms.

Robert Peter Gale <sup>1</sup>✉ and Andreas Hochhaus <sup>2</sup>

<sup>1</sup>Centre for Haematology, Department of Immunology and Inflammation, Imperial College of Science, Technology and Medicine, London, UK. <sup>2</sup>Klinik für Innere Medizin II, Universitätsklinikum Jena, Jena, Germany. ✉email: [robertpetergale@alumni.ucla.edu](mailto:robertpetergale@alumni.ucla.edu)

## REFERENCES

1. Passamonti F, Corrao G, Castellani G, Mora B, Maggioni G, Gale RP, et al. The future of research in hematology: integration of conventional studies with real-world data and artificial intelligence. *Blood Rev.* 2022;54:100914. <https://doi.org/10.1016/j.blre.2021.100914>.
2. Passamonti F, Corrao G, Castellani G, Mora B, Maggioni G, Della Porta MG, et al. Using real-world evidence in haematology. *Best Pract Res Clin Haematol.* 2024;37:101536. <https://doi.org/10.1016/j.beha.2024.101536>.

**ACKNOWLEDGEMENTS**

RPG acknowledges support from the UK National Institute of Health Research (NIHR).

**AUTHOR CONTRIBUTIONS**

RPG and AH conceived and developed the typescript, take responsibility for the content and agreed to submit for publication.

**COMPETING INTERESTS**

RPG is a consultant to Antengene Biotech LLC; Consultant Shenzhen TargetRx; Medical Director, FFF Enterprises Inc.; A speaker for Janssen Pharma, BeiGene and Hengrui Pharma; Board of Directors: Russian Foundation for Cancer Research Support and Scientific Advisory Board, StemRad Ltd. AH Research support from Novartis, BMS, Pfizer, Incyte, Terns, Enliven; Honoraria: Novartis, Incyte.