

Peer Review

Review of: "Strategies for Robust, Accurate, and Generalisable Benchmarking of Drug Discovery Platforms"

Belén Otero Carrasco¹

1. Universidad Politécnica de Madrid, Spain

The article addresses a critical need in drug discovery by improving benchmarking protocols, which is essential for reducing costs and inefficiencies in pharmaceutical development. It introduces novel metrics, such as "nAIA" and "nNDCG," which enhance the early identification of relevant candidates and provide a more accurate reflection of system performance. Furthermore, the study systematically identifies and optimizes key parameters that directly affect platform performance. A comparison between two drug-indication databases demonstrates how data quality influences benchmarking results. The development of *compbench*, an open-source, standardized benchmarking protocol, contributes significantly by enabling consistent and replicable comparisons between platforms. Additionally, the authors conduct detailed correlational analyses that investigate factors influencing performance, such as indication size and chemical similarity, providing a solid foundation for future optimizations. Finally, the optimization of parameters leads to clear improvements in CANDO's performance, with specific configurations showing distinct advantages.

As limitations in this study were detected:

1. The evaluation relies heavily on specific data sources (CTD and TTD), which may limit the applicability of the findings to other datasets or discovery scenarios.
2. While the metrics and tools introduced are robust, it remains unclear whether the methodology is scalable to larger platforms or different configurations.
3. The study highlights the superior performance of TTD due to its stricter FDA-approved associations. However, this reliance on curated data may limit the protocol's applicability to

novel drug discovery.

The article makes valuable contributions to the field of drug discovery benchmarking, particularly through the development of improved metrics and open-source tools. It is suggested that grouping related references could improve clarity and help streamline the argument presented in the article. Additionally, future work could focus on validating the predictions in experimental settings to confirm their practical efficacy. Addressing scalability concerns would also enhance the applicability of the findings across a wider range of platforms, ensuring the broader utility and relevance of the proposed methodologies.

Declarations

Potential competing interests: No potential competing interests to declare.