FDALabel for drug repurposing studies and beyond

To the Editor — Prescription drug labeling (also known as the package insert or prescribing information) accompanies every US Food and Drug Administration (FDA)-approved medicine and is intended as a primary mechanism that communicates science-based information to healthcare professionals[1]. This requirement is embedded in the US Code of Federal Regulations (21 CFR 201.56). Drug labeling is proposed by the manufacturer and approved by the FDA. It must include a summary of information for the safe and effective use of the drug that is informative and accurate and does not include promotional, false or misleading statements. The primary purpose of drug labeling is to give healthcare professionals the information they need to prescribe the medicine appropriately. However, the labeling contains a treasure trove of information that can be used to support drug development and drug repurposing efforts.

The overall challenge in leveraging the information from drug labeling is that the marketplace has thousands of prescription drugs with a corresponding large number of labeling documents, many of which are not available in a machine-readable or text-searchable form. Finding specific labeling information about a given drug or drugs, and their indicated medical conditions, can be very difficult and frustrating. For example, asking the simple question "how many drugs have a Boxed Warning containing a given term (for example, acute liver failure)?" is both a time and labor-intensive challenge. Similarly, it is by no means a trivial task to find information about drug–drug interactions in certain clinical scenarios. With this in mind, the FDA has created FDALabel (https://nctr-crs.fda.gov/fdalabel/ui/search), a database that efficiently captures critical information pertinent to the safe and effective use of medicines provided in drug labeling documents.

FDALabel currently contains over 130,000 drug labeling documents, of which 47,000 are prescription drug and biological product labeling. The data are updated regularly. FDALabel is freely available to the public through Amazon Web Services in an efficient, stable and easy-to-use cloud environment. There are more than 250 daily users of FDALabel and over 2,000 homepage views per month.

FDALabel offers various ways of customizing the access and retrieval of desired information from labeling documents. Specifically, searches can be carried out either for an entire labeling document or documents or within specific sections or subsections. In addition, one can search information for a drug or a class of drugs (for example, serotonin uptake inhibitors) with the option to further narrow the search by application type (for example, New Drug Application or Biologic License Application). The database also provides the user with the ability to make a wide assortment of queries regarding a specific drug (for example, "Does drug X carry a risk of QT prolongation?"), to compare labeling across drugs for differences (for example, "Which drugs are approved for major depressive disorder?"), and to find similarities (for example, "Do all factor Xa inhibitors recommend dose reductions in acute renal failure?"). The results and queries can be saved and shared through
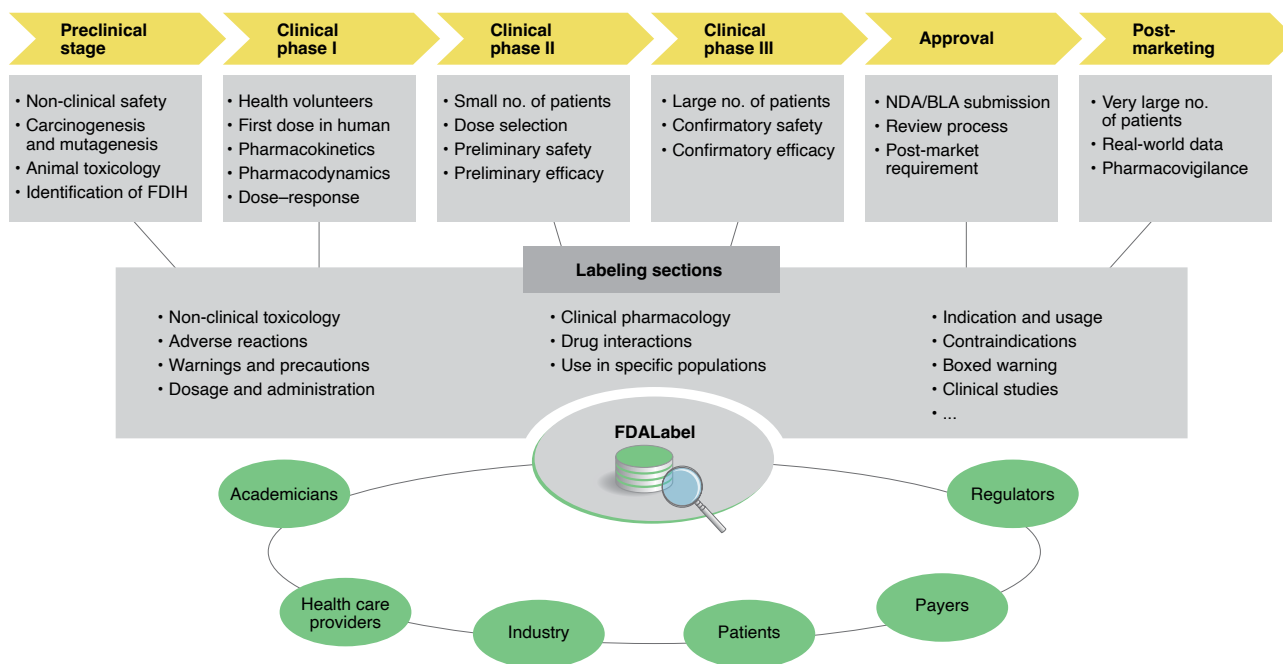

Fig. 1 | Drug labeling in the context of drug discovery and development. The overall stages of the development process are illustrated in the top layer; the data and information generated from this process are captured in the layer below. This information then feeds into drug labeling and is summarized in 17 main labeling sections that can be easily and extensively queried using FDALabel by users (green).

web links to be repeated and reproduced by others.

Some unique functions in FDALabel, not previously available in a single application platform, are particularly useful for addressing broad scientific questions. Drug labeling documents in FDALabel are from FDA's Structured Product Labeling (SPL) archive, submitted by manufacturers. These consist of semi-structured text wherein much of the critical content is contained in the blocks of unstructured free text. To bring a degree of discoverability and insight to the mass of free text parts among all labeling, we have encoded the labeling information with the standard nomenclature of Medical Dictionary for Regulatory Activities (MedDRA), which is a clinically validated international standard of terminology used by regulatory agencies and industry. The terminology is used in the FDA Adverse Event Reporting System (FAERS) and in coding adverse events in clinical trials, as well as in electronic health records. With FDALabel the standardized terminology of drug safety and efficacy data can be retrieved and analyzed across drugs. Chemical structure data are also available for all the drugs in FDALabel for drug similarity analysis and as a linkage to other public databases. In addition, FDALabel allows extraction of information from individual sections or subsections that contain specific data useful to support a variety of research topics, such as safety for individual drugs and drug–drug interaction, mechanistic data, and pharmacogenomics[2].

Using a contemporary example, chloroquine (CQ), hydroxychloroquine (HCQ) and azithromycin (AZI) have been touted for 'off-label' use (that is, not FDA-approved) to treat COVID-19[3,4]. FDALabel was able to elucidate a potential risk associated with taking these drugs not only alone but also in combination in patients with underlying diseases. For example, when AZI is combined with either CQ or HCQ, a search of FDALabel shows that this combination of medications increases the risk of cardiac abnormalities because of additive QT prolongation from combining two drugs with 'QT prolongation' warnings. FDA recently cautioned against use of HCQ or CQ for COVID-19 outside of the hospital setting or a clinical trial due to risk of heart rhythm problems on the basis of case reports in the FAERS database, the published medical literature and the American Association of Poison Control Centers National Poison Data System (https://www.fda.gov/drugs/drug-safety-and-availability/fda-cautions-against-use-hydroxychloroquine-or-chloroquine-covid-19-outside-hospital-setting-or).

Drug labeling information has been used to address a broad range of scientific questions, such as finding up-to-date information on drug safety arising from pharmacovigilance[5,6], providing context for drug repositioning[7], and precision medicine[2,8–10]. We are implementing a variety of functions to enhance such a research endeavor. Specifically, we are developing an artificial intelligence component to allow querying of labeling documents using customized and fine-tuned public language models and algorithms that can enhance the specificity and relevance of labeling data retrieved.

Figure 1 shows how the different stages of drug discovery and development feed information into a drug's labeling. With FDALabel, this information can now be seamlessly and effectively obtained by the research community in the form of a high-quality and searchable drug labeling database. FDALabel can inform drug development and regulatory strategies, clinical decisions and standards of care, and, most importantly, facilitate scientific research. FDALabel complements other sources of drug labeling information (for example, Drugs@FDA and DailyMed) that support the transparency and accessibility of drug labeling for approved products, and now offers the research community a new resource for drug repositioning and repurposing studies and beyond. ❑

Hong Fang[1], Stephen Harris[1], Zhichao Liu[1], Shraddha Thakkar[1], Junshuang Yang ID [1], Taylor Ingle[1], Joshua Xu ID [1], Lawrence Lesko[2], Lilliam Rosario[3] and Weida Tong[1] ✉

[1]National Center for Toxicological Research, US FDA, Jefferson, AR, USA. [2]College of Pharmacy, University of Florida, Orlando, FL, USA. [3]Office of Translational Sciences, Center for Drug Evaluation and Research, US FDA, Silver Spring, MD, USA.
✉e-mail: Weida.Tong@fda.hhs.gov

References
1. Mullard, A. Nat. Rev. Drug Discov. 19, 79–84 (2020).
2. Mehta, D. et al. Drug Discov. Today 25, 813–820 (2020).
3. Gautret, P. et al. Int. J. Antimicrob. Agents 56, 105949 (2020).
4. Borba, M. G. S. et al. JAMA Netw. Open 3, e208857 (2020).
5. Voelker, R. J. Am. Med. Assoc. 315, 1826 (2016).
6. Zhang, V., Olfson, M. & King, M. JAMA Psychiatry 76, 1208–1210 (2019).
7. Pushpakom, S. et al. Nat. Rev. Drug Discov. 18, 41–58 (2019).
8. Bourgeois, F. T. & Kesselheim, A. S. N. Engl. J. Med. 381, 875–881 (2019).
9. Wang, B., Canestaro, W. J. & Choudhry, N. K. JAMA Intern. Med. 174, 1938–1944 (2014).
10. Fang, H. et al. Drug Discov. Today 21, 1566–1570 (2016).

Check for updates

Increased cyber-biosecurity for DNA synthesis

To the Editor — Commercial DNA synthesizers sell billions of nucleotides to customers each year, amounting to hundreds of millions of dollars in sales[1]. As DNA synthesis becomes more widespread, concern is mounting that a cyberattack intervening with synthetic DNA orders could lead to the synthesis of nucleic acids encoding parts of pathogenic organisms or harmful proteins and toxins. Reassuringly, such an attack is non-trivial. Most DNA synthesis providers check each requested sequence across databases of problematic sequences before order fulfillment, as per the 2010 US Health and Human Services guidelines[2]. Unfortunately, however, there are no comprehensive databases of pathogenic sequences, and the guidelines — unenforced outside of US National Institutes of Health (NIH) grantees[3] — are outdated.

Two industry groups have developed additional guidelines to fill the vacuum[4].

The International Association Synthetic Biology's (IASB) 2009 code requires record keeping of suspicious requests and screening of customers and sequences[5]. And the International Gene Synthesis Consortium (IGSC) best practices, revised[6] in 2017, require synthesizers to both scan every sub-sequence of 200 consecutive base pairs (bp) and apply a best-match approach. Suspicious sequences require human inspection to verify safety and legitimacy.