1521-009X/50/1/86–94$35.00
DRUG METABOLISM AND DISPOSITION

dx.doi.org/10.1124/dmd.121.000420
Drug Metab Dispos 50:86–94, January 2022

Special Section on ADME Databases-Minireview

Overview of Current Herb–Drug Interaction Databases

Yufeng Zhang, Chung Man Ip, Yuen Sze Lai, and Zhong Zuo

School of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR

Received February 10, 2021; accepted October 20, 2021

ABSTRACT

An HERB–Drug Interaction (HDI) database is a structured data collection method for HDI information extracted from scattered literatures for quick retrieval. Our review summarized the ten currently available HDI databases, including those databases comprising HDI on the market. A detailed comparison on the scope of monographs, including the nature of content extracted from the original literature and user interfaces of these databases, was performed, and the number of references of fifty popular herbs in each HDI database was counted and presented in a heatmap to give users an intuitive understanding of the focuses of different HDI databases. Since it is well known that the development and maintenance of databases need continuous investment of capital and manpower, the sustainability of these databases was also reviewed and compared. Recently, artificial intelligence (AI) technologies, especially Natural Language Processing (NLP), have been applied to screen specific topics from massive articles and automatically identify the names of drugs and herbs in the literature. However, its application on the labor-intensive extraction and evaluation of HDI-related experimental conditions and results from literature remains limited due to the scarcity of these HDI data and the lack of well-established annotated datasets for these specific NLP recognition tasks. In view of the difficulties faced by current HDI databases and potential expansion of AI application in HDI database development, we propose a standardized format for data reporting and use of Concept Unique Identifier (CUI) for medical terms in the literature to accelerate the structured data collection.

SIGNIFICANCE STATEMENT

The worldwide popularity of botanical and/or traditional medicine products has raised safety concerns due to potential HDI. However, the publicly available HDI databases are mostly outdated or incomplete. Through our review of the currently available HDI databases, a clear understanding of the key issues could be obtained and possible solutions to overcome the labour-intensive extraction as well as professional evaluation of information in HDI database development are proposed.

1. Introduction

Botanical products and/or traditional medicines are an important supplement to our current medical system and may be called herbs, foods, dietary supplements, nutraceuticals, or traditional Chinese medicine (TCM) under different regulation systems around the world (Trovato and Ballabio, 2018; Alostad et al., 2020). For the sake of simplicity, all these products used for health purposes are collectively referred to as "HERBs" to represent a broader concept than "herbs" in this review. Unlike conventional drugs, most HERBs or their preparations have not undergone rigorous assessment on the safety, efficacy, and quality control before their launch into the market (Glisson and Walker, 2010). These HERBs may be intentionally or unintentionally co-administered with western drugs, raising the potential of pharmacokinetic and/or pharmacodynamic HERB–drug interactions (HDIs) (Izzo and Ernst, 2001; Hu et al., 2005) and leading to major safety concerns, especially for drugs with narrow therapeutic indices [e.g., warfarin (Ge et al., 2014) and digoxin (Cheng, 2006)]. Thus, the collection and evaluation of these reported HDIs would be valuable to both frontline healthcare workers and the public. However, most literature about HDIs are either case reports or limited clinical observations that are not well-documented (Fugh-Berman and Ernst, 2001), which usually result in "unable to be evaluated" based on the established reliability criteria for western drugs. Regardless, HDI undoubtedly exists, and its risk is unavoidable to each individual.

After realizing the clinical importance of HDI, a handful of researchers and companies started to build HDI databases using various information technologies (IT) in the late 1990s (Bailey, 2011; Lin, 2011; Vardell, 2015; Kluwer, 2018; Wu et al., 2019; Birer-Williams et al., 2020; Wang et al., 2020; UW, 2021; Zhang and Zuo, 2021). Nowadays, these HDI databases could be divided into two categories based on their availability to the public: (1) freely accessible databases and (2) commercially available databases. In general, building a professional database, including HDI databases, consists of three key steps. The first step is to identify the data sources from which to collect the needed information; the second step is to grab/digitize and save the data locally for further processing; and the third step is to obtain structured data through extraction and evaluation of the original data. The first two steps are

This work is financially supported by Innovation Technology and Commission of the Hong Kong Special Administrative Region, China (Reference number: ITS/099/18FX).
dx.doi.org/10.1124/dmd.121.000420.

ABBREVIATIONS: CMSS, the Chimei Search System; CUI, Concept Unique Identifier; CWMIIN, The Chinese–Western; DHIQW, Drug–Herb Interaction Query Website Medicine Integrative Information Network; DIDB, UW Drug Interaction Database; HDI, Herb–Drug Interaction; LDI, Lexicomp Drug interactions; NMCD, Natural Medicines Comprehensive Database; Probot, Probot Chinese Medicine–Drug Interaction Database; SHMI, Stockley's Herbal Medicines Interactions.

usually the major work in the early stages of database development and require both the opinions of experts/specialists and IT support. However, the third step consists of the most time-consuming and labor-consuming work, which exists throughout the life cycle of database development and maintenance. Thus, the third step on how to obtain structured data from text materials written in natural languages is a rate-limiting step for the database development. HDI information is buried in all kinds of natural language texts, such as abstracts, journal papers, books, or drug evaluation reports. Only well-trained researchers with a medical background can fulfill this kind of work. For freely accessible databases, they tend to stop updating after their publication due to a lack of continuous funding support. On the other hand, successful commercially available databases could cover all related costs for database development and maintenance by charging annual fees. Recently, to solve the labor- and time-consuming problems in HDI database development, artificial intelligence (AI) technologies have been used to automatically extract HDI information from the literature and displaying the obtained evidence on SUPP.AI (Wang et al., 2020). The application of AI highlights the rate-limiting step of HDI database development, although AI is still far from serving as a substitute for the work of experienced medical researchers in extracting and evaluating HDI information from literature.

In this review, we provided a summary of the coverage of HERBs, main features, source, search and export method, and content update frequency for the ten most popular freely accessible and commercially available HDI databases in **Table 1** and illustrated as follows. To further facilitate the choice of HDI database from user point of view, we also compared the number of HDI-related references for 50 selected popular herbs among these databases as shown in **Fig. 1**. Those databases focused on drug-drug interactions (DDI), such as Cortellis Drug Discovery Intelligence, DrugBank, Medscape, ONCHigh, PharmaPendum, and WebMD, were not covered in this review.

2. Freely Accessible HDI Databases

2.1. The Chi Mei Search System (CMSS). CMSS was developed by the Department of Pharmacy, Taiwan Chi Mei Medical Center in October 2004 (Chimei, 2004), and version 2 is its most updated interface. The database embedded a search box for searching herb names, brand, or chemical names of western drugs. A total of 139 herbs and 52 traditional Chinese Medicine (TCM) formulae were included, which resulted in 6,173 interaction pairs with western drugs (**Table 1**). In addition, possible mechanisms of interaction, clinical manifestations, and recommendations were summarized for each herb/TCM-drug interaction pair. CMSS covers a relatively large number of herbs/TCM and their interactions, but no sources of these interactions were provided, and it is hard for others to re-evaluate those potential HDI interactions.

2.2. The Chinese–Western Medicine Integrative Information Network (CWMIIN). The CWMIIN was developed by Prof. Lin from China Medical University with grant support from the Ministry of Health and Welfare of Taiwan (Lin, 2011) and was first published in 2004, followed by subsequent updates in 2008 and 2011. The coverage in the current version 2011 system is 30 TCM formulae, 72 herbs, 12 foods, and 4 herbal components, which interacted with 171 western drugs from 607 references (**Table 1**). Unlike those interaction pairs listed in CMSS, all records in CWMIIN have a link to the source of the HDI information. Users can perform HDI searches by entering an herb list and a drug list or browse all HDI entities from the predefined lists of TCM formulae, herbs, or drugs. The search results include a brief summary of the HDI with the names of the herbs and drugs involved.

2.3. Drug Herb Interaction Query Website (DHIQW). The DHIQW was the outcome of a collaboration between Prof. CS Wu (National Formosa University) and Prof. ZH Wu (Taipei Medical University), with grant support from the Ministry of Science and Technology of Taiwan (Wu et al., 2019). Current popular front-end and back-end frameworks, such as Vue.js and Node.js, were used for the database development, resulting in a user-friendly and responsive platform. The HDI information extracted from literature in DHIQW were rewritten by pharmacists to render the information easy to be understood by general public. Until now, only three herbs (Ginseng, Ginkgo and Dong Quai) and 300 pairs of HDIs were included in this database as shown in **Table 1**. Two types of searches, including single search and smart search, are allowed. The former only requires either the name of an herb or a drug, while the latter allows users to search for HDIs of multiple herb-drug pairs. In the search results, the HDIs are bilingually summarized with a few phrases in both the English and Chinese languages. Other information, such as details of the study that lead to the conclusions on HDI, mechanism behind the HDI, implication of HDI, and details about the sources, are also given in the search results.

2.4. Center of Excellence for Natural Product-Drug Interaction Research (NaPDI). NaPDI is a data repository for pharmacokinetic natural product–drug interactions developed by the National Institutes of Health National Center for Complementary and Integrative Health (Birer-Williams et al., 2020). This database is still under development, and as of July 2021, interactions of 7 herbs with 259 compounds have been included as indicated in **Table 1**. The unique feature of NaPDI is that structured data, including experimental conditions and pharmacokinetic parameters of in-vitro and in-vivo studies, are extracted for four experiment type categories to guarantee FAIR (Findability, Accessibility, Interoperability and Reusability) data in NaPDI. In addition to supporting refined searches, data in NaPDI can be browsed by the titles of Natural Products, Studies, or Compounds. In addition to the published reports included in NaPDI, there are also six unpublished studies in the current database.

2.5. Probot Chinese Medicine–Drug Interaction Database (Probot). Probot aims to collect all published evidence about interactions between Chinese Medicine and drugs to provide factual data for clinicians and the public. This database, supported by Healthy Power Limited and Innovation and Technology Commission, Hong Kong, is still under development (Zhang and Zuo, 2021). As of July 2021, 6,292 interactions between 193 herbs and 726 western drugs originated from 4,342 references are included (**Table 1**). Probot supports bilingual display and query in Chinese and English. For the maintenance of the website, abstracts from PubMed, Wanfang, and CNKI are automatically retrieved by in-house programs and screened for their relevance to HDI using a Naive Bayes model (Precision=0.78, Recall=0.91, F1-score=0.84). Only those HDI-related abstracts are further processed by experienced pharmacists. Such an approach improves the automation of database development so as to allow more focus on manual extraction and evaluation of detailed information from literature that AI may not be able to perform effectively.

2.6. SUPP.AI. In 2019, a team from the Allen Institute for Artificial Intelligence (AI2) developed the SUPP.AI database, with the aim of providing scientific evidence for supplement-drug interactions by automatically extracting supplement information from the scientific literature (Wang et al., 2020). Apart from searching evidence on supplement-drug interactions, users can also download their dataset, as well as access programmatically with their Application Program Interface (API). As shown in **Table 1**, the database contains information about 60,000 interactions, with 195,000 evidence sentences extracted from 22 million articles. Specifically, about 2,044 supplements and about 2,842 drugs are involved in these interactions as of July 2021. Users can perform

TABLE 1.

Comparison of the properties from the ten most popular HDI databases.

Name[1]	Coverage[1]	Features	Sources	Search and export method	Content update method and frequency
CMSS	• 139 herbs, 52 TCM formulae and 251 drugs; • 5482 herb–drug and 691 TCM formula–drug interactions of PK/PD.	• Mechanism of interaction, clinical manifestations and clinical recommendations are included for each interaction.	NA	• Search by keywords of herb or drug; • Cannot export structured data.	No update after 2004.
CWMIIN	• 118 TCMs (30 formulae, 72 herbs, 12 food and 4 herbal components) and 171 drugs; • 396 PK/PD interaction pairs extracted from 607 references.	• A detailed information page for each interaction pair with a one sentence conclusion and related references; • Browse interaction pairs by names of drugs or TCMs.	PubMed	• Search by names of both TCM and drug; • Cannot export structured data.	Updated at 2008 and 2011 with no further update afterward
DHIQW	• 3 herbs (Ginseng, Ginkgo and Dong Quai) and 48 drugs; • 80 PK/PD interaction pairs from 74 references.	• Bilingual search and display for the whole website; • User-friendly interfaces developed using Vue.js.	PubMed	• "Single search" by a keyword of herbs or drugs; • "Smart search" for exploring multiple interactions between drugs and herbs; • Cannot export structured data.	• Content is evaluated and summarized by the pharmacist. • No update since 2019.
NaPDI	• 7 herbs and 80 drugs; • 119 PK interaction pairs from 90 references.	• Experiment conditions and detailed results of each parameter were included; • Links all of the experimental data for a given natural product across different studies; • Search and filter options.	• Natural product-related studies from NaPDI Center; • PubMed; • NDA/BLA reviews at drugs@FDA	• Search by keywords of herb or drug and filter by name of natural product/compound/ drug/enzyme/ transporter, experiment type, measurement and test type, overall effect; • Structured data in search results can be exported to a CSV file.	Data entry based on validated SOPs. It is under development and only partial datasets are publicly available.
Probot	• 193 herbs and 726 drugs; • 6,292 PK/PD interaction entries based on 4,342 references.	• Bilingual support for search and display; • Bilingual browse by drugs or herbs in alphabetical order.	PubMed, CNKI and Wanfang	• Search by Pinyin or names of herbs or drugs in both Chinese and English; • No structured data to be exported.	HDI-related references were automatically fetched and identified from PubMed, CNKI and Wanfang once a week followed by manual curation and verification.
SUPP.AI	• 2,044 supplements and 2,842 drugs; • 195K evidence sentences extracted from 22M references which lead to 60K PK/ PD interactions.	• "Evidence sentences" for each interaction pair were grouped together; • Entities of supplement and drug in "evidence sentences" were highlighted in different colors; • All "evidence sentences" linked to their source papers.	Semantic Scholar	• Search by names of supplements or drugs and further filter of possible interactions by drug name; • Application Programming Interface available to retrieve all interaction information from SUPP.AI.	Updated by programs with no manual validation once in several months.
DIDB	• 2,539 natural products, including herbal medications and food products; • 15,864 PK interaction experiments/studies from 2,501 references.	• Both preclinical (in vitro human) and clinical data included • Results are categorized based on mechanism of interactions • Detailed experimental conditions and results are curated, as well as clinical outcome (pharmacokinetic, safety, and pharmacodynamic effect)	PubMed, Embase, product label and NDA/BLA reviews at drugs@FDA	• Search by compound, enzyme, transporter, therapeutic class, in vitro/in vivo/ pharmacokinetic parameter using over 70 pre-formulated queries • Structured data in search results can be filtered and exported to excel and csv files • Site search also allows retrieval based of any specific keywords	Manual curation by research scientists and faculty with MD, PhD, or PharmD degrees, with new information added every day

TABLE 1. *continued*

Name[1]	Coverage[1]	Features	Sources	Search and export method	Content update method and frequency
		• Provide links to original references, e.g., PubMed, doi.org, Semantic Scholar, Drugs@FDA • Impact of food and disease (renal, hepatic impairment) is also included.			
LDI	• 85 herbs and 2011 drugs; • 295 PK/PD herb-drug/ herb-herb interactions from 902 references.	• Interaction Rating assigned to each interaction as "Avoid combination", "Monitor therapy", "No known interaction", "Consider therapy modification", or "No action needed"; • HDI is summarized based on "Interacting Members". Herbs/drugs under the same pharmacological category are considered as "Interacting Members".	Prescribing information of global marketed drugs and PubMed	• Search by keywords of herb or drug; • Interaction details can be printed with no structured data exported.	NA
NMCD	• 1,286 products including food, herbs and supplements; 433 drug categories; • 3,207 PK/PD interactions of these products with drugs from 4,632 references.	• Drugs having interactions with HERBs were grouped based on their targets or metabolic enzymes; • Significance of interaction labeled as "Minor", "Moderate", and "Major"; • Interactions with disease conditions also included; • Browse by first letter of product name.	Books, conference papers, and PubMed	• Search by: o keywords of products; o keywords on specific product information pages in NMCD; • Printable product information pages with no export of structured data.	Content in product information pages are updated daily by editors on a regular schedule.
SHMI	• 216 monographs of herbal medicines, dietary supplements, and nutraceuticals; 256 drugs/drug categories; • 1,082 PK/PD interactions in all monographs which originated from 1,956 references.	• Experts summary for each interaction on "clinical evidence", "experimental evidence", "mechanism", and "importance and management". • Browse by first letter of monograph titles.	PubMed	• Search by keywords of herbs or drugs; • Cannot export structured data.	Manual update by pharmacists every 1-3 months.

[1]The definition for HERBs was consistent with that in each HDI database.

searches with keywords, such as the name of an herb. The search results include possible HDIs at the entity page and the relevant evidence sentences at the interaction page, and the herbs and drugs are highlighted in each evidence sentence. The sources of the evidence sentence, as well as links to further details of the sources provided by the semantic scholar database, are also given. An important advantage of this database is its automated approach to extract evidence sentences, which not only saves time and avoids manual efforts, but also provides convenience for users to process data with customized computer programs for specific purposes. This database covers a relatively large number of supplements and their interactions with drugs. Limitations of the methods for providing the HDI information are discussed in the paper published by the developers (Wang et al., 2020). The arbitrary distinction between drugs and supplements, the lack of a standardized terminology for supplements, and the weakness of the natural language processing (NLP) tools employed by SUPP.AI all seem to limit SUPP.AI's capability to identify potential HDI information from literature.

3. Commercially Available HDI Databases

3.1. UW Drug Interaction Database (DIDB). The DIDB was founded by Dr. René Levy at the University of Washington in the late 1990s, and the subscription program was started in 2002 (Hachad et al., 2010). DIDB has the largest manually curated collection of in vitro and clinical data related to drug interactions in humans (no data from animal studies), including interacting co-medications, excipients, food products, herbals, tobacco, organ impairment, and genetics, which can affect drug exposure in humans. This database integrates information from the literature, drug labels, FDA drug approval review packages for new drug applications (NDAs), and biologics license applications (BLAs). Relevant information from these resources is manually extracted and presented in DIDB in a well-structured manner based on the mediated mechanism(s), e.g., enzyme or transporter inhibition or induction. Both in vitro kinetic and clinical pharmacokinetic parameters as well as detailed experimental conditions, study design, and dosing regimen are curated. Clinical outcome of each interaction includes pharmacokinetic,

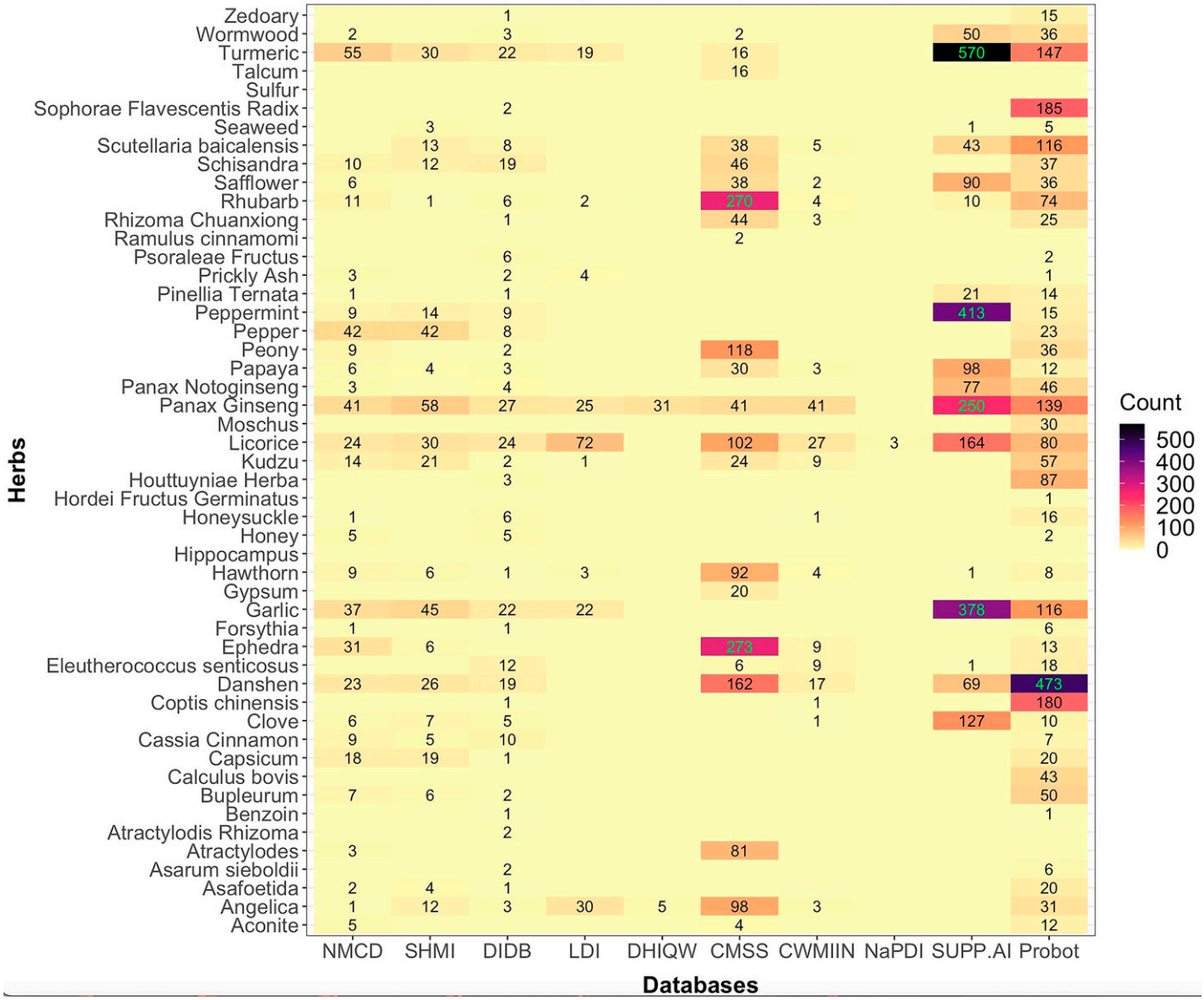

Fig. 1. Comparison of HDI database coverage for the selected 50 herbs. Numbers of references larger than zero were marked at the intersection of herbs and HDI databases.

pharmacodynamic, and safety. The content in DIDB is validated by experts and updated daily. The database can be searched not only using basic keywords (i.e., drug name, enzyme, transporter, therapeutic class, etc.) but also using more specific parameters of interest, such as in vitro parameters, changes in exposure, QT prolongation, etc. A menu of over 70 pre-formulated queries allows users to search and integrate preclinical and clinical data across multiple studies. In addition to data curation, the database also provides drug monographs for recently marketed drugs, with a detailed DDI summary based on available data. Of note, HDI information is a growing fraction of the DIDB. As of June 2021, the application contains a total of 2,539 natural products (herbal medications and food products), with 15,864 drug interaction experiments/studies (**Table 1**).

3.2. Lexicomp Drug Interactions (LDI). Information regarding HDI can be accessed by using the interaction module of the Lexicomp database (Kluwer, 2018). This module is part of UpToDate, which is under Wolters Kluwer, a global provider of professional information for a wide variety of sectors. UpToDate has been marketed as "the most trusted evidence-based clinical decision support resource at the point of care". The target users are medical professionals who need to provide medical advice to clients on a regular basis. Clinical evidence was pre-processed by a team of authors and editors. Data in LDI are mostly for DDI as those in DIDB. Among the 2,096 entities included by LDI, only 85 were herbs, and the total number of herb–drug or herb–herb interaction pairs were 295 which were from 902 references. One of the following interaction ratings is assigned to each interaction: "Avoid combination", "Monitor therapy", "No known interaction", "Consider therapy modification", and "No action needed". LDI organizes all interactions under "Interacting Members", and the member compounds in the same "Interacting Members" are thought to have the same interaction although there is no publication to support.

3.3. Natural Medicines Comprehensive Database (NMCD). The NMCD was developed by the Therapeutic Research

Center (TRC), an organization set up in 1985 with the aim to "positively impact patient care and reduce medication errors in the U.S and beyond". This database is marketed as "the most authoritative resource available on dietary supplements, herbal medicines, and complementary and integrative therapies" (Yacobucci, 2016). NMCD is in fact a collection of databases, and to search for HDI information, users would need to use the "Food, Herbs and Supplements" database. As of July 2021, this database has included more than 1200 products for food, herbs, and supplements as indicated in **Table 1**. Users can either access information by performing keyword-based searches and/or select from a list of food, herbs, and supplements. In the HDI section of each monograph, the interactions are characterized by the interaction rating, severity, likelihood of occurrence, and level of evidence, followed by a description of the interaction. Apart from the rather comprehensive coverage of herbs, this database also provides easy-to-understand monographs that are specifically prepared for consumers rather than professionals. For each monograph, an image of the item, e.g., herb is provided, which gives the users an idea of the appearance of the searched items, or the sources of the searched items. Users may also find products that contain the particular searched items by using the links provided in the "commercial products" section. In addition, each monograph is reviewed at least once per year, and the date of review and any update are given at the bottom of the monograph. Information in the database should be reliable since it is processed and validated by health professionals licensed to practice in their specialty area. Consumer information about herbs and HDIs are also available in French and Spanish as well as from English. Some of the monographs are still under development; thus, users will find "insufficient reliable information available" in some sections.

3.4. Stockley's Herbal Medicines Interactions (SHMI). As one of the databases from MedicinesComplete, published by the UK Royal Pharmaceutical Society, SHMI has been an important online platform for exploring HDI information since its launching in 2004 (Bailey, 2011; Rice, 2014). SHMI aims to provide quick and easy access to core and specialist resources at the point of care. The database contains monographs of about 216 herbal medicines, dietary supplements, and nutraceuticals originating from about 2,000 references as shown in **Table 1**. The general information page of each 216 HERBs includes its synonyms, constituents, indications, pharmacokinetics, and interaction monographs. Under the section of interaction monographs for each herb, all drugs or drug categories having interactions with that HERB are listed with hyperlinks to the HDI monograph for each HDI pair. The database also offers quick access to "related content" of the queried HDI, so users can conveniently look up other relevant HDIs, such as those involving the herb of the queried HDI. This database covers a relatively large number of herbs, and information is regularly updated, processed, and validated. Some pedagogic guidance on the usage and relevant scientific principles are available, which facilitate users' understanding of herbal medicines and potential risks.

4. Summary and Perspectives

4.1. Coverage of Herbs and References. For fair comparison, only HDI-related information were included in **Table 1** and **Fig. 1,** with those interactions between a single herbal component and drugs or two drugs being excluded. An ideal HDI database is to include all HERBs and all HDI-related references; however, no currently available HDI databases meet such requirement. In general, commercial (DIDB, LDI, NMCD, and SHMI) and newly developed (Probot) databases have broader coverage than free ones. To visually compare the coverage of herbs and HDI-related references in the ten reviewed databases, we selected 50 popular herbs from 616 herbs in 2015 China Pharmacopoeia based on number of research papers obtained in Wanfang and PubMed, the two well-known abstract databases for English and Chinese articles, respectively. As a result, the top 50 herbs were identified as shown in **Fig. 1**. To find whether there is a match for each herb, the plant/herb name as well as their synonyms in both English and Chinese were searched in the ten HDI databases. For example, the matched terms for Licorice were (1) "Licorice" in DIDB, LDI, NMCD, NaPDI, and SUPP.AI; (2) "Licorice" in SHMI; (3) "甘草" in CMSS, CWMIIN, and Probot; (4) "Gan Cao" in Probot. The search results for the 50 selected herbs were summarized in **Fig. 1**. Since no source is available for CMSS, the number of HDI entries were used to represent the number of references instead. It is noted that CMSS, Probot, and SUPP.AI had a better coverage of popular herbs than other HDI databases. Although NMCD and SHMI are also good sources for HDI, their relatively lower reference numbers could be mainly due to the insufficient inclusion of literature written in Chinese in PubMed, the source of these two databases. Despite low coverage of herbs, DHIQW and NaPDI are still very useful for HDI information of their included herbs due to unique features, such as structured data for experimental conditions and results.

4.2. Nature of Content Extracted from the Original Literature. Except for CMSS, all other databases provide source links where applicable, and most of the data appears to come from PubMed as indicated in **Table 1**. Some databases also include prescribing information, books, conference papers, even regulatory documents, although these are not common since alternative data sources might be less accessible than PubMed. Many databases use data from post-processed original literature. The post-processing work is typically performed by health professionals and researchers, for the sake of validating raw data and simplifying technical details which would otherwise be difficult for general audiences to understand. An exception is SUPP.AI, in which evidence sentences are extracted directly from original references without further rephrasing nor validation by professionals. The advantage is that it reduces manual effort: however, the quality of the HDI information extracted without verification by professionals might be questionable.

4.3. Content Update Frequency, Sustainability and Liability. Data entities in professional databases need to be constantly revised, supplemented, and added to present the users with comprehensive and correct information at all times. If the information in databases is not updated in time, it may cause users to make wrong judgments, thereby harming the health of patients. Update frequencies of the reviewed ten databases were given in **Table 1**. Three databases (CMSS, CWMIIN, and DHIQW) have, however, already abandoned continuous update, which might be attributed to the costly manual update of their content. LDI, NaPDI, and SUPP.AI have not updated their content for at least several months. Only NMCD, SHMI, DIDB, and Probot have published recent updates. DIDB, NMCD, and SHMI belong to commercial databases, and annual subscriptions are needed for users to access their data, which could be used to cover the high cost of regular updates. Therefore, user-paid access may facilitate the sustainability of these databases. In addition, SUPP.AI offers an automatic protocol for updating the database once in several months, which may serve as a solution for database developers to provide continuous updates without sufficient resources.

Data provided in these HDI databases should be used with care since some databases did not undergo a strict review process and it is suggested to verify the information with their provided original sources before making a crucial decision. It was also noted that most HDI databases lacked convenient ways for users to point out errors in their databases. It is not expected that all the HDI databases can replace health

professionals to provide HDI information in the near future. Currently, many HDI databases include a disclaimer or warning to indicate that information provided by the databases should not be used as a substitute for advice from healthcare professionals. The disclaimer on Stockley's database states that the publisher is not be responsible for errors and omissions, and it is the responsibility of practitioners to interpret SHMI in light of professional knowledge and relevant circumstances. SUPP.AI has also included a disclaimer to indicate that "the information contained herein should not be used as a substitute for the advice of an appropriately qualified and licensed physician or other health care provider".

4.4. User Interfaces. All databases support keyword-based searches. The keywords that users must provide are typically either the name of the herbs or the drugs or both. Some databases provide alphabetical indexes to facilitate searches (e.g., NMCD and SHMI), while analogous indexes for herbs and drugs in Chinese are available in CWMIIN. Almost none of the databases considered provide a search function that allows query based solely on the characteristics of the HDI without explicitly stating the herbs and drugs involved. An exception appears to be DIDB, in which study results of HDI can be retrieved without necessarily providing drug names.

Recently, there has been a surge in the use of graph databases, which allows searches to be performed based on descriptors of relationships. Potentially, a graph database can enhance the search function and some other technical aspects of a typical relational-type HDI database and remove the limitation of confining keywords to be the names of herbs and drugs in HDIs. A graph database stores data in the form of nodes and relationships, which are fundamental elements of graphs. In the context of HDI, the graph can be a representation based on herbs and drugs as nodes and interactions as relationships. The implication of an HDI graph database is that, if users only have knowledge of the symptoms and are ignorant about herbs and drugs, and these symptoms describe the relationships, i.e., the interactions, users might still be able to make use of the database and discover potential herb–drug pairs that cause the symptoms.

Some databases provide advanced search options to facilitate filtration of undesirable search results. NMCD, for example, allows users to exclude certain fields for a particular query. Both NMCD and SHMI are collections of sub-categorized databases of smaller sizes, and both databases allow users to search only a subset of the collection. Surprisingly, few databases provide the functions to allow users to download a record of the search results in a format, e.g., .CSV, which can be electronically edited with ease. Seemingly only NaPDI, DIDB, and SUPP.AI offer this type of function. SUPP.AI is the only database that provides an API for programmatic access of data. Such a feature is particularly convenient for data analysis that requires customization with user-developed computer programs.

4.5. Role of AI in HDI Database Development. The subset of AI that is particularly relevant to the development of HDI database, which typically requires understanding and organizing a large amount of biomedical text data from a large corpus of scientific articles and reports, is NLP (Rodriguez-Esteban, 2009). With the aid of NLP, the expected outcome is that these text data can be efficiently and accurately classified, extracted, translated, and interpreted, minimizing the manual efforts and time required to develop and maintain these HDI databases. Some of the applications of NLP for developing an HDI database include the recognition of HDI-related articles, the recognition of named entities, e.g., herbs and drugs, the recognition of relationships e.g., the interactions between a pair of herb and drug, and to draw conclusions from a piece (or a large corpus) of text. In the context of HDI, the last task can be interpreted as extracting the conclusion of HDI. This is not the same as drawing a conclusion from several studies for a particular HDI, which is desirable for an HDI database as this would produce certainty with regard to the HDI information offered, but difficult to achieve with confidence even done manually by professionals. Databases such as DrugBank (Wishart et al., 2006) and NMCD would use metrics, such as evidence levels, to represent the reliability of the information on HDI and DDI.

In general, the NLP tasks for HDI data can, in principle, be performed with two types of algorithms, rule-based approaches and statistical models. Rule-based approaches use rules derived from linguistics and/or knowledge of medicines to extract the relevant HDI information. For instance, provided with a dictionary and a set of well-defined rules for the nomenclature of herbs and drugs, matching text strings might suffice the task for Named Entity Recognition (NER), a method that does not require high computational cost. A dictionary-based approach, however, can suffer from problems of term variation. Kang et al. discussed that variations can be eliminated using the Unified Medical Language System (UMLS) (Bodenreider, 2004), and they have derived rules to tackle these problems (Kang et al., 2013). In the same study, researchers have derived rules for other NLP tasks, including coordination, abbreviations, boundary corrections, and filtering, and all the rules are related to the use of a concept normalization system. These rule-based NLP methods were able to improve the recognition of disease from relevant corpus. A linguistic rule-based approach was attempted by Segura-Bedmar et al. to extract DDIs from biomedical text, but in this study, the rules were unable to identify many of the interactions, and the authors suggested that this is due to the variability of natural language expression (Segura-Bedmar et al., 2011). Generally, troubleshooting is convenient for rule-based approaches, and there is often high flexibility to improve the effect of the rules, i.e., the performance of the NLP methods. It might be difficult, however, to develop rule-based systems if the data sources are articles with different writing styles, and there is a large variability with regard to the content of these articles.

Another approach would be the use of statistical models, which are normally trained with large and annotated datasets. Techniques for text classification and NER are well-developed in the field of biomedical text mining. Many common classification methods, such as Naïve Bayes (NB) classifiers and support vector machines (SVMs), that without the need to extensively experiment with these methods, may be sufficient to produce results with tolerable inaccuracies, provided the dataset is sufficiently large and well-curated. The linear SVM classifier in Sum Kim et al.'s work was able to achieve an F1-score of 0.67 based on the DDIExtraction 2013 corpus (Kim et al., 2015). A feature-based approach combined with an SVM classifier tested by Quoc-Chinh Bui et al. were able to achieve accuracy of over 80% for extracting DDIs (Bui et al., 2014).

The research community generally believes that deep learning techniques can improve the performance of AI in many different applications. Zhang et al. reviewed a large number of these techniques for the extraction of DDIs, which are either based on convolutional neural network (CNN), recurrent neural network (RNN), or recursive neural network (Re-NN) (Zhang et al., 2020). They have shown that some of the models, such as attention-based RNN and deep CNN were able to produce high F1 scores of over 80%. These methods, however, may have stability issues, and the dependencies of these models on other factors, such as data volume and quality, could discourage choosing this type of technique to extract DDIs. In general, statistical models are particularly useful in unstructured data, e.g., articles with different writing styles and

contents, but difficult for troubleshooting, and require large amounts of data that can be difficult to prepare and collect.

So far, the most representative example of the application of AI for an HDI database is SUPP.AI. The data collected were pre-processed, including entity recognition and entity linking using the ScispaCy library (Neumann et al., 2019), and the generation and clustering of CUIs based on the UMLS Metathesaurus. This step prepares sentences from a large number of abstracts for further classification. The goal of the subsequent classification is to determine for a sentence if an interaction exists or not. SUPP.AI in this respect uses a Bidirectional Encoder Representations from Transformers (BERT) model, RoBERTa (Liu et al., 2019), that was fine-tuned by pre-trained embedding for DDI classification. The model for identifying supplement–drug interactions (SDIs) is based on the presumption that the model trained for identifying DDI can be transferred to identify SDIs. The relatively sophisticated models in SUPP.AI (Precision=0.82, Recall=0.58, F1-score=0.68) have seemingly outperformed some classic classification methods, such as SVM for the extraction of DDI, but when tested for supplements the performance is not as promising.

It can be concluded that the main advantage of AI is its possibility to process a large amount of data so as to achieve a broader coverage of HDI information. However, a number of limitations for AI application in database development remain. First, there is a strong need to improve the accuracy of methods for performing various NLP tasks. In addition, many studies have explored the performance of NER methods on dataset for DDIs instead of HDIs. To apply such method to extract HDI information, a branch of AI known as transfer learning could serve as a potential solution. Moreover, since the development and improvement of both rule-based and probabilistic methods could be time-consuming and cumbersome, developers of HDI databases are therefore advised to plan ahead and perform extensive testing of methods before implementations.

4.6. Further Development of HDI Database. As we mentioned in the introduction, the third step, extracting and evaluating the buried information in literature to obtain structured data, is the rate-limiting step in HDI database development. The application of AI seems to achieve promising results in recognizing the names of herbs/drugs and in text classification (e.g., extracting SDI sentences in SUPP.AI and screening HDI-related abstracts in Probot) and could significantly reduce the manual efforts for database maintenance. However, application of AI in other specific tasks, such as recognizing experimental models, conditions, and parameters, would be more challenging due to the lack of annotated training datasets for these tasks. The lack of such datasets is due to not only the huge amount of work required but also the shortage of published HDI reports so far. According to our searches on the HDI databases shown in **Table 1**, except for SUPP.AI (due to unavailable amount of HDI references), the number of published HDI reports is estimated to be less than 5,000. Even though all the published HDI articles are manually annotated, the resulting datasets may not be sufficient for training NER models for recognizing all kinds of information from literature. Moreover, the annotated datasets are task-specific, and different datasets are needed for different training purposes. Thus, the role of AI in obtaining structured HDI-related data are limited and manual curation is still not dispensable in the near future.

On the other hand, to avoid the difficulty of extracting structured data from HDI literatures, it is suggested that the authors publish their experiments and results in a predefined format, such as those proposed in NaPDI. The data in NaPDI are organized according to "Study" and "Experiment". One "Study" consists of one or more related "Experiments". Each "Experiment" belongs to one of the eleven experiment types with a standard operating procedure (SOPs) for entering data into NaPDI. Apart from this standardized format to reposit research data, we also suggest the authors provide CUIs for related medical terms based on the Unified Medical Language System (UMLS). A CUI contains the letter C followed by seven numbers and the key goal of it is to link different names for herbs, drugs, and experimental models with the same meaning and improve the accuracy in scientific expression. Overall, the sharing of structured data by authors would eliminate the need to develop NLP models as well as the effort of manual curation and thus accelerate the process of HDI database development.

Acknowledgments

The authors would like to thank Drs. Jingjing Yu and Isabelle Ragueneau-Majlessi from University of Washington Drug Interaction Solutions for providing relevant information about UW DIDB.

Note Added in Proof: An unrelated supplemental Figure 1 was accidentally added to the Fast Forward version of the article published October 25, 2021. The supplementary figure has been removed.

Authorship Contributions

Participated in research design: Zhang, Zuo.
Performed data analysis: Zhang, Ip, Lai.
Contributed to the writing of the manuscript: Zhang, Ip, Zuo.

References

Alostad AH, Steinke DT, and Schafheutle EI (2020) Herbal Medicine Classification: Policy Recommendations. *Front Med (Lausanne)* **7**:31.
Bailey D (2011) Stockley's Herbal Medicines Interactions. *Br J Clin Pharmacol* **71**:143.
Birer-Williams C, Gufford BT, Chou E, Alilio M, VanAlstine S, Morley RE, McCune JS, Paine MF, and Boyce RD (2020) A New Data Repository for Pharmacokinetic Natural Product–Drug Interactions: From Chemical Characterization to Clinical Studies. *Drug Metab Dispos* **48**:1104–1112.
Bodenreider O (2004) The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* **32**:D267–D270.
Bui QC, Sloot PMA, van Mulligen EM, and Kors JA (2014) A novel feature-based approach to extract drug–drug interactions from biomedical text. *Bioinformatics* **30**:3365–3371.
Cheng TO (2006) Herb digoxin interactions. *Int J Cardiol* **110**:93.
Chimei (2004) Chimei Herb–Drug Interaction Search System. *Department of Pharmacy, Chi Mei Medical Center,* Available from: http://www.chimei.org.tw/main/cmh_department/55500/DIS/cdi_search.asp [accessed July 9, 2021]
Fugh-Berman A and Ernst E (2001) Herb–drug interactions: review and assessment of report reliability. *Br J Clin Pharmacol* **52**:587–595.
Ge B, Zhang Z, and Zuo Z (2014) Updates on the clinical evidenced herb-warfarin interactions. *Evid Based Complement Alternat Med* **2014**:957362.
Glisson JK and Walker LA (2010) How physicians should evaluate dietary supplements. *Am J Med* **123**:577–582.
Hachad H, Ragueneau-Majlessi I, and Levy RH (2010) A useful tool for drug interaction evaluation: the University of Washington Metabolism and Transport Drug Interaction Database. *Hum Genomics* **5**:61–72.
Hu Z, Yang X, Ho PC, Chan SY, Heng PW, Chan E, Duan W, Koh HL, and Zhou S (2005) Herb-drug interactions: a literature review. *Drugs* **65**:1239–1282.
Izzo AA and Ernst E (2001) Interactions between herbal medicines and prescribed drugs: a systematic review. *Drugs* **61**:2163–2175.
Kang N, Singh B, Afzal Z, van Mulligen EM, and Kors JA (2013) Using rule-based natural language processing to improve disease normalization in biomedical text. *J Am Med Inform Assoc* **20**:876–881.
Kim S, Liu H, Yeganova L, and Wilbur WJ (2015) Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach. *J Biomed Inform* **55**:23–30.
Kluwer (2018) Lexicomp® Drug Interactions. *Wolters Kluwer N.V.,* Available from: https://new.uptomed.ir/Main/d/di.htm [accessed July 8, 2021]
Lin SS(2011) The Chinese-Western medicine integrative information network. *China Medical University,* Available from: https://dhi.cmu.edu.tw/query/ [accessed July 9, 2021]
Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, and Stoyanov V (2019) RoBERTa: A Robustly Optimized BERT Pretraining Approach.
Neumann M, King D, Beltagy I, and Ammar W(2019) ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing. *Proceedings of the 18th BioNLP Workshop and Shared Task*.
Rice JO (2014) *Stockley's Herbal Medicines Interactions: A Guide to the Interactions of Herbal Medicines.* Second edition. *J Med Libr Assoc: JMLA* **102**:221-222.
Rodriguez-Esteban R (2009) Biomedical text mining and its applications. *PLOS Comput Biol* **5**:e1000597.
Segura-Bedmar I, Martínez P, and de Pablo-Sánchez C (2011) A linguistic rule-based approach to extract drug–drug interactions from pharmacological documents. *BMC Bioinformatics* **12** (Suppl 2):S1.
Trovato M and Ballabio C (2018) Botanical Products: General Aspects, in *Food Supplements Containing Botanicals: Benefits, Side Effects and Regulatory Aspects* (Restani P, ed Springer, Cham.

UW (2021) UW Drug Interaction Database (DIDB). *University of Washington,* Available from: https://els2.comotion.uw.edu/product/uw-drug-interaction-database [accessed July 9, 2021]

Vardell E (2015) Natural Medicines: A Complementary and Alternative Medicines Tool Combining Natural Standard and the Natural Medicines Comprehensive Database. *Med Ref Serv Q* **34**:461–470.

Wang L, Tafjord O, Cohan A, Jain S, Skjonsberg S, Schoenick C, Botner N, and Ammar W(2020) SUPP.AI: finding evidence for supplement-drug interactions. 362-371.

Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, and Woolsey J (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* **34**:D668–D672.

Wu CS, Chen YH, Chen CL, Chien SK, Syifa N, Hung YC, Cheng KJ, Hu SC, Lo PT, Lin SY et al. (2019) Constructing a bilingual website with validated database for Herb and Western medicine interactions using Ginseng, Ginkgo and Dong Quai as examples. *BMC Complement Altern Med* **19**:335.

Yacobucci KL (2016) Natural Medicines. *J Med Libr Assoc* **104**:371–374.

Zhang T, Leng J, and Liu Y (2020) Deep learning for drug-drug interaction extraction from the literature: a review. *Brief Bioinform* **21**:1609–1627.

Zhang YF and Zuo Z (2021) Probot Chinese Medicine-Drug Interaction Database. *Healthy Power Limited,* Available from: http://www.probot.hk [accessed July 9, 2021]

Address correspondence to: Prof. Zhong Zuo, School of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR. E-mail: joanzuo@cuhk.edu.hk