

# Similarity and Consistency Assessment of Three Major Online Drug-Drug Interaction Resources

Elpida Kontsioti <sup>a,b</sup>, Simon Maskell <sup>a</sup>, Amina Bensalem <sup>c</sup>, Bhaskar Dutta <sup>d</sup>, Munir Pirmohamed <sup>e</sup>

## Affiliations

<sup>a</sup> Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, United Kingdom

<sup>b</sup> Institute for Risk and Uncertainty, University of Liverpool, Liverpool, United Kingdom

<sup>c</sup> R&D, Ceva Santé Animale, Libourne, France

<sup>d</sup> Patient Safety Center of Excellence, AstraZeneca, Gaithersburg, MD, United States

<sup>e</sup> The Wolfson Centre for Personalized Medicine, MRC Centre for Drug Safety Science, Department of Pharmacology and Therapeutics, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, United Kingdom

## Corresponding authors and contact information:

- Elpida Kontsioti, Department of Electrical Engineering and Electronics & Institute for Risk and Uncertainty, University of Liverpool, Peach Street, Chadwick Building, Liverpool, L69 7ZF, UK
- Prof Sir Munir Pirmohamed, Wolfson Centre for Personalised Medicine, Department of Molecular and Clinical Pharmacology, University of Liverpool, Block A: Waterhouse Buildings, 1-5 Brownlow Street, Liverpool, L69 3GL, UK.

**Running title:** Similarity of DDI online resources

**Keywords:** drug-drug interaction, drug information, drug-drug interaction software, clinical decision support, clinical management of drug interactions

Word count: 4,748

Table count: 4

Figure count: 4

## Summary

### AIM

To explore the level of agreement on drug-drug interaction (DDI) information listed in three major online drug information resources (DIRs) in terms of: (1) inclusion of interacting drug pairs; (2) severity rating; (3) evidence rating and (4) clinical management recommendations.

### METHODS

We extracted information related to DDIs from three online DIRs: the *British National Formulary* (BNF), *Thesaurus* and *Micromedex*. Following drug name normalisation, we calculated the number of common drug pairs amongst the three DIRs and compared severity and evidence ratings across the DIRs. We also assessed the agreement on clinical management recommendations, either using manual annotation (for *BNF* and *Thesaurus*) or, due to the large volume of descriptions, through application of a machine learning algorithm that facilitated annotation (for *Micromedex*).

### RESULTS

The *BNF* contained 51,481 unique drug pairs involved in DDIs, while *Thesaurus* and *Micromedex* contained 38,037 and 65,446 unique drug pairs, respectively. The number of drug pairs that were common to all three DIRs was 6,970 (13.54% of *BNF*, 18.32% of *Thesaurus* and 10.65% of *Micromedex*). *Micromedex* and *Thesaurus* overall showed higher levels of similarity in their severity ratings, while the *BNF* agreed more with *Micromedex* on the critical severity ratings and with *Thesaurus* on the least significant ones. Evidence rating agreement between *BNF* and *Micromedex* was generally poor. Variation in clinical management recommendations was also identified, with some categories (i.e. *Monitor* and *Adjust dose*) showing higher levels of agreement among DIRs compared to others (i.e. *Use with caution*, *Wash-out*, *Modify administration*).

### CONCLUSIONS

There is considerable variation in the drug pairs included in the examined DIRs, together with variability in categorisation of severity and clinical advice given. DDIs labelled as critical are more likely to appear in multiple DIRs. Such variability in information could have deleterious consequences for patient safety, and there is a need for harmonisation and standardisation.

### WHAT IS ALREADY KNOWN ABOUT THIS SUBJECT

- There are a variety of online DIRs, which differ in coverage, content, and inclusion criteria, that are available to clinicians and other prescribers, mainly for prescribing decision support purposes.
- Previous studies have described major discrepancies between widely used DIRs on inclusion of critical DDIs or interactions of specific therapeutic categories, along with discordance in their severity and evidence ratings.

### WHAT THIS STUDY ADDS

- To the best of our knowledge, this is the first study to concurrently compare the similarity among complete datasets from DIRs in terms of inclusion of drug pairs, recommendations for clinical management, severity, and evidence of DDIs.
- Considerable variation was identified in all types of information, which has important clinical implications for patient safety and requires efforts towards harmonisation and standardisation.

# 1 Introduction

Coadministration of multiple drugs increases the risk of drug-drug interactions (DDIs). A DDI can be defined as the modification in the therapeutic effect of one or more medications due to the presence of concomitant medications, and can lead to clinically significant events, caused by either an increase in the effect of the interacting drug leading to an adverse drug reaction (ADR), or a decrease in its effect that results in lack of efficacy. Previous studies have reported that DDIs seemed to be a significant cause of hospitalisation, being responsible for 16.6% of cases where the cause was an ADR and around 1% of all hospital admissions.<sup>1</sup> The risk for DDIs increases during hospitalisation and after discharge, as there is a high prevalence of administration of potentially interacting drug combinations.<sup>2</sup>

As new medicines gain approval each year, the volume of possible drug combinations is constantly growing. At the same time, the rising numbers of people with multimorbidity together with increasing life expectancy around the world are associated with the phenomenon of polypharmacy, which aggravates the impact of DDIs in clinical practice. According to a recent review, over 1 out of 3 people in England aged 60 and older are exposed to at least 5 medicines at the same time, with more than a third of all people above 80 being on 8 or more medicines.<sup>3</sup>

The clinical manifestation of DDIs depends on several factors. Potential DDIs, based on pharmacological knowledge, far outnumber those which lead to clinically significant adverse effect results.<sup>4</sup> Despite the theoretical potential for an ADR to occur due to a DDI, there are several factors that can affect the actual behaviour of drug molecules inside the human body, including dosage and patient characteristics (e.g. age, number and type of morbidities, etc). Also, genetic polymorphisms of drug-metabolising enzymes, drug transporters or drug receptors may be responsible for the appearance of some DDIs.<sup>5</sup> Therefore, it is especially difficult to accurately predict the occurrence of a clinically significant DDI in an individual patient. To overcome this problem, clinicians are commonly aided by drug information resources (DIRs) to assess the risk-benefit ratio of each drug added to the treatment schedule. DIRs can be either open source or commercial, and they are often incorporated in computerised clinical decision support (CDS) tools.

The availability of DIR information related to severity, evidence availability, and clinical options for management of DDIs (e.g. entirely avoid the combination, monitor, adjust dose, etc) are central to the development of CDS.<sup>6</sup> Inconsistencies between DIRs may confuse clinicians and impact clinical decisions.<sup>7</sup> Previous studies have assessed the level of agreement of DIRs, mainly in terms of listing of DDIs and severity ratings. However, most of them were restricted to only DDI listing for a limited number of drugs, specific therapeutic categories, or did not focus solely on clinical resources<sup>8–10</sup>. Moreover, the ability of a DIR to identify clinically relevant DDIs or capture critical DDIs (e.g., FDA black box warnings, ONC high priority list<sup>11</sup>) has also been explored.<sup>12–14</sup> However, it remains unclear to what extent DIRs from different geographic locations agree on their DDI listings as well as DDI-related information.

The aim of this study was to assess the concordance of leading clinical resources for DDIs from three different countries of origin in terms of: (1) inclusion of interacting drug pairs; (2) severity rating; (3) evidence rating and (4) clinical management recommendations. To the best of our knowledge, this is the first comprehensive study that attempts to compare multiple types of information pertinent to DDIs at the same time across entire DIRs. To ensure clinical utility, only clinically relevant resources were included in the present study (see Supplementary **Figure S1** for an overview of online DDI resources). Data sources of potential DDIs (e.g. DrugBank) that are mainly used for scientific research purposes were not taken into consideration.

## 2 Methods

### 2.1 Data sources

Two open-source and one commercial online DIRs were included in our evaluation: the British National Formulary Interactions (hereafter called *BNF*), Interactions Thesaurus by the French Medicines Agency (Agence Nationale de Sécurité du Médicament et des produits de santé, ANSM) (hereafter called *Thesaurus*) and IBM Micromedex (hereafter called *Micromedex*). The *BNF*<sup>15</sup> is extensively used in the UK. *Thesaurus*<sup>16</sup> is maintained and updated annually by ANSM, being considered as the official source of information relevant to DDIs for French clinicians.

*Micromedex*<sup>17</sup> is a leading clinical information resource in the United States (US). The *BNF* and *Thesaurus* are publicly available online, while *Micromedex* can only be accessed via subscription.

## 2.2 Data extraction

Automated web data collection (*web scraping*) was executed for *BNF* and *Micromedex* in Python 3.6<sup>18</sup> with terms of use that permit data collection. *Thesaurus* is a Portable Document Format (PDF) file that is curated and updated annually. An R package (*IMthesaurusANSM*<sup>19</sup>) enabled the automatic data extraction from the original document (version September 2019). The types of extracted information from each DIR are summarised in **Table 1**.

We mapped DDIs from *Thesaurus* at the drug class level (e.g. beta blockers) to their constituent individual drug ingredients using a mapping table available in the ANSM website. We also excluded DDIs from *Micromedex* containing drug combinations (e.g. hydroxyamphetamine/tropicamide), as those simply displayed the union of DDIs from the combination's individual ingredients; hence, only single ingredient drug interactions were considered. Also, cases where drug names of an interacting pair were swapped (i.e. (*D1*, *D2*) and (*D2*, *D1*)) were considered equivalent and duplicate entries were removed from the tables that stored the extracted data (*BNF original table*, *Thesaurus original table* and *Micromedex original table*).

## 2.3 Drug name normalisation

Initial drug names were normalised to RxNorm Ingredients (for US-marketed medicines)<sup>20</sup> and RxNorm Extension Ingredients (for medicines not found in RxNorm)<sup>21</sup> using the OHDSI<sup>i</sup> Usagi tool.<sup>22</sup> Some names were too general to be mapped (e.g. *insulins*), or were not present in either vocabulary. Thus, interacting pairs containing at least one unmapped drug were excluded from the corresponding DIR table. As the scope of this study was limited to DDIs, only interacting pairs containing exclusively drugs were included in the final DIR tables and interactions with herbs, alcohol, food, etc were excluded. The final tables (*BNF final table*, *Thesaurus final table* and *Micromedex final table*) contained drug interacting pairs and associated information based on normalised drug names. Any duplicate entries based on common normalised names were combined into a single entry. For example, *Metoprolol Tartrate* and *Metoprolol Succinate* were both mapped to the RxNorm entity *Metoprolol*, and their interactions were merged to produce a single set.

## 2.4 Comparison of resources

### Listing of DDIs

The pairwise and three-way overlaps of the final DIR tables were estimated by calculating counts of common drug pairs across the DIRs as well as coverage rates (i.e. the percentage of a set A covered by B, where B is a subset of A). The directionality of interacting drug pairs was not taken into account (i.e. (*D1*, *D2*) and (*D2*, *D1*) were considered equivalent). A table (*DDI intersection table*) containing common interacting drug pairs among all three DIRs with their corresponding text descriptions from each source was generated.

### Severity and evidence ratings

All three DIRs included severity ratings (**Table 2a**), while only the *BNF* and *Micromedex* contained separate text fields regarding evidence ratings (**Table 2b**). Some DDIs from *Thesaurus* appeared at the drug class level in the original source, which was associated to multiple severity ratings; thus, individual drugs were assigned all applicable ratings from the drug class during the mapping process. Also, some DDIs were linked to multiple severity ratings, based on the clinical circumstances (e.g. route of administration, dose, etc). In all cases where multiple ratings were available for an individual DDI, the highest one was kept for further analysis.

To explore discrepancies among DIRs related to severity and evidence ratings, we calculated the subset size,

---

<sup>i</sup> OHDSI: Observational Health Data Sciences and Informatics

pairwise coverage rates and Jaccard indices for all possible pairs of DIR ratings.

## Clinical management recommendations

We aimed to explore the consistency among the clinical management recommendations provided by the DIRs by analysing text descriptions from the *DDI intersection table*. The *BNF* provided a succinct description for each drug pair containing all types of available DDI information in a text field, while *Thesaurus* and *Micromedex* contained separate text fields (*Conduit à tenir* and *Clinical Management*, respectively) under each drug pair related to clinical management options.

Basic pre-processing involved text conversion to lowercase, drug name blinding (i.e. replacement of all drug names with a common string), and sentence tokenisation using the Natural Language Toolkit (NLTK) in *Python 3.6*.<sup>23</sup>

The following advice categories were initially considered:

- i. Avoid;
- ii. Use with caution;
- iii. Space dosing times;
- iv. Wash-out;
- v. Monitor;
- vi. Adjust dose;
- vii. Modify administration;
- viii. Use alternative;
- ix. Discontinue.

Cases that recommended clinicians to refer to literature or other resources, without mentioning any concrete clinical advice, were excluded.

The limited number of unique sentences sourced from *BNF* ( $N = 305$ ) and *Thesaurus* ( $N = 387$ ) following drug name blinding enabled manual sentence labelling, with each sentence being classified into one or multiple advice categories.

To annotate *Clinical Management* text descriptions in *Micromedex* ( $N = 4,507$ ), we developed a bespoke text classification process in *Python* using a methodology that has been widely implemented in similar tasks and provided the desired functionality while keeping the level of complexity low (Supplementary **Figure S2**). First, we annotated a subset of randomly selected unique sentences ( $N = 200$ ) by considering the above-mentioned categories. Then, each labelled sentence was tokenised into its constituent tokens (i.e. words) and stemming (i.e. reducing words to their word roots) was applied. We used *term frequency-inverse document frequency* (*tf-idf*) to calculate weights for each word in the annotated sentences. The goal of *tf-idf* is to reduce the impact of very commonly occurring words in a corpus, assuming that they are less informative. Term frequencies are calculated by counting the relative frequency of each word appearing in each of the annotated sentences. Inverse document frequencies of each word (in its root form) are estimates of the overall presence of the word across all sentences (i.e. how commonly or rarely it appears). The formula for calculation of a word's *tf-idf* is the following:

$$tf-idf_{w,s} = tf_{w,s} \times \left( \log \frac{N}{df_w} + 1 \right) \quad (1)$$

where  $tf_{w,s}$  represents the term frequency of the word,  $w$ , in the sentence,  $s$  (i.e. the number of times the word appears in the sentence divided by the total number of words in the sentence);  $N$  is the total number of sentences in the corpus and,  $df_w$  is the “document” frequency of the word,  $w$  (i.e. the number of sentences that contain the specific word). Weights were applied for sentence encoding to feed classifiers that used a supervised machine learning model called *linear kernel Support Vector Machine* (SVM) for binary text classification (i.e. each sentence was classified as to whether it belongs to each of the advice categories under consideration). We applied class

weights to account for the imbalanced training sets (i.e. disproportion between the number of positive and negative instances) and used leave-one-out cross validation to evaluate performance of the difference classifiers through Receiver Operating Characteristic (ROC) analysis. By estimating the positive predictive value (PPV) for the different thresholds, we excluded sentence classifiers with a PPV below 80% due to poor performance; the remaining, unannotated sentences from Micromedex were automatically labelled by the classifiers using the threshold with maximum sensitivity for PPVs above 80%. A subset ( $N = 100$ ) of the automatically annotated sentences (*validation set*) was also manually annotated to independently estimate the classifiers' performance in the total set of Micromedex sentences.

## 3 Results

### 3.1 Comparative assessment in terms of listing

*Micromedex* contained the largest number of DDI drug pairs ( $N = 65,446$ ), as well as normalised ingredients ( $N = 1,967$ ), followed by *BNF* ( $N = 51,481$ ) and *Thesaurus* ( $N = 38,037$ ) that covered 984 and 1,001 normalised ingredients, respectively. The union of the three DIR final tables (hereafter labelled *DIR union*) included 121,351 DDI drug pairs. The counts of initial drug names, normalised drug ingredients, and unique DDIs in each DIR are summarised in **Table 3**.

There were 690 common normalised ingredients across all examined DIRs, with *BNF* and *Micromedex* sharing the largest number ( $N = 906$ ), followed by *Thesaurus* and *Micromedex* ( $N = 894$ ) and, lastly, the *BNF* and *Thesaurus* ( $N = 716$ ) (**Figure 1a**). Almost four out of five DDI drug pairs (78.04%,  $N = 94,708$ ) in the *DIR union* were only mentioned by a single DIR, with 57.19% of *BNF*, 49.58% of *Thesaurus* and 70.91% of *Micromedex* DDI entries missing from the other two DIRs. The percentage of DDIs mentioned in exactly two out of three DIRs was considerably low (16.21%,  $N = 19,673$ ). *BNF* shared 14,576 DDIs with *Thesaurus* (28.31% of *BNF*; 38.32% of *Thesaurus*) and 14,433 DDIs with *Micromedex* (28.04% of *BNF*; 22.05% of *Micromedex*), while *Thesaurus* and *Micromedex* had 11,574 common DDIs (30.43% of *Thesaurus*; 17.68% of *Micromedex*). The intersection of the three DIRs ( $N = 6,970$ ) represented 5.74% of the *DIR union*, 13.54% of *BNF*, 18.32% of *Thesaurus* and 10.65% of *Micromedex* (**Figure 1b**).

In terms of DDIs restricted to common ingredients across the three DIRs ( $N = 44,719$ ), more than half ( $N = 24,951$ ) were only found in a single DIR (33.74% in *BNF* alone, in comparison to 12.86% and 9.20% in *Micromedex* and *Thesaurus*, respectively), while 28.62% were present in two out of three DIRs. In the setting of ingredient-restricted DDIs, the *BNF* intersected with large proportions of both *Thesaurus* (71.40%) and *Micromedex* (59.76%), while less than half of *BNF* (42.81%) overlapped with *Thesaurus*. Finally, the intersection of the three DIRs represented 15.59% of the restricted DDIs (**Figure 1c**).

### 3.2 Comparative assessment of severity rating

The categorisation of DDIs in each DIR in terms of severity rating is outlined in **Table 4**. Regarding critical severity rating categories, almost one quarter (24.56%) of unique DDIs in *BNF* were labelled as *Severe*, compared to 7.75% from *Thesaurus* characterised as *Contraindicated*, and 33.60% as *Not recommended*. In *Micromedex*, 8.76% of unique DDIs were mentioned as *Contraindicated*, while *Major* was the most frequent category (63.73%).

When considering the pairwise DIR overlap using coverage rates (**Figure 2**), the number of DDIs jointly rated as critical was:

- 2,429 between *BNF* and *Thesaurus*, representing 19.21% of the *BNF* and 15.44% of *Thesaurus* critical DDIs;
- 6,026 between the *BNF* and *Micromedex* (47.66% of *BNF* and 31.38% of *Micromedex* critical DDIs);
- 5,014 between *Thesaurus* and *Micromedex*, covering 78.39% of *Thesaurus* and 26.33% of *Micromedex* critical DDIs);
- 1,768 among all three DIRs (25.37% of the *DDI intersection table*).

The percentage of DDIs from the *DDI intersection table* that were considered critical by *BNF*, *Thesaurus* and *Micromedex* was 43.39% (N = 3,024), 52.32% (N = 3,647), and 81.51% (N = 5,681), respectively.

A similarity matrix of the Jaccard index for all DIR severity rating combinations is included in the Supplementary Material (**Supplementary Figure S3**).

### 3.3 Comparative assessment of evidence rating

The *BNF* included evidence ratings in just around one third (30.66%) of its DDIs, with the majority being flagged as *Theoretical* (16.99%), followed by *Study* (11.89%), and *Anecdotal* (1.78%). In *Micromedex*, evidence ratings were consistently present under each DDI description. *Theoretical* was the most common category (70.91%), while *Probable* and *Established* included 20.36% and 8.73% of the DDIs mentioned in *Micromedex*.

Almost half (48.13%) of the DDIs from the *DDI intersection table* contained no evidence rating in *BNF*; the remainder belonged to *Study* (28.05%), *Theoretical* (19.94%) and, lastly, *Anecdotal* (3.87%) evidence categories. According to *Micromedex*, most of them (64.51%) were *Theoretical*, with 19.94% and 15.55% being considered as *Probable* and *Established*, respectively.

**Figure 3** shows the overlap of the different evidence categories between the two DIRs as a two-by-two grid with subset counts and coverage rates. Probable DDIs from *Micromedex* and DDIs with no evidence rating from *BNF* were absent in higher percentages in the other resource. In both DIRs, the percentage of missing DDIs increased as one moved towards DDIs with a “poorer” or no evidence rating in the other resource. Using the Jaccard index, agreement between ratings was generally low in all cases, with the *BNF Study* and *Micromedex Theoretical* categories being the most similar (0.04662), while the *BNF Anecdotal* and *Micromedex Established* had the lowest concordance (0.00573).

### 3.4 Comparative assessment of clinical management advice

In the *BNF*, no instances of the *Discontinue* advice category were identified in the *DDI intersection table*, while in *Micromedex*, counts for seven out of the nine advice categories are provided, as no sentence classifier was applied to extrapolate the remaining labels (i.e. *Space dosing times* and *Modify administration*) due to poor classifier performance (see **Supplementary Tables S1 and S2** for associated metrics). The subset of *Micromedex* descriptions associated with *BNF* cases that belonged to either of those two advice categories were manually annotated as a surrogate measure of concordance.

The classification of *DDI intersection table* entries in each DIR in terms of clinical management advice is shown in **Supplementary Table S3**. In *BNF*, no advice was available in over half (56.41%) of the DDIs under consideration. The most common advice category was *Avoid* (32.12%) and the least frequently mentioned was *Use alternative* (0.01%). In *Thesaurus*, *Monitor* and *Avoid* jointly covered more than half of the total DDIs (34.89% and 30.82%, respectively), while recommendations related to *Space dosing times*, *Use with caution* and *Wash-out* were only found in small percentages (1.98%, 1.15%, and 0.99%, respectively). In *Micromedex*, the labelling process that was facilitated by sentence classifiers provided the following results: 63.43% of the DDIs were characterised as containing an advice related to *Monitor*, 47.02% related to *Avoid*, and 35.77% related to *Adjust dose*; low percentages represented *Use alternative* (3.79%), *Discontinue* (2.74%), and *Wash-out* (0.39%) categories. In 5.38% of the *Micromedex* DDIs, no advice label was assigned.

The DIR overlap in terms of the DDI-related advice labels that were considered in this study for the DIR intersection is illustrated using Venn diagrams (**Figure 4**). The *BNF* and *Thesaurus* did not share any DDIs in their *Modify Administration* and *Use with caution* categories, as opposed to the DDIs found in their *Space dosing times* and *Adjust dose* categories that showed extensive overlap. *Thesaurus* and *Micromedex* did not have any common DDIs classified into their *Wash-out* advice categories. Also, for *Wash-out* and *Use with caution* advice categories, there was little agreement between any two DIRs. The three DIRs overlapped to a high degree in the *Monitor* category. In the majority of *Space dosing times BNF* cases (87.96%), *Micromedex* also contained the respective advice. For *Modify administration*, *Micromedex* included this advice for less than half (43.18%) of the *BNF* cases.

## 4 Discussion

This study reports on the consistency of DDI-related information included in three major clinical DIRs from different geographic locations, namely the *British National Formulary (BNF)*, *Thesaurus* and *Micromedex*. The DIRs differed in size and number of ingredients mentioned. The number of ingredients in *Micromedex* was almost twice that found in the other two DIRs. This is most likely to have been due to the fact that the *BNF* and *Thesaurus* only include medicines licensed in their countries of origin (i.e. UK and France, respectively), while *Micromedex* includes a broader set of medicines. Although DIR ingredients overlapped to a significant extent, especially between *BNF* and *Thesaurus*, this overlap was not reflected on the DDI sets, which generally showed poor agreement. The *BNF* and *Thesaurus* shared the largest number of DDIs, in contrast with *Thesaurus* and *Micromedex*, which had the fewest DDIs in common.

Our study represents the most comprehensive assessment of the overlap in content and advice provided by different DIRs. However, our findings are consistent with previous studies. For instance, a study that analysed DDIs of fewer than 100 medicines reported less than 7% overall agreement among the examined sources.<sup>8,9</sup> A more recent analysis that compared three commercial DIRs in terms of listing and severity ranking of DDIs also identified very poor overlap (5%), although DDIs flagged as minor were not considered.<sup>14</sup>

Severity ratings were not consistently reported in the *BNF*, as opposed to *Thesaurus* and *Micromedex* severity ratings that were available in all cases. DDIs labelled as critical comprised approximately one fourth of *BNF* and more than 70% of *Micromedex*, in contrast to *Thesaurus*, where the least significant category was the most populous. *Micromedex* and *Thesaurus* showed similarity in their ratings across the different levels of severity. Between *BNF* and *Thesaurus*, their least significant categories (i.e. *Unknown* and *Take into consideration*, respectively) appeared to have a higher level of agreement. However, there were weaker signs of concordance between *BNF* and *Thesaurus* on the classifications of DDIs in the high severity ratings, with the DDIs classified as *Severe* in *BNF* being spread across the different *Thesaurus* categories. In terms of *BNF* and *Micromedex*, there was generally better agreement between their critical ratings than with the less severe ones. Apart from the *BNF-Thesaurus* pair, the percentage of DDIs missing from a DIR increased as one moved to DDIs characterised as less severe by another DIR (i.e. increasing trend in the *Not found* column percentages as we go from top to bottom in tables from **Figure 2**). *Micromedex* categorised the largest proportion from the *DDI intersection table* as being critical compared to the other DIRs, while around one fourth of *DDI intersection table* DDIs were labelled as critical simultaneously by all three DIRs. Also, the pairwise intersections of DIRs covered larger proportions of the DDIs from the critical severity levels compared to the corresponding proportions from lower severity categories.

Although early studies concluded that significant discrepancies exist in severity ratings between DIRs, Fung et al.'s study advocated the presence of higher levels of agreement than previously reported, especially for the most severe DDIs.<sup>8,9,14</sup> Our results, suggesting better agreement between critical severity ratings between *BNF* and *Micromedex*, are partially in line with this observation.

Evidence categorisation was not available in *Thesaurus*, thus preventing a comprehensive assessment of the concordance of evidence rating amongst all the DIRs. In *BNF*, evidence ratings were available for around one third of the DDIs, while they were consistently reported in *Micromedex*. In both DIRs, *Theoretical* was the most frequent category. However, the study revealed a lack of consistency between *BNF* and *Micromedex*, with no cases of strong agreement between any pairs of evidence ratings. An interesting observation was related to the DDIs included in one but missing from the other DIR, as the percentage of *Not found* DDIs in both cases increased as the evidence rating in the other DIR decreased. A study by Vitry et al also performed a similarity assessment of evidence ratings, also highlighting inconsistencies in the grading system for evidence among the different sources.<sup>8</sup> In terms of clinical management recommendations, there was significant disagreement among the DIRs related to some types of advice, such as *Use with caution*, *Wash-out*, *Discontinue* and *Modify administration*. Other types (i.e. *Avoid* and *Use alternative*) showed a moderate level of agreement, while *Space dosing time*, *Monitor* and *Adjust dose* demonstrated higher levels of concordance.

### 4.1 Strengths and limitations

As opposed to multiple previous efforts to assess the level of agreement among DIRs in terms of DDI information, this study examined entire resources, thus revealing the relative size of information in each of the DIRs and exploring the stratification of the included DDIs in terms of severity, evidence, and clinical management recommendations. To the best of our knowledge, this is the first comprehensive effort to compare clinical advice for managing DDIs that is provided in multiple DIRs, with a clear focus on clinically oriented sources compared to previous work.<sup>10</sup> While a previous study expanded the comparison of DIRs at multiple levels (i.e. clinical drug, ingredient and drug class)<sup>24</sup>, our analysis was limited at the ingredient level. This standardisation of DIRs enabled a “fair” comparison in terms of the volume of information listed. Also, code availability for data extraction and standardisation will enable reproducibility of the analysis.

However, there are limitations to this study. First, no updates have been taken into consideration since the date of data retrieval (i.e., offline data). Therefore, the results and conclusions of this study provide an overview of their similarity and consistency at that specific point in time, although no major updates usually occur. Second, *Thesaurus* contained a few DDIs originally reported at the drug class level, which were associated with multiple severity ratings. Hence, some standardised DDIs at the ingredient level were assigned more than one severity ratings. In *BNF*, there was a limited number of DDIs having multiple severity levels depending on the described clinical outcome. In both cases, the highest severity rating was considered for further analysis. Moreover, combination drugs were not taken into consideration. We expect this omission to have no actual impact on overall quality of our study, as DDI information about the various constituent drugs were captured separately in our analysis. Other limitations include the comparison of clinical management options only for the DDIs belonging in the intersection of the DIRs and the custom-made labelling process applied to *Micromedex*. Additionally, in *BNF*, there were referrals to guidance section of the website for various drug categories that were left unmapped during the annotation process, e.g. “For FSRH guidance, see contraceptives, interactions.”, or “See ‘serotonin syndrome’ and ‘monoamine-oxidase inhibitor’ under antidepressant drugs for more information and for specific advice on avoiding monoamine-oxidase inhibitors during and after administration of other serotonergic drugs.”. In this way, the overall advice support provided by *BNF* might have been underestimated, although no concrete clinical advice was provided. For future work, it would be interesting to evaluate a larger number of DIRs and include DIRs (e.g. Medscape) which have terms and conditions that currently prohibit the type of analysis we have conducted, as well as more comprehensive efforts to compare clinical management recommendations among entire resources.

## 4.2 Implications and conclusions

It is reasonable to assume that the inclusion of clinically significant DDIs in DIRs would lead to decreasing rates of preventable medication errors related to adverse DDIs. On the other hand, the value of these tools in clinical practice could be diminished in cases where too many minor or clinically insignificant DDIs are included in an effort to limit legal liability.<sup>25</sup> This phenomenon, described as alert fatigue, can also lead to important clinical consequences in patient safety, as practitioners tend to underestimate and ignore the alerts provided by the system due to the sheer volume of generated alerts.<sup>26</sup>

The Evidence workgroup from the DDI CDS Conference Series has highlighted the importance and need for higher-quality information related to DDIs and also suggested the establishment of systematic DDI search criteria in order to determine the existing evidence related to the information provided.<sup>27</sup> Our analysis also shows the need for consistency in the definitions of severity and evidence ratings provided by the various DIRs.

The inclusion of clinical management options for DDIs in CDS tools is also quite important and especially useful in the clinical setting. More focus on this aspect has been suggested in multiple studies, which advocate more detailed and actionable advice (e.g. what and when to monitor) and clear indications of the strength of the recommendation.<sup>28,29</sup> We recommend that, by providing a dedicated section for clinical management recommendations that contains clear, actionable recommendations, information retrieval in the clinic can be facilitated. Also, information about specific patient risk factors for DDIs, such as genetic polymorphisms, could be included in separate text fields to enhance DDI preventability.

The use of pharmacologic drug classes in *Thesaurus* to summarise DDIs might become a source of confusion for clinicians. In many cases, individual drugs from a drug class have different pharmacological profiles (e.g. excretion,

metabolism, etc), which contribute to a markedly different DDI risk when considering the same interacting drug.<sup>28</sup> Consequently, drug ingredient indexing may paradoxically impede searching rather than achieving the aim of providing effective DDI summaries, as a large volume of information is presented in a short space where multiple severity levels are assigned and different recommendations are provided depending on the individual drug from the drug class.

In conclusion, there is a great deal of interest in clinical decision support systems providing information on drug-drug interactions to improve prescribing and reduce the risk of adverse DDIs that affect either efficacy and/or safety of individual medicines. However, there is a lack of consistency and standardisation in the information provided by different DIRs. Our study which has systematically compared three DIRs has identified that there is considerable variation in the DDI information provided in those resources. Such variability in information could have deleterious consequences for patient safety, and there is a need for harmonisation and standardisation.

## Acknowledgement

We would like to thank Ms Angela Hall from the Library and Knowledge Service at the Liverpool University Hospitals NHS Foundation Trust for facilitating access to Micromedex. The authors would like to acknowledge Antoni Wisniewski and Isobel Anderson for providing feedback on this research.

## Conflict of interest statement

EK receives PhD studentship that is jointly funded by AstraZeneca and the EPSRC, and currently works on a fixed-term employment contract for AstraZeneca. AB is an employee of Ceva Santé Animale. MP has received partnership funding for the following: MRC Clinical Pharmacology Training Scheme (co-funded by MRC and Roche, UCB, Eli Lilly and Novartis); and grant funding from Vistagen Therapeutics. He has also unrestricted educational grant support for the UK Pharmacogenetics and Stratified Medicine Network from Bristol-Myers Squibb and UCB. He has developed an HLA genotyping panel with MC Diagnostics, but does not benefit financially from this. MP is part of the IMI Consortium ARDAT ([www.ardat.org](http://www.ardat.org)). BD is an employee of AstraZeneca and holds stocks of AstraZeneca. SM has no conflict of interest to declare.

## Funding information

EK is funded for her PhD through joint funding from EPSRC and AstraZeneca.

## Data availability statement

Data from the DIRs were derived from the following web sources: the *BNF* website (<https://bnf.nice.org.uk/interaction/>, available in the public domain); ANSM website (<https://ansm.sante.fr/documents/reference/thesaurus-des-interactions-medicamenteuses-1>, available in the public domain). Restrictions apply to the availability of Micromedex data, which were used under license for this study. Data are available at <https://www.micromedexsolutions.com/> with the permission of IBM Watson Health. The code that supports the web data extraction and analysis are available at: <https://github.com/elpidakon/CRESCENDDI>.

## References

1. Pirmohamed M, James S, Meakin S, Green C. Adverse drug reactions as cause of admission to hospital. *BMJ Br Med J*. 2004;329(7463):460. doi:10.1136/bmj.329.7463.459-b
2. Egger SS, Drewe J, Schlienger RG. Potential drug--drug interactions in the medication of medical patients at hospital discharge. *Eur J Clin Pharmacol*. 2003;58(11):773-778. doi:10.1007/s00228-002-0557-z
3. Department of Health and Social Care. *Good for You, Good for Us, Good for Everybody: A Plan to Reduce Overprescribing to Make Patient Care Better and Safer, Support the NHS, and Reduce Carbon Emissions.*; 2021.
4. Magro L, Moretti U, Leone R. Epidemiology and characteristics of adverse drug reactions caused by drug--drug interactions. *Expert Opin Drug Saf*. 2012;11(1):83-94. doi:10.1517/14740338.2012.631910
5. Bjornsson TD, Callaghan JT, Einolf HJ, et al. THE CONDUCT OF IN VITRO AND IN VIVO DRUG-DRUG INTERACTION STUDIES: A PHARMACEUTICAL RESEARCH AND MANUFACTURERS OF AMERICA (PhRMA) PERSPECTIVE. *Drug Metab Dispos*. 2003;31(7):815-832. doi:10.1124/dmd.31.7.815
6. Romagnoli KM, Nelson SD, Hines L, Empey P, Boyce RD, Hochheiser H. Information needs for making clinical recommendations about potential drug-drug interactions: a synthesis of literature review and interviews. doi:10.1186/s12911-017-0419-3
7. Hansten PD. Drug interaction management. *Pharm World Sci*. 2003;25(3):94-97. doi:10.1023/A:1024077018902
8. Vitry AI. Comparative assessment of four drug interaction compendia. *Br J Clin Pharmacol*. 2007;63(6):709-714. doi:10.1111/j.1365-2125.2006.02809.x
9. Fulda TR, Valuck RJ, Zanden J Vander, Parker S, Byrns PJ. Disagreement among drug compendia on inclusion and ratings of drug-drug interactions. *Curr Ther Res - Clin Exp*. 2000;61(8):540-548. doi:10.1016/S0011-393X(00)80036-3
10. Ayvaz S, Horn J, Hassanzadeh O, et al. Toward a complete dataset of drug-drug interaction information from publicly available sources. *J Biomed Inform*. 2015;55:206-217. doi:10.1016/j.jbi.2015.04.006
11. Phansalkar S, Desai AA, Bell D, et al. High-priority drug-drug interactions for use in electronic health records. *J Am Med Informatics Assoc*. 2012;19(5):735-743. doi:10.1136/amiainl-2011-000612
12. Vonbach P, Dubied A, Krähenbühl S, Beer JH. Evaluation of frequently used drug interaction screening programs. *Pharm World Sci*. 2008;30(4):367-374. doi:10.1007/s11096-008-9191-x
13. Wang LM, Wong M, Lightwood JM, Cheng CM. Black box warning contraindicated comedications: Concordance among three major drug interaction screening programs. *Ann Pharmacother*. 2010;44(1):28-34. doi:10.1345/aph.1M475
14. Fung KW, Kapusnik-Uner J, Cunningham J, Higby-Baker S, Bodenreider O. Comparison of three commercial knowledge bases for detection of drug-drug interactions in clinical decision support. *J Am Med Informatics Assoc*. 2017;24(4):806-812. doi:10.1093/jamia/ocx010
15. BNF: British National Formulary - NICE. <https://bnf.nice.org.uk/>. Accessed September 5, 2018.
16. Thésaurus des interactions médicamenteuses - ANSM : Agence nationale de sécurité du médicament et des produits de santé. <https://ansm.sante.fr/documents/reference/thesaurus-des-interactions-medicamenteuses-1>. Accessed October 10, 2020.

17. Micromedex® (electronic version). IBM Watson Health, Greenwood Village, Colorado, USA. <https://www.micromedexsolutions.com/>. Accessed August 25, 2018.
18. Python Software Foundation. Python Software Foundation | Python Software Foundation.
19. Cossin S. IMthesaurusANSM: Thesaurus des Interactions Medicamenteuses de l' ANSM. 2016.
20. Nelson SJ, Zeng K, Kilbourne J, Powell T, Moore R. Normalized names for clinical drugs: RxNorm at 6 years. *J Am Med Informatics Assoc.* 2011;18(4):441-448. doi:10.1136/amiajnl-2011-000116
21. RxNorm Extension: an OHDSI resource to represent international drugs. [https://www.ohdsi.org/web/wiki/doku.php?id=documentation:international\\_drugs](https://www.ohdsi.org/web/wiki/doku.php?id=documentation:international_drugs). Accessed February 26, 2020.
22. Usagi: an application to help create mappings between coding systems and the Vocabulary standard concepts. <https://github.com/OHDSI/Usagi>. Accessed February 15, 2020.
23. Natural Language Toolkit (NLTK). <http://www.nltk.org/>.
24. Fung KW, Kapusnik-Uner J, Cunningham J, Higby-Baker S, Bodenreider O. Comparison of three commercial knowledge bases for detection of drug-drug interactions in clinical decision support. *J of the Am Med Informatics Assoc.* 2017;24(4):806-812. doi:10.1093/jamia/ocx010
25. Greenberg M, Ridgely MS. Clinical Decision Support and Malpractice Risk. *Jama.* 2011;306(1):90-91.
26. Phansalkar S, van der Sijs H, Tucker AD, et al. Drug-drug interactions that should be noninterruptive in order to reduce alert fatigue in electronic health records. *J Am Med Informatics Assoc.* 2013;20(3):489-493. doi:10.1136/amiajnl-2012-001089
27. Scheife RT, Hines LE, Boyce RD, et al. Consensus Recommendations for Systematic Evaluation of Drug–Drug Interaction Evidence for Clinical Decision Support. *Drug Saf.* 2015;38(2):197-206. doi:10.1007/s40264-014-0262-8
28. Horn JR, Gumpfer KF, Hardy JC, McDonnell PJ, Phansalkar S, Reilly C. Clinical decision support for drug-drug interactions: improvement needed. *Am J Health Syst Pharm.* 2013;70(10):905-909. doi:10.2146/ajhp120405
29. Tilson H, Hines LE, McEvoy G, et al. Recommendations for selecting drug-drug interactions for clinical decision support. *Am J Heal Pharm.* 2016;73(8):576-585. doi:10.2146/ajhp150565

## Tables

DIR	Extracted fields	Categories
<b>BNF</b> (Accessed in June 2018)	<ul style="list-style-type: none"> <li>- Drug Name</li> <li>- Interactant Name</li> <li>- Description</li> <li>- Severity (where present)</li> <li>- Evidence (where present)</li> </ul>	<p>(a) active pharmaceutical ingredients (APIs) (e.g. atropine);</p> <p>(b) drug classes (e.g. combined hormonal contraceptives);</p> <p>(c) herbs and supplements (e.g. peppermint oil);</p> <p>(d) foods and beverages (e.g., grapefruit juice).</p>
<b>Thesaurus</b> (Accessed in July 2020; September 2019 update)	<ul style="list-style-type: none"> <li>- Drug Name</li> <li>- Interactant Name</li> <li>- Mechanism (if available)</li> <li>- Severity Index</li> <li>- Additional Information (specification for drug class, etc)</li> <li>- Clinical Information (i.e. manifestation, management)</li> </ul>	<p>(a) Drug ingredient;</p> <p>(b) Drug classes.</p>
<b>Micromedex</b> (Accessed: August 2018)  <b>In Depth Answers Database (Detailed evidence-based information)</b>	<ul style="list-style-type: none"> <li>- Drug Name</li> <li>- Interactant Name</li> <li>- Interaction Effect</li> <li>- Summary</li> <li>- Severity</li> <li>- Onset</li> <li>- Substantiation</li> <li>- Clinical Management</li> </ul>	<p>(c) Drug ingredients;</p> <p>(d) Combination drugs;</p> <p>(e) Food;</p> <p>(f) Tobacco;</p> <p>(g) Lab tests.</p>

**Table 1.** Extracted information from the DIRs.

DIR	Level	Definition
<b>(a) Severity</b>		
<b>BNF</b>	1 - Severe*	The result may be a life-threatening event or have a permanent detrimental effect.
	2 - Moderate	The result could cause considerable distress or partially incapacitate a patient; they are unlikely to be life-threatening or result in long-term effects.
	3 - Mild	The result is unlikely to cause concern or incapacitate the majority of patients.
	4 - Unknown	Used for those interactions that are predicted, but there is insufficient evidence to hazard a guess at the outcome.
<b>Thesaurus</b>	1 - Contraindicated*	
	2 - Not recommended*	
	3 - Precautions for use	
	4 - Take into consideration	
<b>Micromedex</b>	1 - Contraindicated*	The drugs are contraindicated for concurrent use.
	2 - Major*	The interaction may be life-threatening and/or require medical intervention to minimize or prevent serious adverse effects.
	3 - Moderate	The interaction may result in exacerbation of the patient's condition and/or require an alteration in therapy.
	4 - Minor	The interaction would have limited clinical effects. Manifestations may include an increase in the frequency or severity of the side effects but generally would not require a major alteration in therapy.
<b>(b) Evidence</b>		
<b>BNF</b>	Study	For interactions where the information is based on formal study including those for other drugs with same mechanism (e.g. known inducers, inhibitors, or substrates of cytochrome P450 isoenzymes or P-glycoprotein).
	Anecdotal	Interactions based on either a single case report or a limited number of case reports.
	Theoretical	Interactions that are predicted based on sound theoretical considerations. The information may have been derived from in vitro studies or based on the way other members in the same class act.
<b>Micromedex</b>	Established (Excellent)	Controlled studies have clearly established the existence of the interaction.
	Theoretical (Good)	Documentation strongly suggests the interactions exists, but well-controlled studies are lacking.
	Probable (Poor)	Available documentation is poor, but pharmacologic considerations lead clinicians to suspect the interaction exists; or, documentation is good for a pharmacologically similar drug.

(\*): critical severity ratings

**Table 2.** (a) DIR severity ratings in descending order (as displayed in the original source); (b) DIR Evidence Ratings.

DIR	Initial drug names	Normalised ingredients	DDI counts
BNF	1,004	984	51,481
Thesaurus	1,049	1,001	38,037
Micromedex	2,602	1,967	65,446

**Table 3.** Number of initial drug names, normalised ingredients and DDI counts per DIR.

Severity rating	BNF	Thesaurus	Micromedex
1	12,644 (24.56%)	2,949 (7.75%)	5,730 (8.76%)
2	4,997 (9.46%)	12,779 (33.60%)	41,713 (63.73%)
3	273 (0.51%)	8,195 (21.54%)	15,890 (24.28%)
4	33,705 (65.47%)	14,114 (37.11%)	2,113 (3.23%)

**Table 4.** Number and percentage of DDIs by severity rating in each DIR.

## Figure legends

**Figure 1.** Venn diagrams illustrating the intersections in terms of: (a) drug ingredients; (b) unique DDI pairs included in the DIRs and (c) DDI pairs included in the DIRs only for the ingredient intersection subset. Each circle represents a DIR and their intersections show the number of ingredients/DDIs they share with the each one of the other DIRs.

**Figure 2.** Pairwise comparison tables for the different DIR severity levels. For tables (a)-(c), row labels contain the severity ratings of the DIR under consideration, while column labels represent the severity ratings of the remaining two DIRs. A separate column has been added to include the numbers of unique DDIs missing from each of the other DIRs. Each row contains the number of unique DDIs per severity rating of the DIR under consideration, subcategorized by the severity ratings of the other DIRs. The numbers in parentheses represent the corresponding percentages of the various sets per severity rating of the DIR under consideration. Colour gradient shows the relative differences in the percentages mentioned among the various overlapping sets.

**Figure 3.** Heatmap for evidence rating comparison between BNF and Micromedex, including counts and coverage rates.

**Figure 4.** Venn diagrams for clinical management advice overlap among the DIRs.