

REVIEW

Open Access



# Diagnosis of Rare Diseases: a scoping review of clinical decision support systems

Jannik Schaaf<sup>1\*</sup> , Martin Sedlmayr<sup>2</sup>, Johanna Schaefer<sup>1</sup> and Holger Storf<sup>1</sup>

## Abstract

**Background:** Rare Diseases (RDs), which are defined as diseases affecting no more than 5 out of 10,000 people, are often severe, chronic and life-threatening. A main problem is the delay in diagnosing RDs. Clinical decision support systems (CDSSs) for RDs are software systems to support clinicians in the diagnosis of patients with RDs. Due to their clinical importance, we conducted a scoping review to determine which CDSSs are available to support the diagnosis of RDs patients, whether the CDSSs are available to be used by clinicians and which functionalities and data are used to provide decision support.

**Methods:** We searched PubMed for CDSSs in RDs published between December 16, 2008 and December 16, 2018. Only English articles, original peer reviewed journals and conference papers describing a clinical prototype or a routine use of CDSSs were included. For data charting, we used the data items "Objective and background of the publication/project", "System or project name", "Functionality", "Type of clinical data", "Rare Diseases covered", "Development status", "System availability", "Data entry and integration", "Last software update" and "Clinical usage".

**Results:** The search identified 636 articles. After title and abstracting screening, as well as assessing the eligibility criteria for full-text screening, 22 articles describing 19 different CDSSs were identified. Three types of CDSSs were classified: "Analysis or comparison of genetic and phenotypic data," "machine learning" and "information retrieval". Twelve of nineteen CDSSs use phenotypic and genetic data, followed by clinical data, literature databases and patient questionnaires. Fourteen of nineteen CDSSs are fully developed systems and therefore publicly available. Data can be entered or uploaded manually in six CDSSs, whereas for four CDSSs no information for data integration was available. Only seven CDSSs allow further ways of data integration. thirteen CDSS do not provide information about clinical usage.

**Conclusions:** Different CDSS for various purposes are available, yet clinicians have to determine which is best for their patient. To allow a more precise usage, future research has to focus on CDSSs RDs data integration, clinical usage and updating clinical knowledge. It remains interesting which of the CDSSs will be used and maintained in the future.

**Keywords:** Rare diseases, Computer-assisted diagnosis, Clinical decision support systems

\* Correspondence: [jannik.schaaf@kgu.de](mailto:jannik.schaaf@kgu.de)

<sup>1</sup>Medical Informatics Group (MIG), University Hospital Frankfurt, Frankfurt, Germany

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

In the European Union (EU), a disease is declared as “rare” if no more than 5 out of 10,000 people are affected [1]. It is estimated that about 7000 different rare diseases (RDs) exist. According to the World Health Organization (WHO), about 400 million people are affected [2]. Many RDs are severe, chronic and life-threatening [3, 4]. 80% of RDs are of genetic origin and pre-dominantly affect children [5–9]. For instance cystic fibrosis as a rare lung disease occurs in the first years of the childhood and is associated with an average life expectancy of 40 years [10]. Other RDs like amyotrophic lateral sclerosis, a degenerative disease of the central and peripheral nervous system, can occur later in life and lead to death within a few years [11]. A big challenge in the management of RDs is finding the right diagnosis. Patients with RDs are sometimes diagnosed too late or not at all. They report many years of a diagnosis odyssey [4].

In the past, several clinical decision support systems (CDSSs) have been developed to support clinicians in finding the right diagnosis for patients with RDs. According to Hunt et al., a CDSS is defined as a system that supports clinical decision-making by comparing characteristics of patients to a knowledge base and collecting and displaying the results [12]. We refer to any system matching this definition as “specific CDSS”. Every other system that a physician might use for decisions, but which does not actively give recommendations based on patient characteristics, is called an “implicit CDSS”.

The information of CDSS is very limited and only two reviews about software for diagnosis support in RDs are currently available. Mueller et al. [13] present an overview of software that can be used to support the diagnosis of RDs. Their article includes different types of software and databases that match both our specific and implicit CDSS categories. In addition, only fully developed systems are presented that are (1) available for download and can be installed on one’s own computer or (2) are only useable online. Systems under development, such as research prototypes or tools in clinical evaluation, have not been considered. However, when developing a new CDSS, software developers require information which prototypes are available and which data and functions they use [12]. The second review by Svenstrup et al. gives an overview of web search, social media and data mining approaches for the diagnosis of RDs. However, this article mainly focusses on their own web search engine FindZebra [14]. Despite their importance, we are not aware of any reviews about developments and current systems of specific RDs CDSSs.

Due to the importance of improving the diagnosis of RDs, we conducted a scoping review in order to map the research performed in this area, to reveal gaps in knowledge as well as to give clinicians an overview of the

specific CDSSs that are currently available. The need for this research is highlighted by the fact that the support of diagnosis of RDs using software is part of national strategy plans for RDs, e.g. in Germany (National Plan of Action for People with Rare Diseases [15] and the United Kingdom (The UK Strategy for Rare Diseases) [16].

The objectives of the scoping review were to show clinicians as well as software-developers (1) which specific CDSSs are available to support the diagnosis of patients with RDs, (2) which functionalities and data are used within the specific CDSSs, (3) which CDSSs can be used by clinicians directly and (4) how data can be entered or automatically integrated into the specific CDSSs.

## Methods

The reporting of this scoping review complies with PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [17]. We considered 19 out of 22 PRISMA-ScR items (shown in Additional file 1). We created and uploaded a review protocol in Open Science (URL: <https://osf.io/de79v>). The author JAS drafted the protocol in December 2018, which was approved by all other authors on December 15, 2018. It was last updated after accessing information about the identified CDSS in January 2020. The final protocol was uploaded for publication retrospectively on 01 March 2020.

## Sources of information and search criteria

To identify relevant articles, we searched PubMed. Unpublished literature was not considered. We did not contact the authors of the articles, but checked the reference lists for further sources of evidence. We retrieved articles published over the course of 10 years, from December 16, 2008 to December 16, 2018, to capture as many relevant publications as possible. The final search was conducted on December 16, 2018.

## Definition process of search terms

The author JAS performed an initial search in PubMed with a combination of the terms “Clinical Decision Support” and “Rare Diseases”. The terms were combined with a logical “AND”. The goal was to identify relevant keywords for a broad search. The results of the search were 165 publications. To obtain keywords, we checked titles and abstracts of the publications to determine whether they described an specific CDSS for RDs and identified five relevant publications [13, 14, 18–20]. Afterwards, we extracted the keywords from these publications in a brainstorming session with all authors and decided which of them were relevant for the search (see Additional file 2 – Part A). Based on the identified keywords, we created a map to establish a relationship between them (see Additional file 2 – Part B-C).

In the next step, the identified keywords were tested and mapped to MeSH terms (Medical Subject Heading) by JAS, validated by MS and approved by all authors (described in Additional file 2 – Part D). We also added non-MeSH terms to cover articles which do not appear in the index of MeSH [21]. This led us to our final search terms (shown in Additional file 2 – Part E). We grouped the terms into four groups “A” to “D”. The groups were combined with a logical “AND”. Terms of group “A” and “B” (MeSH terms) and “C and D” (non-MeSH terms) were combined for the search. The search was conducted by JAS with the final query in Fig. 1.

### Inclusion criteria and study selection

We conducted two screening rounds to select publications: A screening based on bibliographic data and a full-text screening. Eligibility criteria are shown in Table 1. JAS created screening forms with Microsoft Word to handle eligibility criteria (Additional file 3). All authors approved the forms.

In title and abstract screening, we investigated the search result based on the bibliographic data. Publications were included if they contained a peer-reviewed journal or conference paper and an abstract written in English, and if the publication contained primary research of a CDSS in RDs. All other publications were excluded. To test the screening form, we used a random sample of 63 publications (~10%). JAS made the decisions on which publications to include, and these decisions were verified by MS. A revision of the form was not necessary, since all authors agreed on the results. After this step, the complete title and abstract screening was performed by JAS and verified by MS. Any disagreements were discussed with all authors and resolved by consensus.

Where available, full-text publications were screened by JAS regarding they describe a specific CDSS for RDs or not. Publications were included if they described a clinical prototype or a routine use of a CDSS for RDs. Publications about any other types of CDSS implementations (e.g. concepts or software architectures), were not considered.

Similar to the abstract and title screening, a screening form for full-text screening was tested by JAS with two (~10%) of the remaining publications. We discussed the results amongst all authors and agreed that a revision of the screening form was not necessary. Subsequently, JAS screened all full-text publications available. We considered additional sources using the same eligibility criteria. MS verified the results and any disagreements were resolved by discussion and consensus with all authors. After these two screening rounds, all remaining papers were obtained for data charting.

### Data charting

All authors jointly developed and agreed on a data charting form to determine which data items to extract and guide the author through the data charting process (Additional file 4). We selected the data items based on our research questions. The first version of the data charting form covered six data items. JAS tested the form with five available full-text publications. After the discussion of the pilot test between JAS and MS, we agreed to add four more data items, as of further interest for our research. After the revision of the data charting form, the publications were entered into a spreadsheet by JAS, verified by MS and approved by all authors. The data items are described in Table 2.

```
(Rare Diseases[MeSH Terms] AND Medical Informatics Applications[MeSH Terms])
OR (Rare Diseases[MeSH Terms] AND Medical Informatics[MeSH Terms])
OR (Rare Diseases[MeSH Terms] AND Algorithms[MeSH Terms])
OR (Rare Diseases[MeSH Terms] AND Diagnosis, Computer-Assisted[MeSH Terms])
OR (Rare Diseases[MeSH Terms] AND Decision Support Techniques[MeSH Terms])
OR (Rare Diseases[MeSH Terms] AND Decision Support Systems, Clinical[MeSH Terms])
OR (Rare Diseases[MeSH Terms] AND Decision Making, Computer-Assisted[MeSH Terms])
OR (Rare Diseases[MeSH Terms] AND Search Engine[MeSH Terms])
OR ((rare diseases) AND diagnostic decision support)
OR ((rare disease) AND diagnostic decision support)
OR ((orphan diseases) AND diagnostic decision support)
OR ((orphan disease) AND diagnostic decision support)
```

**Fig. 1** Search Query

**Table 1** Inclusion and exclusion criteria for title and abstract screening and full-text screening

Screening round	Inclusion	Exclusion
Title and abstract screening	The publication is primary research in a peer-reviewed journal or a conference paper.	The publication is of any other publication type. For instance, literature reviews, study protocols, commentaries and editorials were excluded.
Title and abstract screening	The publication has an abstract available.	The publication has no abstract available.
Title and abstract screening	The publication is written in English	The publication is written in any other language than English.
Title and abstract screening	The publication contains primary research or a report of a Clinical Decision Support System for Rare Diseases.	The publication contains any other description of software for Rare Diseases and not a Clinical Decision Support System.
Full-text screening	The publication describes a specific Clinical Decision Support System for Rare Diseases.	The publication describes an implicit Clinical Decision Support System for Rare Diseases.
Full-text screening	The publication describes a clinical prototype or a routine use of a Clinical Decision Support System for Rare Diseases.	The publication describes any other types of Clinical Decision Support Systems implementations (e.g. concepts, software architectures).

### Summarizing and reporting the results

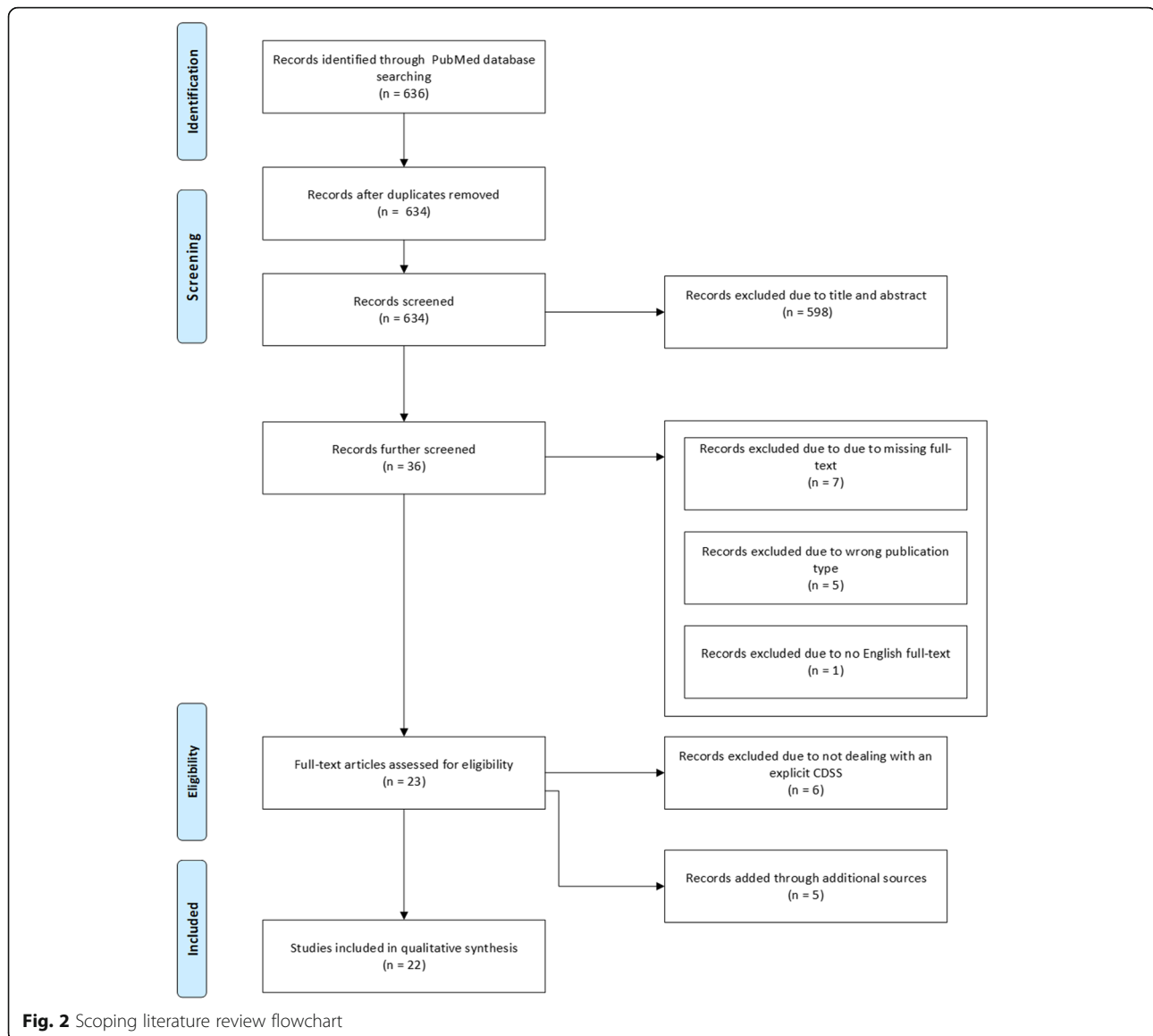
To present the results, we prepared an overview of all results regarding the data items. Furthermore, we grouped each relevant CDSS according to the data item “Functionality”. We described the background of a CDSS and the data used to perform decision support. We also prepared an overview of the development status of each CDSS and stated whether and how the system is available. Furthermore, we described how data can be entered or integrated into the CDSS and provided information about the latest updated software version. We also provided a short summary including all data items at the end of each section.

### Results

The search identified 636 articles in PubMed (shown in Fig. 2). After removing two duplicates, 634 articles were available for title and abstract screening. In the first screening step, 598 articles were excluded and 36 articles were considered relevant. This number was further reduced due to not accessible full-text of seven articles, wrong publication types in five articles and no full-text in English in one article. After assessing the eligibility of the remaining 23 articles, six articles were excluded because they did not deal with a specific CDSS and were neither in clinical nor in routine use. This resulted in 17 articles and an additional of five articles [22–26] were

**Table 2** Data items for data charting

Data item	Description
Objective and background of the publication/project	The reader should identify the relevant key message and the objective of the publication or project. This should give the clinician or researcher of this publication an overview of which specific CDSSs for RDs are available and why the project and publication was developed.
System or project name	The “System or project name” is the name of the RDs CDSS, the related project name or the name of the first author.
Functionality	We define “Functionality” as the technology that performs the decision support (e.g. machine learning). The goal is to show the clinicians or researchers how the decision support is derived.
Type of clinical data	“Type of clinical data” indicates whether the CDSS uses clinical routine data (e.g. lab, reports and documentation) or phenotypic and genetic data.
Rare Diseases covered	“Rare Diseases covered” is defined as which diseases are covered by the CDSS.
Development status	“Development status” describes if the CDSS is a clinical prototype or a fully developed system.
System availability	“System availability” describes whether the system is available for download or whether online usage is possible, as well as any access restrictions.
Data entry and integration	“Data integration” describes whether data can be entered manually or transferred automatically into the CDSS (e.g. via file upload or REST-API).
Last software update	“Last software update” reports when the CDSS was last updated to a new software version.
Clinical usage	“Clinical usage” reports if there are any information about CDSS available regarding clinical usage (e.g. amount of patient cases, amount of users, participating hospitals)?



added after checking the reference list of the publications. A total of 22 articles were available, describing 19 different CDSSs (shown in Additional file 5).

### Overview of the results

Twelve out of nineteen CDSSs use “analysis or comparison of genetic and phenotypic data” as a CDSS functionality [22–25, 27–34]. Three of the CDSSs are based on “machine learning” [20, 26, 35]. Four CDSSs use “information retrieval” [18, 36–38]. The results show that five CDSSs are clinical prototypes [18, 20, 26, 35, 36] whereas 14 are fully developed systems [22–25, 27–34, 37, 38]. Twelve CDSSs use phenotypic and/or genetic data [22–25, 27–34], three use literature databases [18, 37, 38]. Six use clinical data in addition to other data

[24, 29, 30, 32, 34] and two use clinical data only [20, 36]. Two CDSSs use patient questionnaires [26, 35]. In 15 of 19 CDSS, all RDs can be included for decision support [18, 22, 23, 27–34, 37, 38]. Only four CDSS are restricted to a group or different RDs [20, 26, 35, 36].

Regarding system availability, five CDSSs are not available for personal use [18, 20, 26, 35, 36], six CDSSs can be used online and free after completing a registration [27, 28, 30, 31, 33, 34]. Furthermore, three CDSS can be used online and free without any registration [22, 29, 37]. One CDSS can be downloaded without any registration [38]. Four CDSSs can be downloaded, but require registration before access is granted [23, 24, 31, 32]. Data can be entered only via forms in four CDSSs [22, 28, 31, 37], whereas six CDSSs additionally allow the upload of

data and files [23, 25, 29, 32–34]. We could not find exact information on data entry or integration for four CDSSs [18, 20, 26, 35]. Six CDSSs described further ways of a data integration complying with the REST standard [24, 27, 30, 32–34], whereas data can be integrated with ETL processes in one CDSS [36] and with data upload only in one other CDSS [38].

There is no information about software updates for 13 CDSSs [18, 20, 22, 23, 25–28, 33, 35–38], whereas six CDSS were updated between November 2017 and August 2019 [24, 29–32, 34]. One CDSS provided regular updates, but no information about the release date was provided [24]. Furthermore, we could not find any reporting about the current clinical usage for 13 CDSS [18, 20, 22–26, 28, 29, 31, 35, 37, 38], whereas 6 CDSS provide information [27, 30, 31, 33, 34, 36]. Table 3 shows a comparative overview of the results.

### CDSSs using machine learning

Machine learning (ML) allows computer-systems to “learn” from data using statistic methods [39]. CDSSs using ML can be trained using medical data in order to support the decision of a clinician [40]. In this section, we show CDSSs for RDs using ML.

Rother et al. [26], Grigull et al. [35] and Sidiropoulos et al. [20] developed clinical prototypes using ML algorithms for the diagnoses of patients with rare pulmonary diseases, rare neuromuscular diseases and rare cancers. Grigull et al. [35] and Rother et al. [26] focused on rare pulmonary and neuromuscular diseases and used patient-related questionnaires to train ML algorithms. In the results, they achieved a diagnosis rate of 89% respectively 94% [26, 35]. Sidiropoulos et al. [20] developed a real-time decision support system for the diagnosis of rare cancers. The authors used a GPU

**Table 3** Results of the charted data items

Data item	Subcategories	Total/Frequency
Functionality	Analysis or comparison of genetic and phenotypic data	12 (63.2%)
	Information retrieval	4 (21.0%)
	Machine learning	3 (15.8%)
Development status	Fully developed systems	14 (73.7%)
	Clinical prototypes	5 (26.3%)
Type of clinical data	Phenotypic and/or genetic data	12 (63.2%)
	Clinical data	8 (42.1%)
	Literature databases	3 (15.8%)
	Patient questionnaires	2 (10.5%)
Rare Diseases covered	All rare diseases	15 (78.8%)
	Neuromuscular rare diseases	1 (5.3%)
	Pulmonary rare diseases	1 (5.3%)
	Rare brain cancer diseases	1 (5.3%)
	Other	1 (5.3%)
System availability	The system can be used online and free, subject to registration	6 (31.6%)
	The system is not available for personal use	5 (26.3%)
	The system can be downloaded, subject to registration	4 (21.0%)
	The system can be used online and free, no registration necessary	3 (15.8%)
	The system can be downloaded, no registration necessary	1 (5.3%)
Data entry and integration	Data entry with forms and data upload is possible	6 (31.6%)
	REST-API available	6 (31.6%)
	Data entry is only possible with forms	4 (21.0%)
	No information available	4 (21.0%)
	ETL processes	1 (5.3%)
	Data upload is possible	1 (5.3%)
Last software update	No information available	13 (68.4%)
	Information available	6 (31.6%)
Current clinical usage	No information available	13 (68.4%)
	Information available	6 (31.6%)

framework (Graphics Processing Unit) to show a result in real time based on histological clinical data. This allows a faster real-time decision than on a CPU-based system (Central Processing Unit). The system subsequently suggested the correct diagnosis in about 74% of the cases and performed up to 288 times faster than on the CPU [20]. Since all three CDSSs are clinical prototypes, no information about software update, clinical usage, data integration and access for clinicians are available.

#### Summary for clinical usage

Table 4 shows the summary of “CDSSs using machine learning”.

#### CDSSs using information retrieval

Online databases, like PubMed, are consulted by clinicians to search for case reports of patients. Often, case reports are manually compared to identify similar characteristics of patients. This process is time-consuming and inefficient. With the help of methods like Information retrieval (IR), it is possible to find information, especially in large databases or on the internet [41]. IR includes different techniques to retrieve information based on keywords. For instance, search engines like Google use IR methods [42]. In this review, we show CDSSs using IR which support the diagnosis of RDs.

#### Identify relevant information in databases based on symptoms and phenotypes

FindZebra [37] is a search engine and a fully developed system that allows clinicians to enter symptoms in a search field and find corresponding information in databases. The knowledge base of FindZebra is built on 33,144 documents covering approximately 90% of the RDs listed in the Orphanet database. FindZebra uses ten

sources for their dataset, for instance OMIM (Online Mendelian Inheritance in Man) and GARD (Genetic and Rare Diseases Information Center) [37]. OMIM contains descriptions about human genes and their correlation with phenotypes, which are defined in genetics as a set of all visible characteristics of an organism [37]. GARD provides a database about RDs with symptoms, treatments and further research information [37].

To evaluate FindZebra, the authors compared FindZebra with other platforms like PubMed and Google, using 56 search queries with patient symptoms based on expert knowledge. The findings show that FindZebra outperforms Google and PubMed. Especially for queries with a long list of symptoms, FindZebra achieves better results. Google uses an algorithm based on how often a website is visited or linked to other websites. The authors concluded that this would lead to poor results for RDs [37]. However, an information about software updates, clinical usage or further data integration is not provided.

A further CDSS to identify relevant information in a database is the CDSS of Taboada et al. [38]. The authors described a fully developed system which can automatically capture relevant literature data based on phenotypes. Their CDSS uses so-called “text annotation”, a method from the field of Natural Language Processing (NLP), to identify relevant words in a text.

The evaluation of the CDSS was based on a disease “Cerebrotendinous xanthomatosis” (CTX), a rare disorder of bile acid metabolism. The authors extracted 223 abstracts of case reports from PubMed corresponding to CTX. Only the title and relevant parts of the texts were used for annotation. Those were annotated with the Human Phenotype Ontology (HPO) using the software Open Biological and Biomedical Ontologies (OBO) and Biportal. The HPO describes

**Table 4** Summary for clinical usage – CDSSs using machine learning

CDSSs using machine learning	<ul style="list-style-type: none"> <li>• Rother et al. [26]</li> <li>• Grigull et al. [35]</li> <li>• Sidiropoulos et al. [20]</li> </ul>
Development status and system availability	<ul style="list-style-type: none"> <li>• All CDSSs are clinical prototypes</li> <li>• All CDSSs are currently unavailable for usage</li> </ul>
Type of clinical data	<ul style="list-style-type: none"> <li>• Rother et al., Grigull et al.: Patients questionnaire</li> <li>• Sidiropoulos et al.: Clinical data</li> </ul>
Rare Diseases covered	<ul style="list-style-type: none"> <li>• Rother et al.: Pulmonary rare diseases</li> <li>• Grigull et al.: Neuromuscular rare diseases</li> <li>• Sidiropoulos et al.: Rare brain cancer diseases</li> </ul>
Data entry and integration	<ul style="list-style-type: none"> <li>• No information available</li> </ul>
Last software update	<ul style="list-style-type: none"> <li>• No information available</li> </ul>
Current clinical usage	<ul style="list-style-type: none"> <li>• No information available</li> </ul>

the correlation between phenotypes and genetic diseases.

The evaluation was measured between the automatic annotation method and a manual annotation of two neurologists, who extracted the relevant phenotypes manually. The authors evaluated the capability to identify the relevant papers for both methods (F-measure). The CDSS achieved an F-measure of 74%, which is significantly lower than the result of the manual method with 88% [38]. The authors concluded that the annotation method could have a high impact on the quality of the results [38].

FindZebra [37] and Taboada et al. [38] both provide fully developed systems, using literature databases, but different ways to find relevant data about RDs in databases. As with FindZebra, no information about software update or clinical usage is available.

#### **Using data of electronic health records for the recommendation of Rare diseases**

Garcelon et al. [36] developed a clinical prototype to find similar patients to an undiagnosed patient (index patient) in a clinical data warehouse containing about 400,000 patients. The data warehouse is a combination of different sources, e.g. electronic health records (EHR). The similarity is calculated using the Vector Space Model (VSM), by representing patient data in a mathematical vector. The similarity of two patients is measured as the presence or absence of words in the compared patient vectors [36].

Five different rare genetic diseases with 7 to 103 patient cases per disease were used for the evaluation of the CDSS. The authors evaluated its capability to find the patients who were most similar to an undiagnosed patient. Patients were considered to be similar (called true positive similar patients) when they were among the top 30 of the most similar patients and also appeared in the list of diagnosed patients, which was provided by a domain expert. The percentages of index patients, returning at least one true positive similar patient in the list of the top 30 similar patients, were reported as 94% for Lowe Syndrome, 97% for Epidermolysis Bulloas, 86% for Activated PI3K Delta Syndrome, 71% for Dowling Meara and 99% for Rett Syndrome. The average number of patients with the same disease among the top 30 similar patients was 51% [36]. Although the system achieved good results in diagnostics, it cannot be accessed and no information about a software update is available. However, data integration is described with ETL processes.

Shen et al. [18] developed a clinical prototype that uses not only clinical data. The authors merged clinical and literature data. They included clinical data from 13 million unstructured clinical notes on 700,000 patients' electronic health records limited to described problems

and diagnosis. Abstracts from research articles from the SemMedDB were extracted for the literature dataset. SemMedDb is a repository of semantic predications, extracted from titles and abstracts of all PubMed citations. The authors applied HPO and GARD terms to match the representation of both data types, followed by data fusion strategies to include the data into a collaborative filtering model to enable RD recommendation. Data fusion means that different types of data sets are combined into one dataset [18]. The authors used the following data fusion strategies: First, only the patient phenotype information was extracted from the EHR. For the second, the authors combined EHR data with phenotypes and literature. In the third fusion strategy, phenotype-rare disease associations were extracted from literature with the limitation that phenotypes of the literature data were deleted if they did not appear in the EHR data. The authors then evaluated the prediction output for each fusion strategy using a collaborative filtering model to determine which possible combinations provide the best results. This technique is used, for instance, in e-commerce to recommend products to customers based on similar buying preferences of other customers. That this scenario is similar to patients' phenotypic information. If patients have similar phenotypes, their diseases might also be similar. The results are compared with the actual diagnosis of the patient [18]. The results show that the combination of EHR and literature data did not always lead to the best performance. The authors conclude that this may be due to different approaches and expressions in clinical notes varying from physician to physician [18]. Since the CDSS is a clinical prototype, the system cannot be accessed, no information for data entry and integration is available and neither there are any information about software updates and clinical usage.

#### **Summary for clinical usage**

Table 5 shows the summary of "CDSSs using information retrieval."

#### **CDSSs using analysis or comparison of genetic and phenotypic data**

When dealing with complex symptoms of patients with RDs, it is important to identify phenotypes and to combine it with genetic testing to determine the cause of the disease (genotype) [33]. Whole-genome sequencing (WGS) and whole-exome sequencing (WES) provide possibilities to meet these challenges [29]. However, understanding the complexity of the genetic variants which can cause a disease, remains a challenge for clinicians [24]. Different software programs have been developed to tackle these problem. We call these tools "CDSSs using analysis and comparison of genetic and phenotypic



**Table 5** Summary for clinical usage – CDSSs using information retrieval

CDSSs using information retrieval	<ul style="list-style-type: none"> <li>• FindZebra [37]</li> <li>• Taboada et al. [38]</li> <li>• Garcelon et al. [36]</li> <li>• Shen et al. [18]</li> </ul>
Development status and system availability	<ul style="list-style-type: none"> <li>• FindZebra: Fully developed system. The system can be used online and free, no registration necessary</li> <li>• Taboada et al.: Fully developed system. The system can be downloaded, no registration necessary</li> <li>• Garcelon et al., Shen et al.: Clinical prototype. The system is not available for personal use</li> </ul>
Type of clinical data	<ul style="list-style-type: none"> <li>• Taboada et al., FindZebra: Literature databases</li> <li>• Garcelon et al.: Clinical data</li> <li>• Shen et al.: Literature databases and clinical data</li> </ul>
Rare Diseases covered	<ul style="list-style-type: none"> <li>• Taboada et al., Shen et al., FindZebra: All rare diseases</li> <li>• Garcelon et al.: Lowe Syndrome, Dystrophic Epidermolysis Bullosa, Activated PI3K delta Syndrome, Rett Syndrome, Dowling Meara</li> </ul>
Data entry and integration	<ul style="list-style-type: none"> <li>• FindZebra: Data entry is only possible with forms</li> <li>• Taboada et al.: Data upload is possible</li> <li>• Shen et al.: No information available</li> <li>• Garcelon et al.: ETL processes</li> </ul>
Last software update	<ul style="list-style-type: none"> <li>• No information available</li> </ul>
Clinical usage	<ul style="list-style-type: none"> <li>• FindZebra: No information available</li> <li>• Taboada et al.: No information available</li> <li>• Shen et al.: No information available</li> <li>• Garcelon et al.: 400.000 patients included</li> </ul>

data”. To showcase the results, we distinguish between (1) “CDSSs using analysis of genetic and phenotypic data”, which allow the investigation of genetic variants and their correlated phenotypes and (2) “CDSSs using comparison of genetic and phenotypic data”, which enable the identification of similar patients.

#### **Investigation of phenotype and genotype correlations**

Our review includes Phenopolis [31], GEMINI [24] and GenIO [29], which provide different tools for the investigation of genetic variants. All three of these CDSSs are fully developed systems and can be accessed by clinicians. We show further characteristics in Table 6.

Phenopolis [31] provides an open-source web server and different analysis tools like variant filtering and gene prioritization based on phenotypes of a patient using the HPO. Variant filtering allows to identify relevant variants for a diagnosis. With gene prioritization, potentially causative genes can be prioritized. Phenopolis contains 6048 exomes representing the 4,859,971 variants which comprise the data base [31].

GEMINI [24] is a software package which allows researchers to integrate the genetic variations in the Variant Call Format (VCF), a common format for the gene sequence variants. GEMINI provides variant analysis tools for the investigation of variants and a

programming interface to customize the data analysis and exploration [24]. Koile et al. developed GenIO [29] a web interface for clinicians and researchers who do not have the necessary skills to annotate, classify and filter variants. GenIO uses the so-called “GenIO pipeline”, which consists of a variant annotation and phenotype processing [29]. At the end of the phenotyping process, a list of genes with matches to the patient’s phenotype is shown [43].

#### **Summary for clinical usage**

Table 6 shows the summary of “CDSSs using analysis of genetic and phenotypic data”.

#### **Finding similar patients and sharing patient cases**

The identification of similar patients and the sharing of patient cases is possible with GeneMatcher [27], GeneYenta [28], Phenotips [32], PhenomeCentral [34], MatchMaker Exchange [30], DECIPHER [33], PhenoDB [23] and GenomeConnect [25]. All of these CDSSs are fully developed systems and can be accessed in different ways. They allow clinicians and researchers to find similar patients in a database based on genetic or phenotypic data. We show further details for clinical usage in Table 7.

**Table 6** Summary for clinical usage – CDSSs using analysis of genetic and phenotypic data

CDSs using analysis of genetic and phenotypic data	<ul style="list-style-type: none"> <li>• GenIO [29]</li> <li>• Phenopolis [31]</li> <li>• GEMINI [24]</li> </ul>
Development status and system availability	<ul style="list-style-type: none"> <li>• GenIO: Fully developed system. The system can be used online and free, no registration necessary</li> <li>• Phenopolis: Fully developed system. The system can be downloaded, subject to registration</li> <li>• GEMINI: Fully developed System. The system can be downloaded, subject to registration</li> </ul>
Type of clinical data	<ul style="list-style-type: none"> <li>• GenIO: Phenotypic, genetic and clinical data</li> <li>• Phenopolis: Phenotypic data and genetic data</li> <li>• GEMINI: Phenotypic, genetic and clinical data</li> </ul>
Rare Diseases covered	<ul style="list-style-type: none"> <li>• GenIO: All rare diseases</li> <li>• Phenopolis: All rare diseases</li> <li>• GEMINI: All rare diseases</li> </ul>
Data entry and integration	<ul style="list-style-type: none"> <li>• GenIO: Data entry with forms and data upload is possible</li> <li>• Phenopolis: Data entry is only possible with forms</li> <li>• GEMINI: Data upload and REST API available</li> </ul>
Last software update	<ul style="list-style-type: none"> <li>• GenIO: Version 1.0, 22nd of November 2017</li> <li>• Phenopolis: Version 1.0.2, 12 h of November 2017</li> <li>• GEMINI: Version 0.20.1, no date available</li> </ul>
Clinical usage	<ul style="list-style-type: none"> <li>• GenIO: No information available</li> <li>• Phenopolis: No information available</li> <li>• GEMINI: No information available</li> </ul>

PhenomeCentral, DECIPHER and GeneYenta are connected to the Matchmaker Exchange Project (MME) which connects organizations and projects through a federate network of databases of genotypes and rare phenotypes using a REST API [30]. A REST (Representational State Transfer) API (Application Programming Interface) is a web architecture style and provides the opportunity to shift data to another software system over the internet [44]. The MME enables searches across multiple databases from different platforms by making requests to all databases, e.g. to find similar matches of patients. In order to use MME, a clinician must therefore be part of a participating MME project. MME itself does not provide a user interface, but only connects existing platforms via the MME API [30].

Another CDSS which uses the similarity of phenotypes by not comparing different patient cases is Phenomizer [22]. Phenomizer is a fully developed system and facilitates differential diagnoses by using the HPO for entering phenotypes. The software classifies all diseases listed in OMIM, Orphanet and DECIPHER and uses a semantic similarity metric to measure the similarity between phenotypes and genetic diseases. Several symptoms of a patient can be entered and combined to describe the entire spectrum of a patient's symptoms. All related

diagnoses with their statistical probability are shown to rank the candidate's disease [22].

#### Summary for clinical usage

Table 7 shows the summary of "CDSSs using comparison of genetic and phenotypic data". For interested developers, we provide links to the respective REST APIs showing which data can be integrated (Additional file 6).

## Discussion

### Summary of evidence and interpretation

Our scoping review is the first to summarize the evidence of specific CDSSs for the diagnosis support for RDs. We identified 19 CDSSs between 2008 and 2018. Our findings show that most used methods of CDSSs are analysis or comparison of genetic and phenotypic data, followed by information retrieval and machine learning. However, we could not identify many publications considering machine learning, although it plays an increasing role in healthcare [45]. In other medical fields, a higher number of CDSSs can be found. For example, a review found 60 CDSSs for infectious diseases [46], while a systematic review in cardiology identified 331 relevant studies [47]. This might have been caused by the fact that machine learning for RDs is currently a

**Table 7** Summary for clinical usage – CDSSs using comparison of genetic and phenotypic data

CDSs using comparison of genetic and phenotypic data	<ul style="list-style-type: none"> <li>• DECIPHER [33]</li> <li>• GeneMatcher [27]</li> <li>• GeneYenta [28]</li> <li>• Matchmaker Exchange [30]</li> <li>• PhenomeCentral [34]</li> <li>• PhenoTips [32]</li> <li>• Phenomizer [22]</li> <li>• PhenoDB [23]</li> <li>• GenomeConnect [25]</li> </ul>
Development status and system availability	<ul style="list-style-type: none"> <li>• DECIPHER, GeneMatcher, GeneYenta, Matchmaker Exchange, PhenomeCentral, GenomeConnect: Fully developed system. The system can be used online and free, subject to registration</li> <li>• PhenoDB, PhenoTips: Fully developed System. The system can be downloaded, subject to registration</li> <li>• Phenomizer: Fully developed system. The system can be used online and free, no registration necessary</li> </ul>
Type of clinical data	<ul style="list-style-type: none"> <li>• DECIPHER, GeneMatcher, PhenoDB, GenomeConnect: Phenotypic and genetic data</li> <li>• GeneYenta, Phenomizer: Phenotypic data</li> <li>• Matchmaker Exchange, PhenoTips, PhenomeCentral: Phenotypic, genetic and clinical data</li> </ul>
Rare Diseases covered	<ul style="list-style-type: none"> <li>• DECIPHER, GeneMatcher, GeneYenta, Matchmaker Exchange, PhenomeCentral, PhenoTips, Phenomizer, PhenoDB, GenomeConnect: All rare diseases</li> </ul>
Data entry and integration	<ul style="list-style-type: none"> <li>• DECIPHER, PhenomeCentral, PhenoDB, GenomeConnect, PhenoTips: Data entry with forms and data upload is possible</li> <li>• GeneMatcher, GeneYenta, Phenomizer: Data entry is possible with forms</li> <li>• Matchmaker Exchange, DECIPHER, GeneMatcher, PhenomeCentral, PhenoTips: REST API available</li> </ul>
Last software update	<ul style="list-style-type: none"> <li>• DECIPHER, GeneMatcher, GeneYenta, Phenomizer, PhenoDB, GenomeConnect: No information available</li> <li>• Matchmaker Exchange: Version 1.1, 20th of August 2019</li> <li>• PhenomeCentral: Version 1.2.0, 14th of August 2019</li> <li>• PhenoTips: Version 1.4.7, 17 h of May 2019</li> </ul>
Clinical usage	<ul style="list-style-type: none"> <li>• DECIPHER: 270 centres, 36.000 patient cases</li> <li>• GeneMatcher: 8807 users from 90 countries</li> <li>• GeneYenta: No information available</li> <li>• Phenomizer: No information available</li> <li>• PhenoDB: No information available</li> <li>• Matchmaker Exchange: Information about connected platforms are available (see <a href="https://www.matchmakereexchange.org/statistics.html">https://www.matchmakereexchange.org/statistics.html</a>)</li> <li>• PhenomeCentral: 10.000 patient cases</li> <li>• PhenoTips: Used in over 60 countries</li> <li>• GenomeConnect: No information available</li> </ul>

problem due to a lack of data pertaining to RD patients. Garcelon et al. even recommended in their study to focus on other methods for clinical decision support [36].

Since most CDSSs use analysis or comparison of genetic and phenotypic data, this is also the most used data, followed by clinical data, literature databases and patient questionnaires. From the clinician's point of view, describing phenotypes is a challenging task. Individual patients are often not comprehensively described, for instance when a patient does not report all symptoms. If an anomaly in an individual

patient is not described, it does not follow that this anomaly does not exist. The description of the phenotype features also depends on the clinician's experience. Another problem is that the same phenotype can be caused by multiple genetic defects [29].

Furthermore our review shows, that 15 of 19 CDSS can be used for all different kinds of RDs. Only 4 CDSS are limited to different RDs. However, these CDSS have been assessed with this RDs. Therefore it is unclear whether these CDSS can be used with other diseases.

Our review includes 14 CDSSs that are fully developed systems and can be used by the clinicians directly. All of these systems can be used online or via free download. For some of them, a registration is required. Only five systems are clinical prototypes and cannot be accessed.

### Identified gaps of knowledge

Our findings indicate a lack of CDSSs that allow automatic data integration. Only seven of 19 CDSSs use ETL processes or REST interfaces. However, four studies did not describe how the data can be entered into their CDSS. We conclude this is an essential factor for the acceptance of a CDSS. Redundant data entries into several systems should be avoided [48, 49]. There are studies available specifically dealing with data integration into CDSSs [50–52], which appears to be a major challenge due to the heterogeneity of the information systems used in healthcare [53, 54]. As a possible solution, CDSSs could utilize interoperable standards such as FHIR (Fast Healthcare Interoperability Resources) for data integration [66, 67].

Furthermore, the timeliness of the data in a CDSS is important, and there should be information on when it was last updated and maintained. CDSSs should be updated regularly to integrate a new balance of evidenced-base medicine or end user preferences [55, 56]. For instance, the Roadmap for National Action on Clinical Decision Support of the American Medical Informatics Association recommends that CDSS knowledge bases and methods should be improved continuously [57]. Regarding the timeliness of data in a CDSS, only six CDSSs provide information, e.g. on their websites. These CDSSs were updated between November 2017 and August 2019.

However, since 14 CDSS are fully developed systems, we could not find any information of clinical usage for 13 CDSS. It was not possible to find sufficient and sustainable information. For instance, no evaluation reports, investigating the clinical usage, were available. It remains interesting how often and by how many users CDSS for RDs are used.

### Limitations

This work provides a broad overview of 19 different CDSSs which employ different approaches and functionalities to support the diagnosis of RDs. Not every CDSS could be explained in detail, and each system's background, medical but also technical aspects are only touched upon to provide an overview. However, this scoping review is a starting point to show clinicians and software developers what is known in the context of CDSSs for RDs. Further studies may take up this review and carry out further investigations.

Furthermore, the literature of this scoping review is up-to-date as of December 2018. We accessed the information of the topicality of the CDSS in January 2020. However, it was not possible to establish the topicality of most of the systems, since we did not contact the authors for more information. It should also be mentioned that scoping reviews do not address the risk of bias. Using only PubMed as a data source and not covering unpublished literature can have an influence on the completeness of the search. Furthermore, the study selection and data charting were only performed by one author, although they were approved by all authors. However, addressing a high methodological standard with PRISMA-ScR helped us minimize a possible bias across the study.

As this review is intended to give a broad overview of CDSSs for RDs, especially which CDSSs are available and can be used, our work is limited to collecting data about the usability of the CDSSs and effectivity of the RDs diagnosis. Some publications addressed their effectivity, but none addressed their usability in the clinical settings. We consider it necessary to involve users in the development of CDSSs, especially for clinical prototypes [40, 58–60]. We recommend to use a User-Centered Design Process (UCD), which defines an iterative process to include user requirements, needs and limitations, develop designs, prototypes and evaluate and refine it in several steps together with the users [61]. Several studies stated the importance of user-centered design in the clinical context, especially in CDSSs, to have an impact on usability and effectiveness of the systems [59, 62–64]. The use of a UCD process can help to ensure that systems are used more often in clinical routine.

Furthermore, it could be of interest at what point in the diagnosis process the CDSSs are used, since the CDSSs presented here use a variety of different approaches. More studies on the diagnosis success are needed to determine how useful these CDSS are in clinical practice.

### Conclusion

The aim of this scoping review was to give an overview of the current literature of CDSSs in RDs. The study has identified several CDSS, using different functionalities and data to support the diagnosis of RDs (e.g. analysis and comparison of phenotypic or genetic data). We have noticed that most of the CDSS are fully developed systems, which means, that they can be downloaded or used online by the clinicians. Most of the CDSS can be used for all possible diseases.

However, several improvements in the systems are useful. For instance, studies should focus on data integration to allow automatic data transfer from other

systems like EHR. CDSSs developers should provide regular updates to keep the knowledge base of their CDSSs up to date.

In summary, this study shows an important overview of which CDSSs are available, by including clinical prototypes but also full developed systems, which should be interesting for developers of new CDSSs and clinicians. In the end, clinicians have to decide which system can be used for which purpose and at what stage of the diagnosis process, based on their experience and the respective patient case. Looking ahead, it remains interesting which of the CDSSs will be further developed and actively used. It is also considered important to involve clinicians in the development of the CDSSs and investigate the diagnostic success and clinical usage of CDSS in further studies.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s13023-020-01536-z>.

**Additional file 1.** PRISMA-ScR checklist.

**Additional file 2.** Search process.

**Additional file 3.** Screening forms.

**Additional file 4.** Data charting form.

**Additional file 5.** Final relevant publications.

**Additional file 6.** Overview of REST-APIs.

## Abbreviations

API: Application Programming Interface; CDSS: Clinical Decision Support System; CPU: Central Processing Unit; CTX: Cerebrotendinous xanthomatosis; EHR: Electronic Health Record; ETL: Extract, Transform, Load; EU: European Union; FHIR: Fast Healthcare Interoperability Resources; GARD: Genetic and Rare Diseases Information Center; GPU: Graphics Processing Unit; HPO: Human Phenotype Ontology; ICD: International Statistical Classification of Diseases and Related Health Problems; IR: Information retrieval; MeSH: Medical Subject Heading; ML: Machine learning; MME: Matchmaker Exchange Project; NLP: Natural Language Processing; OBO: Open Biological Ontologies; OMIM: Online Mendelian Inheritance in Man; PRISMA-ScR: Preferred Items for Systematic Reviews and Meta Analyses extension for Scoping Reviews; RD: Rare Disease; REST: Representational State Transfer; UMLS: Unified Medical Language System; VCF: Variant Call Format; VSM: Vector Space Model; WES: Whole-exome Sequencing; WGS: Whole-Genome Sequencing; WHO: World Health Organization; XML: Extensible Markup Language

## Acknowledgements

Not applicable.

## Authors' contributions

JAS and MS designed the review and phrased the research questions. JAS conducted the search in PubMed and performed the study selection, data charting and analysis. JAS summarized and reported the results of this scoping review. HS and JOS advised the process by reviewing the work of JAS and MS. The first draft of the article was written by JAS, and MS reviewed this process. HS and JOS revised the article and provided valuable input and comments. The final manuscript was written by JAS and approved by all authors.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Medical Informatics Group (MIG), University Hospital Frankfurt, Frankfurt, Germany. <sup>2</sup>Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine Technische Universität Dresden, Dresden, Germany.

Received: 1 March 2020 Accepted: 7 September 2020

Published online: 24 September 2020

## References

- Nabarette H, Oziel D, Urbero B, Maxime N, Ayme S. Use of a directory of specialized services and guidance in the healthcare system: the example of the Orphanet database for rare diseases. *Rev Epidemiol Sante Publique*. 2006;54:41–53.
- World Health Organization. Priority diseases and reasons for inclusion. 2013. [https://www.who.int/medicines/areas/priority\\_medicines/Ch6\\_19Rare.pdf](https://www.who.int/medicines/areas/priority_medicines/Ch6_19Rare.pdf).
- Taruscio D, Florida G, Salvatore M, Groft SC, Gahl WA. Undiagnosed diseases: Italy-US collaboration and international efforts to tackle Rare and common diseases lacking a diagnosis. *Adv Exp Med Biol*. 2017;1031:25–38.
- Evans WR, Rafi I. Rare diseases in general practice: recognising the zebras among the horses. *Br J Gen Pract*. 2016;66:550–1.
- Guillem P, Cans C, Robert-Gnansia E, Aymé S, Jouk P. Rare diseases in disabled children: an epidemiological survey. *Arch Dis Child*. 2008;2:115–8.
- Zuryski Y, Frith K, Leonard K, Elliot E. Rare childhood diseases: how should we respond? *Arch Dis Child*. 2008;93:1071–4.
- Denis A, Mergaert L, Fostier C, Cleemput I SC. A comparative study of European rare disease and orphan drug markets. *Health Policy*. 2010;97:173–9.
- Griffon N, Schuers M, Dhombres F, Merabti T, Kerdelhue G, Rollin L, et al. Searching for rare diseases in PubMed: a blind comparison of Orphanet expert query and query based on terminological knowledge. *BMC Med Inform Decis Mak*. 2016;16:101.
- Rare Disease UK. What is a Rare Disease. 2018. <https://www.raredisease.org.uk/what-is-a-rare-disease>. Accessed 20 Sept 2019.
- McCarthy C, Lara Gallego B, Trapnell BC, McCormack FX. Epidemiology of Rare lung diseases: the challenges and opportunities to improve research and knowledge. *Adv Exp Med Biol*. 2017;1031:419–42.
- Ng L, Khan F, Young CA, Galea M. Symptomatic treatments for amyotrophic lateral sclerosis/motor neuron disease. *Cochrane Database Syst Rev*. 2017. <https://pubmed.ncbi.nlm.nih.gov/28072907/>.
- Hunt D, Haynes R, Hanna S, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA*. 1998;280:1339–46.
- Muller T, Jerrentrup A, Schafer JR. Computer-assisted diagnosis of rare diseases. *Internist*. 2018;59:391–400.
- Svenstrup D, Jorgensen HL, Winther O. Rare disease diagnosis: a review of web search, social media and large-scale data-mining approaches. *Rare Dis Austin Tex*. 2015;3:e10831.
- Geschäftsstelle des Nationalen Aktionsbündnisses für Menschen mit Selteneren Erkrankungen (NAMSE). National action league for people with rare diseases. 2010. <http://www.namse.de/images/stories/Dokumente/Aktionsplan/national%20plan%20of%20action.pdf>. Accessed 11 Jan 2019.
- Department of Health. The UK Strategy for Rare Diseases. 2013. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/260562/UK\\_Strategy\\_for\\_Rare\\_Diseases.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/260562/UK_Strategy_for_Rare_Diseases.pdf). Accessed 11 Sept 2019.
- Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169:467–73.

18. Shen F, Liu S, Wang Y, Wen A, Wang L, Liu H. Utilization of electronic medical records and biomedical literature to support the diagnosis of Rare diseases using data fusion and collaborative filtering approaches. *JMIR Med Inform.* 2018;6:e11301.
19. Pearce F, Lanyon PC, Watts RA. Can prediction models in primary care enable earlier diagnosis of rare rheumatic diseases? *Rheumatol Oxf Engl.* 2018;57:2065–6.
20. Sidiropoulos K, Glotsos D, Kostopoulos S, Ravazoula P, Kalatzis I, Cavouras D, et al. Real time decision support system for diagnosis of rare cancers, trained in parallel, on a graphics processing unit. *Comput Biol Med.* 2012;42:376–86.
21. McKeever L, Nguyen V, Peterson SJ, Gomez-Perez S, Braunschweig C. Demystifying the search button: a comprehensive PubMed search strategy for performing an exhaustive literature review. *JPEN J Parenter Enteral Nutr.* 2015;39:622–35.
22. Kohler S, Schulz MH, Krawitz P, Bauer S, Dolken S, Ott CE, et al. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Hum Genet.* 2009;85:457–64.
23. Hamosh A, Sobreira N, Hoover-Fong J, Sutton VR, Boehm C, Schiettecatte F, et al. PhenDB: a new web-based tool for the collection, storage, and analysis of phenotypic features. *Hum Mutat.* 2013;34:566–71.
24. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput Biol.* 2013;9:e1003153.
25. Kirkpatrick BE, Riggs ER, Azzariti DR, Miller VR, Ledbetter DH, Miller DT, et al. GenomeConnect: matchmaking between patients, clinical laboratories, and researchers to improve genomic knowledge. *Hum Mutat.* 2015;36:974–8.
26. Rother A-K, Schwerk N, Brinkmann F, Klawonn F, Lechner W, Grigull L. Diagnostic support for selected Paediatric pulmonary diseases using answer-pattern recognition in questionnaires based on combined data mining applications—a monocentric observational pilot study. *PLoS One.* 2015;10:e0135180.
27. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat.* 2015;36:928–30.
28. Gottlieb MM, Arenillas DJ, Maithripala S, Maurer ZD, Tarailo Graovac M, Armstrong L, et al. GeneYenta: a phenotype-based rare disease case matching tool based on online dating algorithms for the acceleration of exome interpretation. *Hum Mutat.* 2015;36:432–8.
29. Koile D, Cordoba M, Sousa Serro M, Kauffman MA, Yankilevich P. GeneIO: a phenotype-genotype analysis web server for clinical genomics of rare diseases. *BMC Bioinformatics.* 2018;19:25.
30. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Budro M, et al. The Matchmaker exchange: a platform for rare disease gene discovery. *Hum Mutat.* 2015;36:915–21.
31. Pontikos N, Yu J, Moghul I, Withington L, Blanco-Kelly F, Vulliamy T, et al. Phenopolis: an open platform for harmonization and analysis of genetic and phenotypic data. *Bioinforma Oxf Engl.* 2017;33:2421–3.
32. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chenier S, et al. PhenoTips: patient phenotyping software for clinical and research use. *Hum Mutat.* 2013;34:1057–65.
33. Bragin E, Chatzimichali EA, Wright CF, Hurler ME, Firth HV, Bevan AP, et al. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res.* 2014;42(Database issue):D993–1000.
34. Buske OJ, Girdea M, Dumitriu S, Gallinger B, Hartley T, Trang H, et al. PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Hum Mutat.* 2015;36:931–40.
35. Grigull L, Lechner W, Petri S, Kollwe K, Dengler R, Mehmecke S, et al. Diagnostic support for selected neuromuscular diseases using answer-pattern recognition and data mining techniques: a proof of concept multicenter prospective trial. *BMC Med Inform Decis Mak.* 2016;16:31.
36. Garcelon N, Neuraz A, Benoit V, Salomon R, Kracker S, Suarez F, et al. Finding patients using similarity measures in a rare diseases-oriented clinical data warehouse: Dr. warehouse and the needle in the needle stack. *J Biomed Inform.* 2017;73:51–61.
37. Dragusin R, Petcu P, Lioma C, Larsen B, Jorgensen HL, Cox IJ, et al. FindZebra: a search engine for rare diseases. *Int J Med Inf.* 2013;82:528–38.
38. Taboada M, Rodriguez H, Martinez D, Pardo M, Sobrido MJ. Automated semantic annotation of rare disease cases: a case study. *Database J Biol Databases Curation.* 2014;2014:1–13.
39. Deo RC. Machine learning in medicine. *Circulation.* 2015;132:1920–30.
40. Fraccaro P, O'Sullivan D, Plastiras P, O'Sullivan H, Dentone C, Di Biagio A, et al. Behind the screens: clinical decision support methodologies - a review. *Health Policy and Technology.* 2015;4:29–38.
41. Baeza-Yates R, Ribeiro-Neto B. Modern information retrieval. 2nd ed. New York: Addison-Wesley; 1999.
42. Nadkarni PM. An introduction to information retrieval: applications in genomics. *Pharmacogenomics J.* 2002;2:96–102.
43. Yang H, Robinson PN, Wang K. Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. *Nat Methods.* 2015;12:841–3.
44. Mark M, API REST. Design Rulebook: Designing Consistent RESTful Web Service Interfaces. Newton: O'Reilly Media; 2011.
45. Lysaght T, Lim HY, Xafis V, Ngiam KY. AI-assisted decision-making in healthcare. *Asian Bioeth Rev.* 2019;11:299–314.
46. Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Pantelis G, Lescure F-X, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin Microbiol Infect.* 2019.
47. Safdar S, Zafar S, Zafar N, Khan NF. Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. *Artif Intell Rev.* 2018;50:597–623.
48. Gadd C, Baskaran P, Lobach D. Identification of design features to enhance utilization and acceptance of systems for internet-based decision support at the point of care. *Proc AMIA Symp.* 1998;91–5. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2232383/>.
49. Zheng K, Padman R, Johnson MP, Diamond HS. Understanding technology adoption in clinical care: clinician adoption behavior of a point-of-care reminder system. *MedInfo.* 2005;2004(74):535–43.
50. El-Sappagh SH, El-Masri S. A distributed clinical decision support system architecture. *J King Saud Univ - Comput Inf Sci.* 2014;26:69–78.
51. Kopanitsa G. Integration of hospital information and clinical decision support systems to enable the reuse of electronic health record data. *Methods Inf Med.* 2018;56:238–47.
52. Abouzahra M, Sartipi K, Tan J. Integrating Data from EHRs to Enhance Clinical Decision Making: The Inflammatory Bowel Disease Case. In: 2014 IEEE 27th International Symposium on Computer-Based Medical Systems; 2014. p. 531–2.
53. Parimbelli E, Sacchi L, Bellazzi R. Decision support through data integration: strategies to meet the big data challenge. *Eur J Biomed Inform.* 2016;12:10–4.
54. Huang M, Han H, Li L, Zhang Y, Bhatti UA. A clinical decision support framework for heterogeneous data sources. *IEEE J Biomed Health Inform.* 2018;22:1824–33.
55. Wasylewicz A, Scheepers-Hoeks A. Clinical decision support systems. In: Fundamentals of clinical data science. Cham: Springer; 2018. <https://www.ncbi.nlm.nih.gov/books/NBK543516>.
56. Aljarboa S, Miah S, Kerr D. Perceptions of the adoption of clinical decision support Systems in the Saudi Healthcare Sector; 2019.
57. Osheroff J, Teich J, Middleton B, Steen E, Wright A, Detmer D. A roadmap for National Action on clinical decision support. *J Am Med Inform Assoc JAMIA.* 2007;14:141–5.
58. Marcy TW, Kaplan B, Connolly SW, Michel G, Shiffman RN, Flynn BS. Developing a decision support system for tobacco use counselling using primary care physicians. *Inform Prim Care.* 2008;16:101–9.
59. Brunner J, Chuang E, Goldzweig C, Cain CL, Sugar C, Yano EM. User-centered design to improve clinical decision support in primary care. *Int J Med Inf.* 2017;104:56–64.
60. Stanzola E, Uznayo M, Simón M, Otero C, Campos F, Luna D. User-centered Design of Health Care Software Development: towards a cultural change. *Stud Health Technol Inform.* 2015;216:368–71.
61. LeRouge C, Wickramasinghe N. A review of user-centered design for diabetes-related consumer health informatics technologies. *J Diabetes Sci Technol.* 2013;7:1039–56.
62. Chokshi SK, Belli HM, Troxel AB, Blecker S, Blaum C, Testa P, et al. Designing for implementation: user-centered development and pilot testing of a behavioral economic-inspired electronic health record clinical decision support module. *Pilot Feasibility Stud.* 2019;5:28.
63. Nguyen KA, Patel H, Haggstrom DA, Zillich AJ, Imperiale TF, Russ AL. Utilizing a user-centered approach to develop and assess pharmacogenomic clinical decision support for thiopurine methyltransferase. *BMC Med Inform Decis Mak.* 2019;19:194.
64. Horsky J, Schiff GD, Johnston D, Mercincavage L, Bell D, Middleton B. Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. *J Biomed Inform.* 2012;45:1202–16.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.