How to Design a Registry for Undiagnosed Patients in the Framework of Rare Disease Diagnosis – Suggestions on Software, Data Set and Coding System

Alexandra Berger (alexandra.berger@kgu.de)
Goethe-Universitat Frankfurt am Main Fachbereich 16 Medizin https://orcid.org/0000-0002-0982-6613
Anne-Kathrin Rustemeier
University Hospitals Giessen and Marburg Campus Giessen: Universitatsklinikum Giessen und Marburg GmbH Standort Giessen
Jens Göbel
Hospital of the Goethe University Frankfurt: Klinikum der Johann Wolfgang Goethe-Universitat Frankfurt
Dennis Kadioglu
Hospital of the Goethe University Frankfurt: Klinikum der Johann Wolfgang Goethe-Universitat Frankfurt
Vanessa Britz
Goethe-Universitat Frankfurt am Main Fachbereich 16 Medizin
Katharina Schubert
University Hospital Magdeburg: Universitatsklinikum Magdeburg
Klaus Mohnike
University Hospital Magdeburg: Universitatsklinikum Magdeburg
Holger Storf
Hospital of the Goethe University Frankfurt: Klinikum der Johann Wolfgang Goethe-Universitat Frankfurt
Thomas OF Wagner
Goethe University Frankfurt Faculty 16 Medicine: Goethe-Universitat Frankfurt am Main Fachbereich 16 Medizin

---

Research

Keywords: registry, undiagnosed patients, rare diseases, HPO

Posted Date: December 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-122656/v1

Abstract

Background: About 30 million people in the EU and USA, respectively, suffer from a rare disease. Driven by European legislative requirements, national strategies for the improvement of care in rare diseases are being developed. To improve timely and correct diagnosis for patients with rare diseases, the development of a registry for undiagnosed patients was recommended by the German National Action Plan. In this paper we focus on the question on how such a registry for undiagnosed patients can be built and which information it should contain.

Results: To develop a registry for undiagnosed patients, a software for data acquisition and storage, an appropriate data set and an applicable terminology/classification system for the data collected is needed. We used the open source software OSSE (Open Source Registry System for Rare Diseases) to build the registry for undiagnosed patients. Our data set is based on the minimal data set for rare disease patient registries recommended for European Rare Disease Registries Platform. We extended this Common Data Set to also include symptoms, clinical findings and other diagnoses. In order to ensure findability, comparability and statistical analysis, symptoms, clinical findings and diagnoses have to be encoded. We evaluated three medical ontologies (SNOMED – CT, HPO and LOINC) for their usefulness. With exact matches of 98% of tested medical terms, a mean number of five deposited synonyms, SNOMED CT seemed to fit our needs best. HPO and LOINC provided 73% and 31% of exacts matches of clinical terms respectively. Allowing more generic codes for a defined symptom, with SNOMED-CT 99%, with HPO 89% and with LOINC 39% of terms could be encoded.

Conclusions: With the use of the OSSE software and a data set, which, in addition to the Common Data Set, focuses on symptoms and clinical findings, a functioning and meaningful registry for undiagnosed patients can be implemented. The next step is the implementation of the registry in centres for rare diseases. With the help of medical informatics and big data analysis, case similarity analyses could be realized and aid as a decision-support tool enabling diagnosis of some patients.

Background

Rare diseases are defined by prevalence. In the European Union (EU), a disease is considered to be rare, if it affects less than 5 of 10,000 people. In the United States of America (USA) a rare disease is defined as affecting less than 200,000 inhabitants, translating to a prevalence of about 8–9 out of 10,000 people.(1) About 30 million people in both the EU and the USA are suffering from a disease that is considered a rare disease.(2, 3)

Orphanet is a 37-country network, aiming to increase knowledge of rare diseases. It was cofounded by the European Commission in 1997. As of 2020, classification and descriptions of 6,172 rare diseases are included in the Orphanet database; 71.9% being genetic and the onset of symptoms occuring in 69.9% during childhood. About 85% of rare diseases are ultra-rare with a prevalence of less than 1 per 1,000,000. (1)

Many rare diseases are severe chronic conditions with a complex clinical presentation and a negative impact on life expectancy and quality of life.(4) Prevention and cure as well as adequate therapies exist only for a minority of rare diseases.(5)

Implications for Patients

Patients with rare diseases face a multitude of disease-related problems. Starting with delayed diagnosis, multiple doctor's visits before a diagnosis is made, misleading diagnosis, lack of comprehensive information provided at the time of diagnosis, insufficient coordination of care, inadequate transition from paediatric to adult care, and low or non-existent access to medication due to poor knowledge or lacking research and clinical trials. Patient organizations play a vital role in improving these circumstances.(6, 7)

Implications for Physicians

Due to the large number of rare diseases it is impossible for a single doctor to be familiar with all of them. Especially general practitioners, who are the first contact for many patients, feel insufficiently trained in detecting rare diseases and often lack close interdisciplinary collaboration.(8)

Implications for Researchers

For researchers, one of the main obstacles is to include an adequate number of patients in clinical trials. This is a problem with any rare disease but especially affects the ultra-rare diseases. To improve this situation, collaboration of multiple centres of expertise nationwide, in some cases internationally, is needed. Therefore, research networks have started to coordinate research projects and implement highly standardized structures of data collection and sharing.(2, 9, 10)

Implications for the Economy

Patients with rare diseases generate a lot of health care expenditures. Unnecessary costs occur especially during the time before a diagnosis is established: Multiple health care contacts over a period of up to 30 years have been documented.(7) Inadequate utilization of costly therapies due to incomplete diagnosis or false indication is another reason for waste of resources.

Actions for Rare Diseases

Initiated by patient organizations, rare diseases have gained attention in politics over the last decade. Driven by European legislative requirements, national strategies for the improvement of care in rare diseases had to be developed.(11) In Germany, the National Action Plan for People with Rare Diseases implemented 52 measures to improve health care for patients with rare diseases. Some examples are: Recommendations for the implementation of national centres of expertise, specific measures to accelerate time to diagnosis, research support, improvement of information management as well as suggestions on financing of these measures.(12)

Concerning research, the development of a registry toolbox for creating individual disease-specific registries was requested. This registry toolbox should make use of an open source software with a defined minimal data scheme and an emphasis on interoperability on a national and international level as well as metadata management.(12) This project was conducted collaboratively by the Institute of Medical Biometrics, Epidemiology and Informatics of the University Medical Center of the Johannes Gutenberg University Mainz and the University Hospital Frankfurt in 2013 as part of the German National Action Plan and yielded the “Open Source Registry System for Seltene Erkrankungen (OSSE)”. OSSE is an easily scalable and customizable template for developing disease specific rare disease registries automatically connected to a meta data repository and fulfilling FAIR criteria. Further development is since ongoing by the Medical Informatics Group (MIG) of the University Hospital Frankfurt.(12–15)

To improve timely and correct diagnosis for patients with rare diseases, the development of a ‘registry for undiagnosed patients’ was also recommended by the German National Action Plan, taking into account that a high percentage of these ‘undiagnosed patients’ eventually are diagnosed to have a rare disease. (12)

Similar National strategies have been developed in most member states of the European Union as well as Norway, Switzerland and the UK.(16) Some international examples are: The National Institutes of Health Undiagnosed Diseases Program, which started in 2008(17); the “Nan-Byo” (which translates as “difficult and illness”), which was established in 1972 in Japan and extended in 2015 as Japan’s Initiative on Rare and Undiagnosed Diseases(18); Just recently, in February 2020, the Australian government announced to provide funding for activities to implement the National Strategic Action Plan for Rare Diseases, which was developed by Rare Voices Australia.(19)

Registries for Rare Diseases

Registries in general and especially in the field of rare diseases can help to connect data from multiple health care providers (HCP), thus enlarging the data base for research questions, including epidemiology of rare diseases. However, disease-specific ICD-10 codes are not available for most rare diseases and Orpha-codes, OMIM-codes or alpha-IDs are not used in routine clinical care. Therefore, prevalence calculated from disease-specific registries have limited accuracy.(20, 21) And, of course, usually academia driven registries do not achieve sufficient representation of the whole disease population to allow calculation of prevalence.

Due to the fact that undiagnosed patients present with a wide variety of symptoms at different levels and specialities within the health care system, it is even more complicated to assess the number of undiagnosed patients. With a nationwide registry, the impact of undiagnosed patients on the health care system could be estimated. In addition, it could accompany patients on their way to diagnosis, pointing out structural problems in the health care system. As soon as a patient is diagnosed and agrees to data-sharing, the collected data set could be transferred to a disease-specific registry, if such a registry exists. This would help in gaining patients and data for disease-specific research questions as well as connecting different centres of expertise to work together more closely.

With the help of medical informatics and big data analysis, case similarity analysis could be realized and aid as a decision-support tool probably facilitating diagnosis for some patients.

As most medical registries focus on one specific disease or group of diseases, they contain disease-specific and disease-relevant data. Patients, who are not yet diagnosed do not fit into these registry schemes. Therefore, in this paper we focus on the question on how such a registry for undiagnosed patients can be built and which information it should contain.

Methods

To develop a registry for undiagnosed patients, a software for data acquisition and storage, an appropriate data set and an applicable terminology/classification system for the data used is needed.

Registry Software

We have used the open source software OSSE(13) as a framework for the registry for undiagnosed patients. This software enables users - even with limited IT-knowledge - to create registry data schemes for the individual purpose. The data items are specified as data elements in a metadata storage, where they can be retrieved to be re-used as templates in future registries. This openly accessible metadata, and the possibility for researchers to get an impression of the data a registry collects, without forcing the registry to centrally disclose their data allow for a wide interoperability with other registries and research facilities. This in turn enables the researcher on the one hand to decide, which registries can provide appropriate data and on the other hand formulate a detailed inquiry for data using a so called OSSE decentral search inquiry.(15)

Another strong point of the OSSE software is data protection. For pseudonymization, OSSE uses a broadly established open source software, Mainzelliste, developed by the University Medical Center of the Johannes Gutenberg University Mainz.(22, 23) OSSE also offers templates for patient information and declaration of consent. For further information on the software see https://www.osse-register.de/en/.(13)

Data Set and Coding

The basis of our work was the minimal data set for rare disease patient registries recommended for European cooperation Version 3.0 (see. Table 1). Version 3.0 differs only in minor aspects from Version 0.1 (24), which itself has been built based on the French minimal data set *RD MDS v1.08*(25).

Table 1
Data Set for Rare Disease Patient Registries Recommended for European Cooperation (Version 3.0), based on the French minimal data set RD MDS v1.08

Item Group	Item No.	Item Concept	Question	Content Coding	Data Collection (one-time/ repreatedly)	Comment
1. Pseudonym	1.1	Patient's Pseudonym (PID)	Patient's Pseudonym (as definied in the meta-data-set)	String	One-time	
2. Personal Information	2.4	Patient's date of birth	Patient's date of birth as recorded on the birth certificate	Date	One-time	
	2.5	Gender	Patient's gender	Female Male Undetermined Unknown (for the foetus)	One-time	
3. Family Information	3.1	Patient born from a relationship between related parties	Is the patient born from a relationship between related parties	Yes No Unknown	One-time	
4. Vital status	4.1	Patient's vital status	Is the patient still alive?	Yes No Lost to follow-up Discharged from registry	Repeatedly	Update of the data base at least once per year
	4.2	Patient's date of death	Patient's date of death	Date	One-time	Update of the data base at least once per year

Item Group	Item No.	Item Concept	Question	Content Coding	Data Collection (one-time/ repreatedly)	Comment
	4.3	Death due to rare disease	Is the death due to the rare disease the patient is suffering from?	Yes No Unknown		
5. Care pathway	5.1	Patient's date of inclusion in the registry	Date at which the patient was included in the registry	Date		
Item Group	Item No.	Item Concept	Question	Content Coding	Data Collection (one-time/ repreatedly)	Comment
6. Disease history	6.1	Age at onset	Age at which symptoms first appeared	Antenatal At birth XX year(s) and XX month(s) Undetermined		
	6.2	Age at diagnosis	Age at which the diagnosis was made	Antenatal At birth XX year(s) and XX month(s) Undetermined		
7. Diagnosis	7.2	Diagnosis of the rare disease	Diagnosis retained by the RD center	Alpha code		
8. Research	8.1	Agreement to be contacted for a protocol	Does the patient give permission to be contacted for a research protocol?	Yes No		

Item Group	Item No.	Item Concept	Question	Content Coding	Data Collection (one-time/ repreatedly)	Comment
	8.2	Patient non-opposition to the reuse of data	Does the patient give permission for his / her data to be reused for other research purposes?	Yes No		
	8.3	Patient having previously given a biological sample for research	Has the patient already given a biological sample for research?	Yes No		
	8.4	Patient having previously given a biological sample for molecular diagnosis	Has the patient already given a biological sample for molecular diagnosis?	Yes No		

As rare diseases are heterogenous and complex in their clinical presentation, we decided to extend the minimal data set by (subjective) symptoms and (objective) clinical findings together with the time of their first presentation as well as established or suspected diagnoses in this patient. In order to ensure comparability and statistical evaluation, symptoms, clinical findings and diagnoses have to be encoded.

As the ICD-10 code does not have sufficient specificity and granularity for rare diseases, we included the Alpha ID(26) and Orpha Code(27) to encode rare diseases in the registry for undiagnosed patients.

We evaluated three terminologies resp. ontologies (in the following coding systems), i.e. Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT)(28), Human Phenotype Ontology (HPO)(29) and Logical Observation Identifiers Names and Codes (LOINC)(30) with regard to usefulness and feasibility for a registry for undiagnosed patients.

The Frankfurt Reference Centre for Rare Diseases (FRZSE), among other activities, runs a students' clinic for patients without a diagnosis, where patient cases are discussed in interdisciplinary teams to eventually find a diagnosis.

We used 10 random patient files from this students' clinic to evaluate the developed data set and to compare the 3 different coding systems. All identifying data of the files were removed, thus ensuring data

protection by anonymization.

All symptoms, diagnoses and clinical findings mentioned in the medical records were extracted and translated into English. Each symptom was extracted only once, regardless of how often it was mentioned in the file. Overall 80 medical terms were extracted. These terms were entered in the browsers of each of the 3 coding systems. In some cases, more than one possible translation was entered into the browser to increase the chance of a match.

We evaluated the classification systems by numbers/percentages of matches for the extracted medical terms. A term could be an exact match, a more general match, a match that is too specific, a match for which a quantitative figure is needed, or no match.

Results

Evaluation of coding systems

With SNOMED CT, 98% of the 80 medical terms could be coded correctly. For one term, *increase in waist circumference*, an exact quantitative information is mandatory. Elevated gamma-glutamyltransferase levels could not be coded with SNOMED CT. For every medical term that could be coded correctly a mean number of 5 synonyms (range 2–15) was provided by SNOMED CT.

HPO provided exact matches for 73% of the terms. More general codes are available for 16% and codes that were too specific for 3% of the terms. 9% of the medical terms could not be coded. Only about 2 synonyms for each term are available so that further synonyms had to be entered manually to increase finding matches.

Only 31% of medical terms were coded correctly by the LOINC nomenclature. Generic codes are available for 8%, too specific ones for 33% of the terms. 8% of the medical terms could only be coded with an exact quantitative measurement. 21% of medical terms could not be coded at all. Therefore, symptoms and clinical findings cannot be coded sufficiently with LOINC. Lab values can only be represented with their exact value. Basic changes in lab values, for example hyponatraemia, cannot be described with LOINC. Medical terms and their synonyms had to be entered manually in most instances because LOINC only provides a mean of one synonym per term.

Under the assumption that also more generic coding terms are acceptable, 99% of medical terms are matched with SNOMED CT, 89% with HPO and 39% with LOINC (see Fig. 1).

Data Set

The development of the data set for the undiagnosed patients' registry was based on the minimal data set for rare disease patient registries recommended for European cooperation Version 3.0.(24) (see Table 1) This data set had been developed for the EUCERD-Action of the European Commission and has

been the basis for the Set of Common Data Elements for Rare Diseases Registration, which was released by the EU Rare Disease platform with only some minor changes.(31)

We extended the minimal data set by symptoms, clinical findings and prior diagnoses and generated two forms: a basic form, which is filled out initially when a patient is included in the registry and an episodic or longitudinal form, which can be filled out several times. (see Table 2) The basic form can only be altered by an administrator once the data set has been saved. The episodic form functions as a follow-up tool.

Table 2
Suggested data Set for a registry for undiagnosed patients

Item Group	Item No.	Item Concept	Concept Description	Content Coding
1. Personal data / information	1.1	Patient's first name	Patient's first name as specified on the birth certificate or identity card	
	1.2	Patient's (married) last name	Patient's (married) last name	
	1.3	Patient's birth name	Patient's birth name	
	1.4	Patient's date of birth	Patient's date of birth as recorded on the birth certificate and whether Information is recorded for a foetus	
	1.5	City of residence	Patient's current city of residence	
	1.6	Gender	Patient's gender	Female Male Undetermined Unknown (for the foetus)
2. Family information	2.1	Patient born from a relationship between related parties	Is the patient born from a relationship between related parties?	Unknown No, suspected No, conformed Yes, suspected Yes, confirmed
3. Vital Status	3.1	Patient's vital status upon inclusion into the registry	Is the patient still alive?	Alive Dead
	3.2	Patient's date of death	Patient's date of death	
4. Care Pathway	4.1	Patient's date of inclusion in the RD centre	Date at which the patient was recorded in the RD center. Please enter the date when the patient was included in the internal medical information system	

Item Group	Item No.	Item Concept	Concept Description	Content Coding
5. Disease history	5.1	Point in time at onset	when were the symptoms first noticed? (Only fill date input fields if "lifetime" was selected)	Antenatal At birth Undetermined lifetime '*Year of first manifestation*' *'Month of first manifestation'* *'Day of first manifestation'*
Item Group	Item No.	Item Concept	Concept Description	Content Coding
6. Diagnosis upon inclusion into the registry	6.1	Diagnosis code	Prior Diagnosis (Code)	
	6.2	Type of code	Specify which type of code is used - use ICD-10 if possible-	ICD-10 Alpha-ID Orphacode
	6.3	Description of the chosen Code	Description of the chosen code. Please copy the EXACT text belonging to the code. Do NOT enter free text.	
	6.4	Status of diagnosis	Specify whether the diagnosis is already confirmed or only suspected.	Unknown Confirmed Suspected
7. Symptom history	7.1	Diagnosis code (Symptom)	Diagnosis (code) of the symptom	
	7.2	Type of code (Symptom)	Specify which type of code is used - use HPO if possible-	HPO SNOMED-CT
	7.3	Symptom Ontology description	Description from the selected code.	

Item Group	Item No.	Item Concept	Concept Description	Content Coding
	7.4	Symptom priority	Symptom priority	Unknown High Medium Low
	7.5	Year of first manifestation	Year of first manifestation	
	7.6	Month of first manifestation	Month of first manifestation	Unknown January February March April May June July August September October November December
Item Group	**Item No.**	**Item Concept**	**Concept Description**	**Content Coding**
E1. Vital Status	E1.1	Patient's vital status	Is the patient still alive	Alive Dead Lost to follow up Discharged from registry
	E1.2	Patient's date of death	Patient's date of death	

Item Group	Item No.	Item Concept	Concept Description	Content Coding
	E1.3	Death due to the rare disease	Is the death due to the rare disease the patient is suffering from?	Yes No Unknown
E2. Further Symptoms		Diagnosis code (Symptom)	Diagnosis (code) of the symptom	
	E2.1	Type of code (Symptom)	Specify which type of code is used - use HPO if possible-	HPO SNOMED-CT
	E2.2	Symptom Ontology description	Description from the selected code.	
	E2.3	Symptom priority	Symptom priority	Unknown High Medium Low
	E2.4	Year of first manifestation	Year of first manifestation	
	E2.5	Month of first manifestation	Month of first manifestation	Unknown January, February March, April May, June July, August September October November December
E3. Further Diagnosis	E3.1	Diagnosis code	Prior Diagnosis (Code)	

Item Group	Item No.	Item Concept	Concept Description	Content Coding
	E3.2	Type of code	Specify which type of code is used - use ICD-10 if possible-	ICD-10 Alpha-ID Orphacode
	E3.3	Description of the chosen Code	Description of the chosen code. Please copy the EXACT text belonging to the code. Do NOT enter free text.	
	E3.4	Status of diagnosis	Specify whether the diagnosis is already confirmed or only suspected.	Unknown Confirmed Suspected
Item Group	Item No.	Item Concept	Concept Description	Content Coding
E4. Causal Diagnosis (final)	E4.1	Diagnosis code	Diagnosis (Code)	
	E4.2	Type of code	Specify which type of code is used - use ICD-10 if possible-	ICD-10 Alpha-ID Orphacode
	E4.3	Description of the chosen Code	Description of the chosen code. Please copy the EXACT text belonging to the code. Do NOT enter free text.	
	E4.4	Status of diagnosis	Specify whether the diagnosis is already confirmed or only suspected.	Unknown Confirmed Suspected
	E4.5	Rare Disease	Is the newly found diagnosis a rare disease?	(Clickbox)
	E4.6	Year of diagnosis	Year of diagnosis	

Item Group	Item No.	Item Concept	Concept Description	Content Coding
	E4.7	Month of diagnosis	Month of diagnosis	Unknown January February March April May June July August September October November December
E5. Research	E5.1	Agreement to be contacted for a protocol	Does the patient give permission to be contacted for a research protocol?	Yes No Unknown
	E5.2	Patient non-opposition to the reuse of data	Is the patient non-opposite to the reuse of data?	Yes No Unknown
	E5.3	Patient having previously given a biological sample for research	Has the patient already given a biological sample for research?	Yes No Unknown

The basic form

Personal information such as name, surname, date of birth and current address, more precisely the postal code, is used to create the personal ID (pseudonym) with the Mainzelliste.

Gender of the patient is documented.

As over 70% of rare diseases have a genetic origin, we ask for the patient´s parents' consanguinity in the family information.

We ask for vital status upon inclusion in the registry. If the patient has already died, the date of death is to be entered, too. Date of inclusion into the registry is reported.

For the disease history, we ask for the time when the first symptoms were noticed.

Previously diagnosed diseases are prompted. These diagnoses shall be coded with an international disease classification, preferably with the ICD-10. When a more specific code is needed, for example when a coexisting rare disease is to be coded, Alpha-ID or Orpha Code, can be used. Therefore, the type of code as well as its description is to be entered as well. A statement whether the diagnosis has been confirmed or is still being suspected is asked for.

Symptoms and clinical findings are to be entered together with the type of nomenclature with which they are coded and an exact description of the term in the coding system. Year and month of the symptoms' first appearance are to be entered as well as its impact on the patient`s life.

The episodic form

Together with the basic form, the first longitudinal form has to be completed. When an episodic form is created, the data collection date has to be entered. Then vital status, date of death and if death was caused by a rare disease, if applicable, are to be filled out.

Newly developed symptoms and further diagnoses as well as those that were forgotten to be mentioned during the first data collection can be entered. The number of symptoms and diagnoses that can be entered is not limited.

If a causal diagnosis is found which potentially explains all symptoms, this information can be recorded as well. The diagnosis itself shall again be coded preferably using the ICD-10. Type and description of the code, status of diagnosis, year and month of diagnosis and whether the disease is a rare one shall be entered.

Lastly, information for research questions is to be entered. As these can change over time, only the latest episodic form is applicable. The questions concerning research are:

1. Has the patient given consent to be contacted for a study protocol?
2. Has the patient been informed concerning and not opposed to a future re-use of his/her de-identified data for other research purposes?
3. Has the patient already given a biological sample for research?

Test version

After determination of the data set and a coding system, the Medical Informatics Group Frankfurt installed a test version of the registry for undiagnosed patients. This test version was evaluated in terms of content and ease of use. Multiple users entered data of randomly selected anonymized health records of the students' clinic for patients without a diagnosis of the FRZSE. Technical problems as well as issues concerning the content were listed and discussed. The data set and user interface were optimized according to the identified problems. As a result, a tested and proven to work version of the registry for undiagnosed patients has been set up.

Discussion

To ensure comparability of 'undiagnosed patients' registry entries, findings and symptoms have to be encoded. We evaluated three medical coding systems (SNOMED – CT, HPO and LOINC) for their usefulness and feasibility.

For our purposes, finding exact matches for symptoms described by patients and clinical findings as mentioned in the health record is necessary.

With exact matches of 98% of tested medical terms, a mean number of five deposited synonyms, SNOMED CT seemed to fit our needs best.

HPO and LOINC provided 73% and 31% of exact matches of clinical terms respectively. Bringing in more generic terms for a defined symptom, with SNOMED-CT 99%, with HPO 89% and with LOINC 39% of terms could be coded. One has to consider, though, that by using more generic or too specific terms the precise meaning of a symptom can be lost. For example: Raynaud's phenomenon could be coded exactly with the SNOMED-CT code "SCTID 266261006". HPO provides only the more general codes for "cyanosis" (*HPO-Code* 0000961) or "abnormality of blood circulation" (*HPO-Code* 0011028) whereas the LOINC-code 67732-8 encodes only a very specific clinical situation, namely white finger syndrome or Raynaud's syndrome caused by excessive vibration from pneumatic hammers or drills.

With a hit ratio of 31%, LOINC was not suitable for sufficiently coding symptoms and clinical findings in our test-cases.

The HPO ontology is extended continuously.(32) So, we expect the fraction of adequate hits of HPO to improve significantly over time. HPO is widely used for deep phenotyping in the field of rare diseases. The phenotype profile can be compared with computational disease profiles in the HPO database with the aim of identifying genetic diseases with comparable phenotypic profiles. Also, HPO provides for interoperability with other ontologies and it plays a key role with the Exomizer tool, which identifies potential disease-causing variants from whole-exome or whole-genome sequencing data.(32, 33)

Taking into account HPO's acceptable match rate of medical terms, the fact that it is available free of charge and especially its wide application and interoperability in the field of rare diseases, we think that HPO is the ontology of choice for an undiagnosed patients' registry.

The European Common Data Set for Rare Disease Registration also recommends the phenotype of patients to be recorded with HPO.(31) In cases where a symptom cannot be coded adequately with HPO, a request can be sent to the developers of HPO to ask for the definition and addition of a new more suitable code to the HPO ontology for future use.

Although SNOMED CT proved to be the best fit for our needs, one limitation is the requirement of a national license, which is available in Germany only since the beginning of 2020 and for now only in the context of the Medical Informatics Initiative in Germany(34). Furthermore, the current national license and use of SNOMED CT is still undergoing evaluation. However, as the German policy clearly strives for a permanent adoption, we take SNOMED CT into account in the context of future operations and further development of our registry.

Since OSSE as a registry toolkit allows for an uncomplicated modification of the registry's forms and data elements, one of the first adjustments should be the inclusion of date of first contact with a specialized centre and a genetic diagnosis (coded by the international classification of mutations (HGVS) as suggested by the European Common Data Set for Rare disease registries. The inclusion of the patient's disability profile according to the international classification of functioning and disability does not seem practical to us, as it is far too comprehensive to be implemented into the routine service of centres for rare diseases.(31)

As OSSE is an open-source software, further developments and adjustments could be performed to meet specific needs of the distinctive nature of a registry for undiagnosed patients. Such changes could for example include statistics that would be calculated dynamically as the data base grows to show progress in the process of diagnosing patients or other key values.

Conclusions

With the use of the OSSE software and a data set which focuses on symptoms and clinical findings, a functioning and meaningful registry for undiagnosed patients can be implemented. The next step is the implementation of the registry in centres for rare diseases. The FRZSE is currently creating a retrospective registry containing the data of all its previous patients. This project will show, among other things, if the design of the registry suggested by us meets the needs in the clinical routine. After possible adjustments, we also plan to implement a multi-centre prospective registry.

Declarations

Ethics approval and consent

This study was approved by the ethics committee of the Universitätsklinikum Frankfurt.

Number: 20-26, Date: May 5th, 2020

Consent for publication

Not applicable

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors state that they have no competing interests

Funding

Not applicable

Author’s contribution

AB wrote this manuscript based on ARs thesis draft; she put the work into perspective regarding recent literature and research activities. AR extracted the medical terms from the patients’ files, encoded them and analyzed the match rates for the 3 nomenclatures. Also, she wrote her doctoral thesis on this topic. JB, DK and HS set up the registry with the OSSE-Software and further adapted the software to our needs and together with VB, KS and KM discussed and revised the manuscript.

TOFW initiated the research project, contributed substantially to the conception of the work and supervised all other contributors.

All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. European Journal of Human Genetics. 2019;28(2):165-73.

2. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, Krischer J, et al. Clinical research for rare disease: Opportunities, challenges, and solutions. Molecular Genetics and Metabolism. 2009;96(1):20-6.
3. Union RdE. Empfehlung des Rates vom 8. Juni 2009 für eine Maßnahme im Bereich seltener Krankheiten Amtsblatt der Europäischen Union Nr. C 1512009 [3. Juli 2009]. Available from: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32009H0703%2802%29:DE:HTML.
4. Frank M, Eidt-Koch D, Aumann I, Reimann A, Wagner TO, Graf von der Schulenburg JM. [Measures to improve the health situation of patients with rare diseases in Germany. A comparison with the National Action Plan]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2014;57(10):1216-23.
5. Taruscio D, Gentile AE, Evangelista T, Frazzica RG, Bushby K, Montserrat AM. Centres of Expertise and European Reference Networks: key issues in the field of rare diseases. The EUCERD Recommendations. Blood Transfus. 2014;12 Suppl 3(Suppl 3):s621-5.
6. Molster C, Urwin D, Di Pietro L, Fookes M, Petrie D, Van Der Laan S, et al. Survey of healthcare experiences of Australian adults living with rare diseases. Orphanet Journal of Rare Diseases. 2016;11(1):30.
7. Lauren Limb SN, Alev Sen. Experiences of Rare Diseases: An Insight from Patients and Families. Rare Diseases UK; 2010.
8. de Vries E, Fransen L, van den Aker M, Meijboom BR. Preventing gatekeeping delays in the diagnosis of rare diseases. Br J Gen Pract. 2018;68(668):145-6.
9. Pommerening K, Debling D, Kaatsch P, Blettner M. [Registries for rare diseases. Compliance and data protection]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2008;51(5):491-9.
10. Rodwell C, Aymé S, eds. 2014 Report on the State of the Art of Rare Disease Activities in Europe 2014 [01.07.2020]. Available from: http://www.eucerd.eu/?page_id=163#StateArt.
11. COUNCIL RECOMMENDATION of 8 June 2009 on an action in the field of rare diseases Official Journal of the European Union C1512009 [Available from: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2009:151:0007:0010:EN:PDF.
12. Nationaler Aktionsplan für Menschen mit Seltenen Erkrankungen (2019) Handlungsfelder, Empfehlungen und Maßnahmenvorschläge.: BMG, BMBF, ACHSE e.V.; [Available from: https://www.namse.de/fileadmin/user_upload/download.
13. Mainz U. OSSE - Open Source Registry System for Rare Diseases in the EU [Available from: osse-register.de.
14. Schaaf J, Kadioglu D, Goebel J, Behrendt CA, Roos M, van Enckevort D, et al. OSSE Goes FAIR - Implementation of the FAIR Data Principles for an Open-Source Registry for Rare Diseases. Stud Health Technol Inform. 2018;253:209-13.
15. Storf H, Schaaf J, Kadioglu D, Gobel J, Wagner TOF, Uckert F. [Registries for rare diseases : OSSE - An open-source framework for technical implementation]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2017;60(5):523-31.

16. EUROPLAN - European Project for Rare Diseases National Plans Development [Available from: http://www.europlanproject.eu/NationalPlans?idMap=1.
17. Gahl WA, Mulvihill JJ, Toro C, Markello TC, Wise AL, Ramoni RB, et al. The NIH Undiagnosed Diseases Program and Network: Applications to modern medicine. Mol Genet Metab. 2016;117(4):393-400.
18. Adachi T, Imanishi N, Ogawa Y, Furusawa Y, Izumida Y, Izumi Y, et al. Survey on patients with undiagnosed diseases in Japan: potential patient numbers benefiting from Japan's initiative on rare and undiagnosed diseases (IRUD). Orphanet Journal of Rare Diseases. 2018;13(1):208.
19. MP THGH, Health Mf. First National Action Plan for rare diseases 26 February 2020 [Available from: https://www.health.gov.au/ministers/the-hon-greg-hunt-mp/media/first-national-action-plan-for-rare-diseases.
20. COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS on Rare Diseases: Europe's challenges [Available from: https://ec.europa.eu/health/ph_threats/non_com/docs/rare_com_en.pdf.
21. Marx MM, Dulas FM, Schumacher KM. Verbesserung der Sichtbarkeit seltener Erkrankungen in Gesundheitssystemen durch spezifische Routinekodierung. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz. 2017;60(5):532-6.
22. Mainzelliste as an Open Source Service [Mainzelliste]. Available from: https://www.unimedizin-mainz.de/imbei/informatik/ag-verbundforschung/mainzelliste.html?L=1.
23. Lablans M, Borg A, Ückert F. A RESTful interface to pseudonymization services in modern web applications. BMC Medical Informatics and Decision Making. 2015;15(1):2.
24. ACTION EJ. MINIMUM DATA SET FOR RARE DISEASE REGISTRIES 2015 [Available from: http://www.eucerd.eu/wp-content/uploads//2015/03/WP8_Registries_MDS.pdf.
25. Choquet R, Maaroufi M, De Carrara A, Messiaen C, Luigi E, Landais P. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. Journal of the American Medical Informatics Association. 2015;22(1):76-85.
26. Thun S, Jakob R, Schlutius S, Dauben HP. Alpha-ID – Anwendungsbereiche und internationale Verwendung. 50 Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds), 12 Jahrestagung der Deutschen Arbeitsgemeinschaft für Epidemiologie; September 8, 2005; Freiburg im Breisgau,.
27. Orphanet. Orphadata - Rare Diseases and Classifications [updated 01.11.2020. Available from: http://www.orphadata.org/cgi-bin/rare_free.html.
28. Wang Y, Halper M, Wei D, Gu H, Perl Y, Xu J, et al. Auditing complex concepts of SNOMED using a refined hierarchical abstraction network. Journal of Biomedical Informatics. 2012;45(1):1-14.
29. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease. The American Journal of Human Genetics. 2008;83(5):610-5.

30. LOINC from Regenstrief: A universal code system for tests, measurements, and observations [Available from: https://loinc.org/.
31. SET OF COMMON DATA ELEMENTS FOR RARE DISEASES REGISTRATION: EUROPEAN PLATFORM ON RARE DISEASE REGISTRATION (EU RD Platform); [Available from: https://eu-rd-platform.jrc.ec.europa.eu/sites/default/files/CDS/EU_RD_Platform_CDS_Final.pdf.
32. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine JP, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res. 2019;47(D1):D1018-d27.
33. Köhler S, Øien NC, Buske OJ, Groza T, Jacobsen JOB, McNamara C, et al. Encoding Clinical Data with the Human Phenotype Ontology for Computational Differential Diagnostics. Curr Protoc Hum Genet. 2019;103(1):e92.
34. Pressemitteilung 033/2020: Digitalisierung: Medizinische Daten sprechen zukünftig eine gemeinsame Sprache [press release]. 2020.

Figures

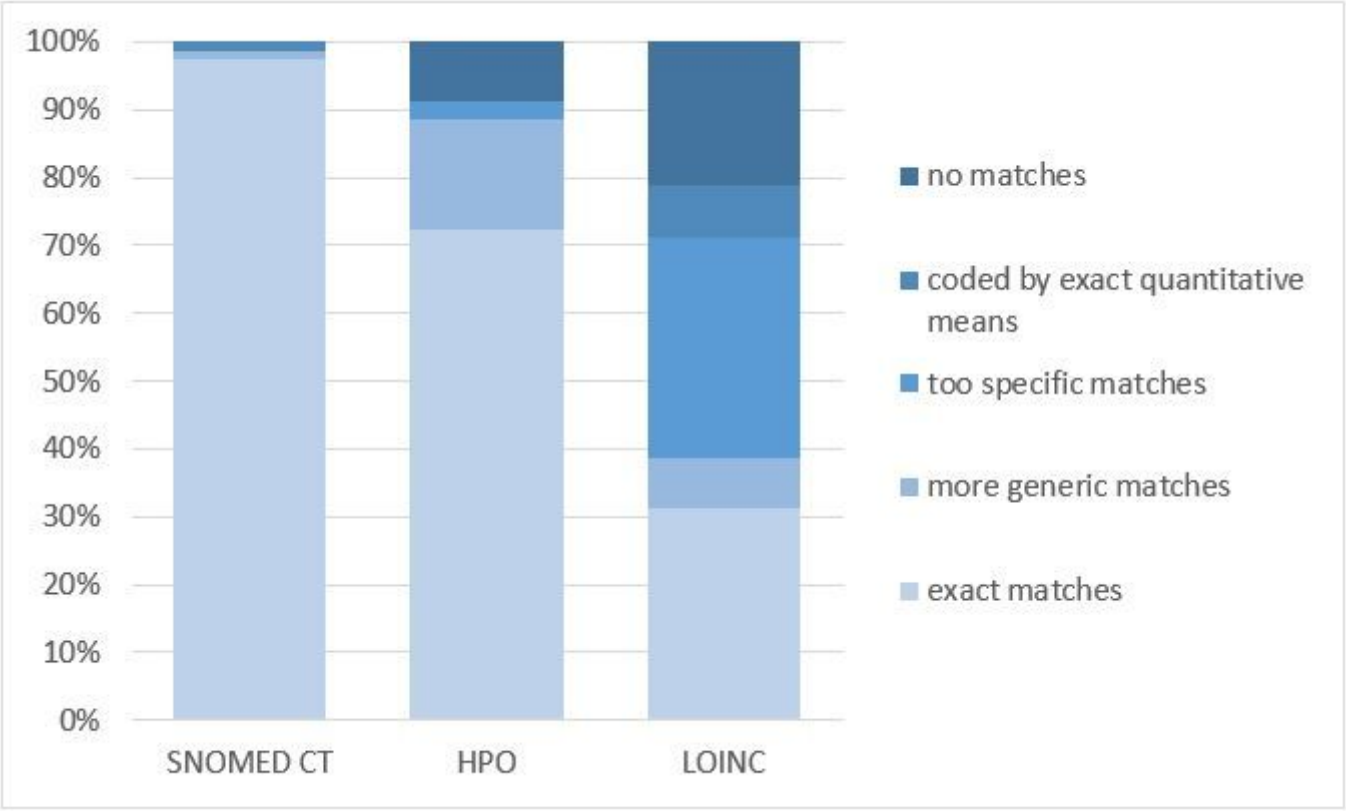

Figure 1

Evaluation of classification systems regarding matches for extracted medical terms