The Status Quo of Rare Diseases Centres for the Development of a Clinical Decision Support System – a Cross-Sectional Study

Jannik SCHAAF[a,1] Martin SEDLMAYR[b], Hans-Ulrich PROKOSCH[c], Thomas GANSLANDT[d], Carmen SCHADE-BRITTINGER[e], Michael VON WAGNER[f], Dennis KADIOGLU[a], Katharina SCHUBERT[g], Min Ae LEE-KIRSCH[h], Bernhard K. KRAEMER[i], Beate WINNER[j], Tobias MUELLER[k], Juergen R. SCHAEFER[k], Thomas O. F. WAGNER[l], Leena BRUCKNER-TUDERMAN[m], Oliver TUESCHER[n], Martin BOEKER[o] and Holger STORF[a]

[a] *Medical Informatics Group, University Hospital Frankfurt, Frankfurt, Germany*
[b] *Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technical University of Dresden, Dresden, Germany*
[c] *Chair of Medical Informatics, Department of Medical Informatics, Biometrics and Epidemiology, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany*
[d] *Heinrich-Lanz-Centre for Digital Health, Department of Biomedical Informatics, University Medicine Mannheim, Mannheim, Germany*
[e] *Chair of the Coordinating Centre for Clinical Trials, Philipps University Marburg, Marburg, Germany*
[f] *Executive Department for Medical IT-Systems and Digitalisation, University Hospital Frankfurt, Frankfurt, Germany*
[g] *Central German Competence Network for Rare Diseases, University Hospitals Magdeburg & Halle, Germany*
[h] *University Centre for Rare Diseases, University Hospital Carl Gustav Carus Dresden, Dresden, Germany*
[i] *Mannheim Centre for Rare Diseases, University Medicine Mannheim, Medical Faculty Mannheim of the University of Heidelberg, Mannheim, Germany*
[j] *Centre for Rare Diseases Erlangen, University Hospital Erlangen, Erlangen, Germany*
[k] *Centre for undiagnosed and rare diseases, University Hospital Gießen and Marburg, Marburg, Germany*
[l] *Frankfurt Reference Centre for Rare Diseases, University Hospital Frankfurt, Frankfurt, Germany*
[m] *Freiburg Centre for Rare Diseases, Medical Faculty and Medical Centre – University of Freiburg, Freiburg, Germany*
[n] *Centre for Rare Diseases of the Nervous System, University Medicine Mainz, Mainz, Germany*
[o] *Institute of Medical Biometry and Statistics, Medical Faculty and Medical Centre – University of Freiburg, Freiburg, Germany*

[1] Corresponding Author: Jannik Schaaf, Medical Informatics Group, University Hospital Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany; E-Mail: jannik.schaaf@kgu.de

Abstract. Clinical decision support systems (CDSS) help to improve the diagnostics and treatment of rare diseases (RD). As one of four funded consortia of the Medical Informatics Initiative supported by the Federal Ministry of Education and Research (BMBF, Germany), MIRACUM develops a clinical decision support system (CDSS) for RD based on distributed data of ten university hospitals. The CDSS will be developed at the Rare Diseases Centres (RDC) of the MIRACUM consortium. Since it is essential to deliver decision support at the right time and place in the clinician's workflow, this study aimed to capture relevant information of the RDCs regarding patient admission and diagnostic process. Additionally, we investigated how patient documentation and digitalisation is performed at the centres. Therefore, we conducted a cross-sectional survey involving experts in the RDs domain to capture relevant information for the further development of a CDSS in RD. For each centre, one expert on RDs participated in the study (n=8). The survey identified several challenges regarding the reuse of patient data, e.g. the paper-based documentation of a patient's medical history and coding of diagnoses using ICD-10. However, we noticed a relevant use of current software diagnosis support and a similarly performed diagnostic process in all RDC. Further studies are needed to get more detailed insights and to define specific requirements.

Keywords. rare diseases, clinical decision support, quantitative analysis

1. Introduction

In Europe, a disease is defined as "rare" if its prevalence amounts to less than 5 out of 10,000 people. There are about 7,000 different rare diseases (RDs); it is challenging to identify patients with RDs. Patients often report years or decades of a diagnosis odyssey [1]. As per definition, gathering a sufficient amount of data poses a problem for RD research. Hence, it is useful to link datasets to large research networks in order to gain new insights for research or even diagnosis support. In Germany, the German Ministry of Education and Research (BMBF) funds large research networks to create data integration centres (DICs) in context of the Medical Informatics Initiative (MI-I). The aim is to make data interoperable for research and patient care. One of these consortia is MIRACUM (Medical Informatics in Research and Care in University Medicine), which comprises ten university hospitals [2]. This data sharing approach is evaluated with different use cases, including a Clinical Decision Support System (CDSS) for RDs. Sim et al. define a CDSS as a software system which matches characteristics to a knowledge base and presents recommendations or assessments for clinical decision making to clinicians [3]. The CDSS will be developed based on the datasets of previously diagnosed cases of RDs [4]. The system will suggest a diagnosis for undiagnosed patients based on similar patients in hospital DICs throughout the consortium.

Clinicians should be involved at all stages of the development and the designing of medical systems such as a CDSS [5]. It is necessary to understand the environment in which clinicians work and how it affects the requirements of a CDSS. The users of our planned CDSS work in Rare Diseases Centres (RDCs) that belong to the hospitals of MIRACUM. These centres would benefit from a CDSS supporting the diagnostic process, as patients often visit them without a diagnosis.

This paper aims to identify the status quo of MIRACUM RDCs to gather relevant information for the further development of a CDSS supporting the diagnostic process in RD. Kawamoto has shown in a systematic review on CDSS that decision support should be provided as part of the clinician's workflow to deliver decision support at the right

time and place [6]. Furthermore, it is essential to investigate what clinical data are available to develop a CDSS.

The first objective of this paper is to investigate under which conditions a patient (contact to the centre) is diagnosed at a centre (organisation). We call this objective "Patient admission and diagnostic process". The second objective is the status quo of patient documentation for RDs (e.g. on electronic or paper or use of standards). We call this objective "Patient documentation and digitalisation".

2. Methods

To address our objectives, we conducted a cross-sectional survey. We applied the STROBE reporting guideline (Strengthening the Reporting of Observational Studies in Epidemiology) to report this study [7]. We considered 20 of 22 STROBE items. We designed the survey according to Jenn [8] with a conceptual framework based on our research objectives. Based on these two objectives, we defined 16 questions (shown in section 3) for the survey. The items 1 to 7 belong to the research objective "Patient admission and diagnostic process", whereas the items 8 to 16 belong to the research objective "Patient documentation and digitalisation". We classified each question as an open-ended question, a binary question, multiple selections or an open question. To avoid bias in the questionnaire, we applied the recommendations by Choi et al. [9]. We sent an invitation e-mail to one expert of each RDC in MIRACUM to participate in the survey. However, there is no standardised definition of the term "RD expert". Hence, experts were selected based on their experience in the field of RDs, following the advice of experts within the author's institution and based on the author's own experience. An expert in the context of this study was defined as a "member of the MIRACUM consortium with a completed degree in medical studies and completed specialist training in human medicine". A total of eight experts were contacted. The survey was conducted with a closed questionnaire published online during May 2018. The language of the survey was German. Each participant received a personal access code which was valid for four weeks.

2.1. Synthesis of results

We determined the absolute frequencies and percentages of each question for the analysis of the results. We then translated answers and questions from German to English for the synthesis of the results. For the presentation and linking of the results, we used an Entity Relationship Model (ERM) according to Chen notation [10].

3. Results

Eight out of ten university hospitals in MIRACUM have established an RDC. All participants (n=8) completed the questionnaire within four weeks. The results of the survey are shown in Table 1. The ERM is shown in the following figure.

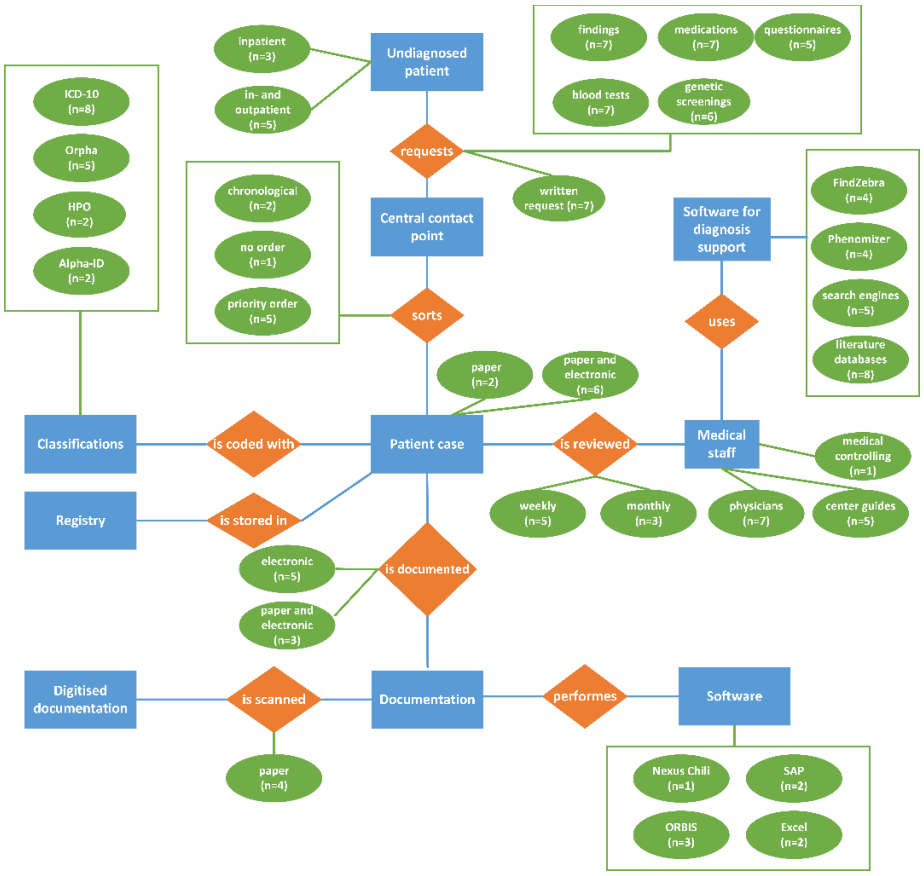

Figure 1: Entity-Relationship Model of RDCs' domain.

3.1. Description of the ERM

An undiagnosed patient contacts an RDC (entity “Undiagnosed patient”). At almost all centres (n=7), the request is entered using medical records. Patients can be accepted as outpatients in three centres. Five centres accept patients as outpatients or inpatients.

With the written request, patients send their medical history to a central contact point (entity “Central Contact Point”), which is available at all centres. Seven centres indicated that in terms of medical history, the medication of the patient is essential, followed by the findings (n=7), blood tests (n=7), genetic screening (n=6) and medical questionnaires (n=5). Afterwards, five centres perform a prioritisation of the patient cases (entity “Patient case”). Two centres process patient cases in chronological order. In contrast, only one centre does not process in a specific order.

To identify a diagnosis for a patient, a weekly (n=5) or monthly (n=3) case conference is conducted (entity “Medical staff”). The centres also use the following software for diagnosis support (entity “Software for diagnosis support”): medical literature databases (n=8), search engines (n=5), FindZebra (n=4) and Phenomizer (n=4). The following medical staff mainly use the software: physicians (n=7), centre guides (n=5), medical controlling (n=1). The centres receive the patient’s documents mainly on

paper and electronically (n=6). Two centres stated that the data are only available on paper. The documentation of a patient case is performed electronically in five centres and electronically as well as on paper in three centres (entity "Documentation"). Documentation is carried out at the centres with different software (entity "Software"): ORBIS (n=2), SAP i.s.h. med (n=2), Excel (n=2), Nexus Chili (n=1). Diagnoses are coded (entity "Classifications") with ICD-10 (n=8), followed by Orpha Number (n=5), Human Phenotype Ontology (n=2) and Alpha-ID (n=2). The records available on paper are digitised in four centres via scanning devices (entity "Digitised documentation"). In contrast, four other centres do not perform a digitalisation of paper-based records. At five centres, patient cases are recorded in a registry (entity "Registry").

Table 1. Questions and results of the survey with given answers in total and frequency

Question	Answers	Total/ Frequency
1. Is a central contact point for patients available in your centre?	Yes	8 (100%)
	No	0 (0%)
2. How many patient cases were received or treated in the centre since its foundation?	25	1 (12.5%)
	100	1 (12.5%)
	460	1 (12.5%)
	500	2 (25.0%)
	670	1 (12.5%)
	700	1 (12.5%)
	7500	1 (12.5%)
3. Inquiries from patients with an unclear diagnosis can only be submitted in writing.	Yes	7 (87.5%)
	No	1 (12.5%)
4. Please show which documents are relevant to receive from a referring doctor.	Medication	7 (87.5%)
	Blood Test	7 (87.5%)
	Findings	7 (87.5%)
	Genetic Screening	6 (75.0%)
	Others: Medical questionnaire filled out by the patient's practitioner or by the patient	5 (62.5%)
5. Is an order of processing patient cases with no diagnosis available?	Urgent cases are preferred	5 (62.5%)
	The processing is carried out according to the order of file entries	2 (25.0%)
	Others: No explicit approach for an order	1 (12.5%)
6. Do interdisciplinary regular case reviews take place in the centre to discuss patient cases? (e.g. interdisciplinary case conferences)	No	0 (0%)
	Weekly	5 (62.5%)
	1-2 times a week	0 (0%)
	3-4 times a week	0 (0%)
	Monthly	3 (37.5%)
7. How do patients get to the centre?	Inpatient only	3 (37.5%)
	Outpatient only	0 (0%)
	Inpatient and outpatient	5 (62.5%)
8. How does the centre receive the patient's documents (e.g. findings)?	Electronic	0 (0%)
	On paper	2 (25.0%)
	Both	6 (75.0%)
9. Are the transmitted patient documents digitised?	Yes, there is a digitisation	4 (50%)
	No, the data remains on paper	4 (50%)
10. How are the paper documents digitised?	Scanning	4 (50.0%)
	No digitalisation applied	4 (50%)
11. How is the further documentation of the patient done in the centre?	Electronic	5 (62.5%)
	On paper	0 (0%)
	Both	3 (37.5%)
12. Are patients with unclear diagnoses documented in a subject-independent registry?	Yes	5 (62.5%)
	No	3 (37.5%)

13. Is the centre using software tools or the internet to support the diagnosis?	FindZebra	4 (50.0%)
	Phenomizer	4 (50.0%)
	Search engines	5 (62.5%)
	Literature databases	8 (100%)
	Others	0 (0%)
	No	0 (0%)
14. Who uses the tools mentioned in the previous question?	Physicians	7 (87.5%)
	Medical controlling	1 (12.5%)
	Centre guides	5 (62.5%)
15. How are the diagnoses of rare diseases documented?	ICD-10	8 (100%)
	Orpha-Number	5 (62.5%)
	Human Phenotype Ontology	2 (25.0%)
	Alpha-ID	2 (25.0%)
16. Which software tools are used for documentation?	ORBIS	3 (37.5%)
	SAP i.s.h. med	2 (25.0%)
	Nexus Chili	1 (12.5%)
	Excel	2 (25.0%)

4. Discussion

In this study, we investigated the status quo of RDCs in MIRACUM with a cross-sectional survey. The objectives were to determine how patients are admitted the centres and how the diagnostic process is performed, as well as to describe the current status quo of patient documentation and digitalisation of documentation.

4.1. Patient admission and diagnostic process

Our study noticed that patient admission and diagnosis processes are similar at all surveyed RDCs. The results show that all centres have a central contact point. To contact a centre, seven out of eight centres only allow contact in writing. For further investigation, it might be of interest to know the further role and tasks of the central contact point.

Patients must provide their medical history when they contact the RDC. The results show that important information for the clinicians are medications, blood tests, findings from physical examinations, genetic screenings, and also questionnaires that are filled out by the referring clinician or the patient. Therefore, different clinical data is available which can be used for decision support. Furthermore, it is necessary to determine which clinical data is available in a structured format. The existence of unstructured text, e.g. in doctor's letters, suggests the usage of Natural Language Processing (NLP) to structure the data [11]. Another heterogeneity is the distribution of existing patient cases since a centre with many cases (n=7500) and a centre with few cases (n=25) are available. A further study should investigate how this heterogeneity affects the entire picture of the data. Overall, the availability of patient data in the RDCs is about 10,000 patient cases, which appears low considering approximately 7,000 different RDs. In this context, it would be useful to know how many different diagnoses exist in these 10,000 cases. The ratio of the number of different diagnoses to the number of cases per diagnosis is relevant in this context. For instance, if only 200 different diseases are diagnosed, this would equate to approximately 500 cases per disease.

Furthermore, the results show that five centres process patient cases according to urgency. A CDSS could support an evaluation of the urgency of a patient case based on previous cases. However, the need for such a function needs to be discussed together with the experts for RDs.

The results also show that case discussions are used to investigate patient cases. These are weekly or monthly conferences with clinicians from various disciplines. Our planned CDSS could provide condensed patient summaries and relevant information to support these conferences, for example, with patient timelines or sunburst plots. This approach could be tested in a further study. Furthermore, the results show that many RDCs treat both inpatients and outpatients. However, it is not clear whether these patients all get admitted through the central contact point. It must be determined whether or not this path is only available for outpatients admissions.

4.2. Patient documentation and digitalisation

Our results show that the ICD-10 is predominately used to code diagnoses. ICD-10 is the official classification for coding diagnoses in Germany. However, some RDCs use Orpha-Number. Using ICD-10 poses a problem in terms of the reusability of data because most RDs are not coded in enough detail in the ICD-10 format. Only about 500 out of approximately 7,000 RDs have a specific ICD-10 code, which makes it challenging to identify a RD unambiguously. A further problem is that the ICD code is non-specific, i.e. several RDs cannot be assigned to a single ICD code. In contrast, the Orpha-Number classification contains more than 7,000 RDs [12]. A comprehensive coding with Orpha-Number across all RDCs would be desirable. However, coding of RDs with Orpha-Number will be addressed in the project Collaboration on Rare Diseases (CORD) in the MI-I.

Furthermore, various software systems are used for documentation. Different software systems will probably use different interfaces for communication. It must be determined which standards are used for communication (e.g., HL7-FHIR), in order to be able to transfer data automatically to a CDSS [13].

Another issue is the documentation sent by patients to the centres, which contains essential information that may be relevant to the diagnosis. Some RDCs offer the possibility to send medical history electronically, but some centres only allow paper-based records. Only 50 % of the RDCs regularly digitise documents received on paper. However, to reuse the information for a CDSS, complete electronic data must be provided. Another source of data could be the patient registries at the five centres which store information of patients with RDs. Nevertheless, it is necessary to determine what kind of data is stored.

Despite these challenges, all centres apply software tools for diagnosis support. This statement is consistent with Mueller et al., who found that diagnosis support systems are helpful in the diagnosis of patients with RDs [14]. The results show that centre guides and physicians use these tools. Therefore, it is of further interest in which phase of the diagnostic process the systems are used.

4.3. Limitations and further research

The results are limited to participating centres in the MIRACUM consortium. There are 32 RDCs in Germany. For higher generalisability of the results, it would be interesting to extend the study to other RDCs outside MIRACUM. We are not aware of any studies that involved CDSS implementation in RDCs. Further studies are necessary to gain more detailed insights. To this end, we will use qualitative methods such as expert interviews and focus groups. The created ERM can be used as a basis for discussion and further refined in the next steps.

5. Conclusion

This study aimed to investigate the status quo of RDCs in Germany as part of our requirements analysis phase. We were able to show that the patient admission and diagnostic process are very similar at the different RDCs. This means for the development of a CDSS in RDs, that a CDSS could be integrated into the RDCs diagnostic process in a similar way. However, the heterogeneity of patient cases across the centres as well as the disparity of medical documentation poses problems for the reuse of patient data and therefore for the creation of a common knowledge base of the CDSS. Furthermore, the comprehensive documentation of RDs on paper and coding of diagnosis using ICD-10 constitutes a barrier for the reuse of the data. Despite the challenges, we noticed a corresponding use of various software for diagnosis support in the RDC.

6. Acknowledgments

MIRACUM is funded by the German Federal Ministry of Education and Research (BMBF) within the "Medical Informatics Funding Scheme" (FKZ 01ZZ1801A, 01ZZ1801B, 01ZZ1801C, 01ZZ1801E, 01ZZ1801F, 01ZZ1801G, 01ZZ1801H,

References

[1] Evans W.R., Rafi I, Rare diseases in general practice: recognising the zebras among the horses, *Br J Gen Pract.* **66** (2016), 550–551.

[2] Prokosch HU, Acker T, Bernarding J, Binder H, Boeker M et al., MIRACUM: Medical Informatics in Research and Care in University Medicine - A Large Data Sharing Network to Enhance Translational Research and Medical Care, *Methods Inf Med*. **57** (2018), 82–91.

[3] Schaaf J, Boeker M, Haverkamp C, Hermann T, Kadioglu D, Prokosch HU et al., Finding the Needle in the Hay Stack: An Open Architecture to Support Diagnosis of Undiagnosed Patients, *Stud Health Technol Inform.* (2019), 1580–1581.

[4] Sim I, Gorman P, Greenes R.A., Haynes R.B., Kaplan B, Lehmann H, Tang P.C., Clinical decision support systems for the practice of evidence-based medicine, *Journal of the American Medical Informatics Association : JAMIA*. **8** (2001), 527–534.

[5] 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*. **4** (2015), 29–38.

[6] Kawamoto K, Houlihan C.A., Balas E.A., Lobach D.F., Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success, *BMJ*. **330** (2005), 765.

[7] von Elm E, Altman D.G., Egger M, Pocock S.J., Gøtzsche P.C, Vandenbroucke J.P., The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies, *International Journal of Surgery*. **12** (2014), 1495–1499.

[8] Jenn N.C., Designing A Questionnaire, *Malays Fam Physician*. **1** (2006), 32–35.

[9] Choi B.C.K., Pak A.W.P., A catalog of biases in questionnaires, *Prev Chronic Dis*. **2** (2005), A13.

[10] Song I.-Y., Chen P.P., Entity Relationship Model, in: L. LIU, and M.T. ÖZSU (Eds.), Encyclopedia of Database Systems, Springer US, Boston, MA, 2009: pp. 1003–1009.

[11] Demner-Fushman D, McDonald CJ, What can natural language processing do for clinical decision support?, *J Biomed Inform*. **42** (2009), 760–772.

[12] Bearryman E, Does your Rare Disease Have a Code?, (2015). http://www.eurordis.org/news/does-your-rare-disease-have-code (accessed January 20, 2020).

[13] Smits M, Krawer E, Harthoorn M, Ronald C, A comparison of two Detailed Clinical Model representations: FHIR and CDA, *EJBI*. **11** (2015), 7–17.

[14] Mueller T, Jerrentrupp A, Schaefer J, Computerunterstützte Diagnosefindung bei seltenen Erkrankungen, *Internist*. **59** (2018), 391–400.