Exploring Complexity in Health: An Interdisciplinary Systems Approach
A. Hoerbst et al. (Eds.)

doi:10.3233/978-1-61499-678-1-670

The Use of Tools, Modelling Methods, Data Types, and Endpoints in Systems Medicine: A Survey on Projects of the German e:Med-Programme

Matthias GIETZELT[a,1], Thomas HÖFER[b], Petra KNAUP-GREGORI[a], Rainer KÖNIG[c,d], Martin LÖPPRICH[a], Alexandra POOS[c,d,e] and Matthias GANZINGER[a]

[a] Institute of Medical Biometry and Informatics, Heidelberg University, Heidelberg, Germany
[b] German Cancer Research Center (DKFZ), Heidelberg, Germany
[c] Integrated Research and Treatment Center, Center for Sepsis Control and Care (CSCC), Jena University Hospital, Jena, Germany
[d] Network Modeling, Leibniz Institute for Natural Product Research and Infection Biology - Hans Knöll Institute (HKI), Jena, Germany
[e] Genome Organization & Function, German Cancer Research Center (DKFZ) Bioquant Center, Heidelberg, Germany

Abstract. Systems medicine is the consequent continuation of research efforts on the road to an individualized medicine. Thereby, systems medicine tries to offer a holistic view on the patient by combining different data sources to highlight different perspectives on the patient's health. Our research question was to identify the main data types, modelling methods, analysis tools, and endpoints currently used and studied in systems medicine. Therefore, we conducted a survey on projects with a systems medicine background. Fifty participants completed this survey. The results of the survey were analyzed using histograms and cross tables, and finally compared to results of a former literature review with the same research focus. The data types reported in this survey were widely diversified. As expected, genomic and phenotype data were used most frequently. In contrast, environmental and behavioral data were rarely used in the projects. Overall, the cross tables of the data types in the survey and the literature review showed overlapping results.

Keywords. Systems medicine, survey, tools, data types, endpoints

1. Introduction

Systems medicine is a novel approach for supporting personalized treatment of patients [1]. Thereby, many different data sources such as genotype, phenotype, and lifestyle data for each. patient are taken into account [2]. These data sources are of

[1] Corresponding Author: Matthias.Gietzelt@med.uni-heidelberg.de

heterogeneous data types and they are used to highlight different perspectives on the patient's health. The gathered data can be fused and interpreted by e.g. decision support systems [3, 4]. The challenging problem for Medical Informatics is to provide therapy suggestions based on these data and evidence-based knowledge.

The e:Med initiative was established by the German Federal Ministry of Education and Research (BMBF) to encourage and support national developments in systems medicine. The e:Med initiative consists of 31 research projects [5].

The aim of this paper is to present and discuss the results of a prospective survey, in which the members of the e:Med initiative were invited to participate in an online survey about the topic and content of their work.

Our research question was to identify the main data types, modelling methods, analysis tools and endpoints currently used in systems medicine in Germany within the projects of the e:Med initiative and put these results in a broader perspective to compare them with the current field of research worldwide facing systems medicine.

2. Methods

Members of the e:Med initiative were invited to take part in a survey about the use of tools, modelling methods, data types, and endpoints of their dedicated projects. The study was conducted during Nov-Dec, 2015 using the online survey tool LimeSurvey version 2.06 [6]. Questions to the participants were about:

- the individual educational background
- the primary role in the project
- diseases covered in the project (as ICD codes)
- data extraction and transformation tools
- data integration tools (e.g. data warehouses)
- data management tools
- data analysis and evaluation tools
- data types
- data sources
- categories of modelling methods
- endpoints studied
- use of biomaterials.

Only a few of these questions will be addressed in this paper. In case of the modelling method categories, the participants were asked to choose from main categories (regression, classification, clustering, time series, network analysis, differential equations, and other) and to specify the methods more precisely afterwards. The results of the survey were analyzed using histograms and cross tables.

3. Results

Fifty members of the e:Med initiative completed this survey, whereas the main amount of responses was given by principal investigators (64 %).

The analysis of the ICD codes of the diagnoses shows that there is a strong emphasis on neoplasms (C00-D48), mental and behavioral disorders (F01-F99), diseases of the nervous system (G00-G99), and the circulatory system (I00-I99). Figure 1 shows the findings of the survey and the comparison to the literature [7].

ICD-10 classes	A00-B99	C00-D49	D50-D89	E00-E89	F01-F99	G00-G99	H00-H59	H60-H95	I00-I99	J00-J99	K00-K95	L00-L99	M00-M99	N00-N99	O00-O9A	P00-P96	Q00-Q99	R00-R99	S00-T88	V00-Y99	Z00-Z99
literature review [7]	5	11	1	9	4	6	3	0	11	5	3	3	2	2	1	0	1	5	1	1	1
e:Med survey	0	28	0	1	9	4	0	0	7	1	4	2	1	0	0	0	0	0	2	0	0

Figure 1. The distribution of ICD-10-CM codes of the survey and the related literature review [7].

In the survey, we also asked for the data types used in the projects. Transcriptomic and genomic data were used most often by the participants, followed by phenotype and laboratory data (see figure 2).

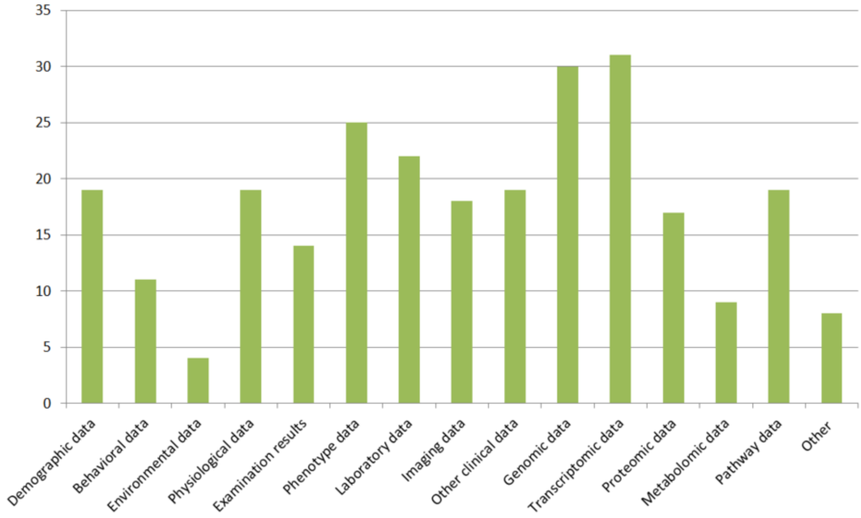
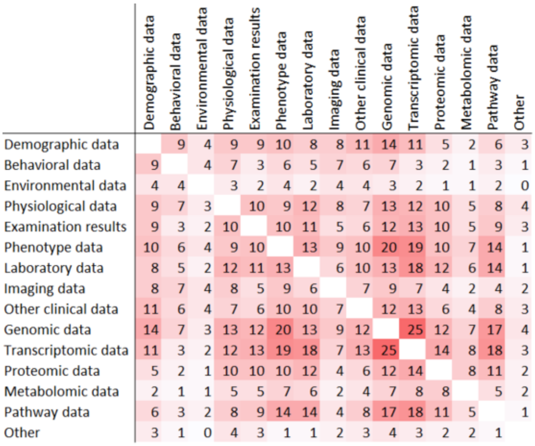


Figure 2. Data types used in the e:Med-projects.

We further investigated data type combinations, because we expected data type compositions, which are characteristic for systems medicine. Figure 3 shows the resulting cross table. It shows, for example, that transcriptomic data was most frequently used in combination with clinical, phenotype, laboratory and pathway data. Another interesting finding is that physiological data are often used in combination with examination results, laboratory data, and genomic data.

We also analyzed the modelling methods which were used within the e:Med initiative. Network analysis was used most often (66 % of the participants), followed by regression and classification (58 %) as well as clustering (54 %). Most participants (37 out of 50) used different kinds of modelling methods, where regression, classification and network analysis are used most often together.

Furthermore, we were interested which software was used for the modelling approaches. Thereby, we found that R in general (with Bioconductor and other packages) was used most frequently for modelling, especially for regression, classification, clustering and network analysis. Besides this, there was also a high amount (40 %) of own developed tools.



	Demographic data	Behavioral data	Environmental data	Physiological data	Examination results	Phenotype data	Laboratory data	Imaging data	Other clinical data	Genomic data	Transcriptomic data	Proteomic data	Metabolomic data	Pathway data	Other
Demographic data		9	4	9	9	10	8	8	11	14	11	5	2	6	3
Behavioral data	9		4	7	3	6	5	7	6	7	3	2	1	3	1
Environmental data	4	4		3	2	4	2	4	4	3	2	1	1	2	0
Physiological data	9	7	3		10	9	12	8	7	13	12	10	5	8	4
Examination results	9	3	2	10		10	11	5	6	12	13	10	5	9	3
Phenotype data	10	6	4	9	10		13	9	10	20	19	10	7	14	1
Laboratory data	8	5	2	12	11	13		6	10	13	18	12	6	14	1
Imaging data	8	7	4	8	5	9	6		7	9	7	4	2	4	2
Other clinical data	11	6	4	7	6	10	10	7		12	13	6	4	8	3
Genomic data	14	7	3	13	12	20	13	9	12		25	12	7	17	4
Transcriptomic data	11	3	2	12	13	19	18	7	13	25		14	8	18	3
Proteomic data	5	2	1	10	10	10	12	4	6	12	14		8	11	2
Metabolomic data	2	1	1	5	5	7	6	2	4	7	8	8		5	2
Pathway data	6	3	2	8	9	14	14	4	8	17	18	11	5		1
Other	3	1	0	4	3	1	1	2	3	4	3	2	2	1	

Figure 3. The cross table of data types used in the e:Med-projects.

Figure 4 shows the distribution of endpoints studied in the particular projects. Most frequently, disease subtypes and biomarkers were studied. Other endpoints were driver mutations, heart specific parameters, and relapse prediction for harmful alcohol use.

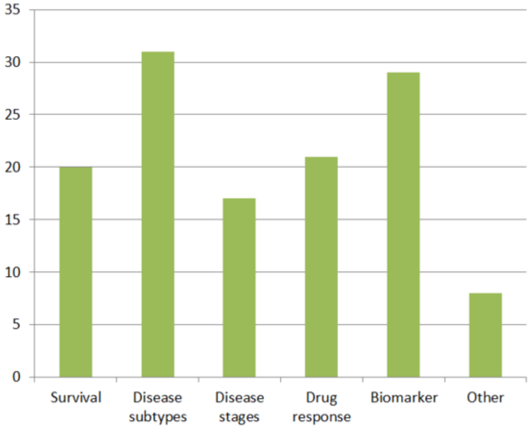


Figure 4. Endpoints studied in the e:Med-projects.

4. Discussion

The projects of the e:Med initiative may focus on particular diseases and do not reflect the whole spectrum of projects in systems medicine worldwide (as shown in figure 1). However, the comparison showed that there are a lot of similarities in data types, modelling methods, analysis tools, and endpoints, regardless of the specific disease addressed.

The data types reported in this survey were widely diversified. As expected, genomic and phenotype data were used most frequently. In contrast, environmental and behavioral data were rarely used in the projects. This was also found as a result in a systematic literature review [7]. Surprisingly, physiological data types were commonly used in the survey (29 times), but infrequent in the literature (only once). In comparison with the literature review, the distribution of data types is more diversified in the survey. This might be an effect of the analysis of the publications, in which this particular information is missing or described too briefly. Overall the cross tables of the data types in both studies showed strong overlapping results. Microsoft Excel was mentioned quite often as a data analysis tool in the survey (46 %), but was only once found in the literature review. This result might be biased, because Excel might not be regarded worth being mentioned in a scientific publication. The main research focus in both, the survey and the literature review was on ICD codes with the highest Disability Adjusted Life Years (DALY) rankings [8, 9]. This could imply that the impact of systems medicine is expected to be a significant progress on the road to an individualized medicine. There are some limitations of our survey. First, only members of the e:Med initiative were invited to take part in this survey. This may have biased the results, because only research projects, consortia, and alliances, which were granted by the German BMBF e:Med initiative participated. Second, it cannot be excluded that a participant completed the survey twice, because the forwarded link was not personalized. However, this scenario can be considered unlikely.

Acknowledgement

This survey was done within the projects CLIOMMICS and CancerTelSys. The projects are funded by the German Federal Ministry of Education and Research (BMBF, grant id: 01ZX1309A and 01ZX1302B) as part of the e:Med initiative.

References

[1] R.S. Wang, B.A. Maron, J. Loscalzo, Systems medicine: evolution of systems biology from bench to bedside, Wiley Interdiscip Rev *Syst Biol Med.* **7** (2015): 141-61.

[2] M. Benson, Clinical implications of omics and systems medicine: focus on predictive and individualized treatment, *J Intern Med.* **279** (2016): 229-40.

[3] L.K. Wiley, P. Tarczy-Hornoch, J.C. Denny, R.R. Freimuth, C.L Overby, N. Shah, R.D. Martin, I.N. Sarkar, Harnessing next-generation informatics for personalizing medicine: a report from AMIA's 2014 Health Policy Invitational Meeting, *J Am Med Inform Assoc.* **23** (2016): 413-9.

[4] M. Rasoolimoghadam, R. Safdari, M. Ghazisaeidi, M. Maharanitehrani, S. Tahmasebiyan, Designing Decision Support System to Detect Drug Interactions Type 2 Diabetes, *Acta Inform Med.* **23** (2015): 336-8.

[5] e:Med Systems Medicine [homepage on the internet]. Federal Ministry of Education and Research (BMBF), Germany; 2015. [updated 2015; cited 2016 Mar 1]. Available from: http://www.sysmed.de/en/.

[6] LimeSurvey [homepage on the internet]. LimeSurvey GmbH, Germany; 2016. [updated 2015 Oct 18; cited 2016 Fev 9], Available from: https://www.limesurvey.org/.

[7] M. Gietzelt, M. Löpprich, C. Karmen, P. Knaup, M. Ganzinger, Models and Data Sources Used in Systems Medicine: A Systematic Literature Review, *Methods Inf Med.* **55** (2016), [Epub].

[8] F. Sassi, Calculating QALYs, comparing QALY and DALY calculations, *Health Policy Plan.* **21** (2006): 402-8.

[9] E.E. Groessl, R.M. Kaplan, C.M. Castro Sweet, T. Church, M.A. Espeland, T.M. Gill, N.W. Glynn, A.C. King, S. Kritchevsky, T. Manini, M.M. McDermott, K.F. Reid, J. Rushing, M. Pahor, LIFE Study Group, Cost-effectiveness of the LIFE Physical Activity Intervention for Older Adults at Increased Risk for Mobility Disability, *J Gerontol A Biol Sci Med Sci.* **pii: glw001** (2016), [Epub].