Evaluation of Data Quality in the Estonian National Health Information System for Digital Decision Support

Markus Bertl[1,*,†], Kristian Juha Ismo Kankainen[1,†], Gunnar Piho[2], Dirk Draheim[2] and Peeter Ross[1,3]

[1]Department of Health Technologies, Tallinn University of Technology, Ehitajate tee 5, Tallinn, 12616, Estonia
[2]Department of Software Science, Tallinn University of Technology, Ehitajate tee 5, Tallinn, 12616, Estonia
[3]East Tallinn Central Hospital, Ravi 18, Tallinn, 10138, Estonia

Abstract

Following the implementation of Electronic Medical Records (EMR), the amount of digital health data has increased significantly in recent decades. This trend creates an opportunity to share data between different healthcare parties for primary and secondary use. However, the quality of this data is often questioned, and data reuse is still rare. This study evaluates the frequency of the use and quality of health data stored in the Estonian Health Information System (EHIS), which is one of the most advanced digital health platforms (DHP) in the world. We collected usage data of the EHIS from its initial release in 2008 till 2021. Comparing 2016 to 2021, the number of documents per year pushed into the EHIS has nearly doubled. But also approximately nine times more patients and five times more health professionals queried data from the EHIS. This increase in read access indicates that both groups find valuable information from the system. To investigate this further, data from patients with common diseases like stroke, cancer, or diabetes have been queried, analyzed, and compared against the actual data needs from the point of healthcare professionals and natural persons. Contradictory to the claim mentioned above, the manual analysis of the queried data sometimes showed poor data quality and missing information, especially discrepancies between the structured and unstructured parts of the documents shared through DHP. As an example of varying data quality, we looked at how smoking behavior is reported, both in structured form and in free text form in the queried data. We analyzed how the data quality of smoking behavior data shifts from document to document using the nine data quality dimensions of the Data Quality Vector. The data quality is shown to shift in 7 dimensions. While humans seem to be able to screen the data and resolve inconsistencies effectively, the data quality issues present make data reuse for tasks like AI training for digital decision support systems challenging.

Keywords

Data Quality, EHR (Electronic Health Record), Estonian National Health Information System, Digital Decision Support (DDSS), Artificial Intelligence (AI), Machine Learning (ML), Medical Data Reuse, Primary Use, Secondary Use

HEDA 2023: the 3rd International Workshop on Health Data (https://conf.researchr.org/home/staf-2023/heda-2023). Co-located with STAF 2023, 18–21 July, Leicester, United Kingdom.

*Corresponding author.

†These authors contributed equally.

markus.bertl@taltech.ee (M. Bertl); kristian.kankainen@taltech.ee (K. J. I. Kankainen); gunnar.piho@taltech.ee (G. Piho); dirk.draheim@taltech.ee (D. Draheim); peeter.ross@taltech.ee (P. Ross)

0000-0003-0644-8095 (M. Bertl); 0000-0002-0551-927X (K. J. I. Kankainen); 0000-0003-4488-3389 (G. Piho); 0000-0003-3376-7489 (D. Draheim); 0000-0003-1072-7249 (P. Ross)

1. Introduction

Estonia is a country in the north of Europe with 1.3 million citizens, approximately 4,500 physicians, and healthcare costs, which made up for about 7.5% of the annual GDP in 2021 [1]. The Estonian health system is based on mandatory, solidarity-based insurance and healthcare providers which operate under private law [2]. The Estonian digital health platform (DHP), called the Estonian nation-wide health information system (EHIS), has been operational since 2008 and allows secure and trusted online access to medical data, different kinds of medical documents, prescriptions, and medical images of virtually every Estonian resident from birth to death. It is fully integrated into the Estonian e-government systems, which provides a digital identity to every citizen, secure authentication methods, the possibility to link data according to the once-only principle, and other mature e-services [3, 4]. Instead of one large, centralized database, the EHIS comprises different federated and mutually independent systems. One is the nationwide electronic health record (EHR) system, which began the ongoing standardization of health-related data in Estonia [5]. In the central EHR, patient data is saved based on international standards like HL7 CDA[1], DICOM[2], LOINC[3], ICD-10[4] and SNOMED-CT[5]. The EHIS uses HL7 CDA as its data collection format. The CDA structure not only permits data capture in structured form but also allows to add medical data in unstructured free text format. Data sent to the EHIS is digitally signed or stamped by either the physician or the healthcare institution, which ensures accountability of the provided information. The data can be queried either directly from the data warehouse for statistical purposes and research, via API for eHealth applications like Hospital Information Systems (HIS), or via Web UIs like the patient portal over which residents of Estonia can view medical data from healthcare providers, referral letters, prescriptions, or fill out health declarations before an appointment. An overview of the first ten years of the EHIS can be found in [6].

As of today, the data collection process works as follows – The primary data sources for the EHIS are the electronic medical records (EMRs) of healthcare providers. Data is entered into the EMR by doctors and nurses or automatically transmitted from digital data sources such as laboratory equipment, etc. The data are entered in different modes: as free text, numeric data, including different codes (ICD-10, etc.), as graphs (ECG, etc.), or images (radiology, endoscopy, etc.). In order to share data with other institutions, the EMR exports data and digital documents in accordance with established standards (HL7 CDA, LOINC, etc.) for nationwide use and pushes them to the applications of various data consumers. One data consumer is the EHIS. Another data consumer is the Estonian Health Insurance Fund (EHIF), to which the ICD-10-coded diagnoses from the EMR are transmitted for billing purposes.

Digital decision support describes computer-based systems that bring together information from various sources, assist in the organization and analysis of information, and facilitate the evaluation of assumptions underlying the use of specific models [7]. Digital Decision Support Systems (DDSSs) could be divided by their goal of using them either as data capture aids or data

[1] http://www.hl7.org/
[2] https://www.dicomstandard.org/
[3] https://loinc.org/
[4] https://icd.who.int/browse10/
[5] https://www.snomed.org/

analysis and presentation tools. They can, for instance, be based on summarizing or visualizing data like the patient summary (Andmevaatur - data viewer in Estonian) functionality of the EHIS, or based on AI-based decision technology like rule-based expert systems [8], machine learning [9], or deep learning [10]. Regardless of the implementation flavor, data is needed for them to work accurately. One would expect that a sufficiently large amount of data to train and operate DDSSs is available through DHPs. Nevertheless, adoption rates of DDSSs are rather low [11]. AI algorithms for DDSS in healthcare itself, however, seem to perform sufficiently accurately [12, 13]. Besides having a holistic approach that includes domain experts from both the medical and the IT side, insufficient data quality has been found as one of the main barriers [11, 14]. Until now, there are two DDSSs operational in Estonia: the drug-drug interaction alert service (Inxbase[6]) and clinical decision support for primary healthcare physicians (EBMeDS[7]). Inxbase is part of the e-prescription services and uses manually defined rules to alert physicians if they prescribe medication that could interact with pharmaceuticals prescribed by other physicians [15]. There is currently no AI-based DDSS trained on data from the EHIS. Therefore, this research investigates the data quality of the EHIS in Estonia and assesses if the data stored there would even be usable for DDSSs.

2. Method

The EHIS has been chosen as the study object of this research because it is one of the most advanced nationwide DHPs in the world [16]. Therefore we assume it to be representative of the state-of-the-art in terms of data capture and data quality. We analyzed two parameters of the EHIS in this research:

- **Use** of saved health data measured by counting all queries made to the EHR from healthcare professionals through their EMRs and patients through the online accessible patient portal[8] of the EHIS. Queries can be, for instance, access to lab results, patient documentation, prescribed medication, or vaccination certificates.
- **Quality** of the captured data, especially to analyze the difference between structured and unstructured data, was measured by a Data Quality Vector (DQV). For this analysis, we decided to apply the DQV to the data of patients whose smoking status has been captured. Smoking is a highly relevant health factor and, in the EHIS case, can be documented both in free text in the EMR or in structured form in the health declaration of the EHR. We analyzed the entries of five randomly selected patients (12 documents in total) in this research to obtain preliminary results about the data quality.

The primary use of data is defined as data used directly for patient care and/or healthcare activities (including self-care). In contrast, secondary use (also called data reuse, multiple use, and further use) is defined as all data use that is not directly linked to patient care [17].

The Data Quality Vector (DQV) [18] offers a multi-dimensional view of data quality. Its nine data quality dimensions (Table 1) unify, according to its authors, all data quality dimensions

[6] https://www.medbase.fi/en/professionals/inxbase
[7] https://www.ebmeds.org/en/
[8] https://www.digilugu.ee/login?locale=en

Table 1
The nine dimensions of data quality according to the Data Quality Vector [18]

Dimension	Description
Completeness	The degree to which relevant data is recorded
Consistency	The degree to which data satisfies specified constraints and rules
Duplicity	The degree to which data contains duplicate registries representing the same entity
Correctness	The degree of accuracy and precision where data is represented with respect to its real-world state
Timeliness	The degree of temporal stability of the data
Spatial stability	The degree to which data is stable among different populations
Contextualization	The degree to which data is correctly/optimally annotated with the context in with it was acquired
Predictive value	The degree to which data contains proper information for specific decision-making purposes
Reliability	The degree of reputation of the stakeholders and institutions involved in the acquisition of data

proposed by other researchers previous to 2012. We used the DQV to assess in which dimensions data quality shifts occur between documents over time. Shifts were assessed between unstructured text and structured data, as well as between sequential documents.

The analyzed data concerns the smoking behavior of the subject of care, either in a structured form as part of health declarations or as free text excerpts as part of clinical reports (discharge summaries, referrals, etc.). The data is grouped by individual and includes all clinical documents about the person that was reported to the EHIS during the year 2019. The detailed inclusion criteria were: age 30–70 years, diagnosis of chronic disease (ICD-10 codes I00–I99, C00–C97, E10–E14). The initial sample size was 90 randomly selected individuals but evenly distributed across the diagnosis groups. The sample size was further decreased to 59 patients by filtering out only those with available data on smoking behavior. Data on smoking behavior was discovered by text search and annotated semantically by hand, otherwise as structured data in the health declaration form. The health declaration is a patient-reported questionnaire and is the basis for health certificates. Of the 59 patients with data on smoking behavior, only five had health declarations, whereas 57 had smoking behavior mentioned in free text. Three health declarations out of the five overlapped with information from free text. Of the two health declarations that provided smoking behavior without it being also mentioned in free text, one expressed smoking, and one expressed non-smoking status. We set up the DQV framework as follows. The analyzed documents were characterized as time-stamped and reported by different healthcare providers. Our analysis considers this time dependency, and although our data has been gathered in retrospect, we analyze it as if it was collected in a continuous data flow. To emulate decision support from the point of view of the document writer, we impose an imaginary constraint on whether the data has been available during the writing. This was judged by the look of the text, e.g., whether it is a copy-paste. The DQV was then used to analyze the documents in the following way. For each patient with data available on smoking behavior, the documents were ordered according to time. Thereafter the smoking behavior of each document was assessed and compared with the information mentioned in the succeeding document using each of the nine data quality dimensions.

The Estonian Human Research Ethics Committee (TAIEK) of the Institute for Health Development (Decision No. 1.1-12/186) approved the research design and data usage for this study.

3. Results

3.1. Use of Health Data

Figure 1 presents the number of documents added to the EHIS per year, the number of documents accessed by patients over the patient portal, and the number of queries from healthcare professionals from the initial launch of the EHIS in 2008 until 2021. While the number of documents pushed to the EHIS seems to reach a peak, the queries from both patients and doctors are still increasing sharply. The red line, which represents the number of queries from health professionals, only accounts for queries that have been actively and knowingly performed to get information from the EHR. The number does not contain system requests which are automatically performed during the clinical process. The blue line represents the number of queries performed through the patient portal of the EHIS, so it shows how much EHR data the patients view. It is worth mentioning that the number of queries is not equal to the number of natural persons logged into the patient portal, as several queries are usually made during a single patient portal session. Also, it is important to consider the COVID-19 pandemic when interpreting the numbers in Fig. 1. Vaccination certificates and lab test results are also part of the EHIS. Accessing them also contributed to the rise in patient queries in 2020 and 2021.

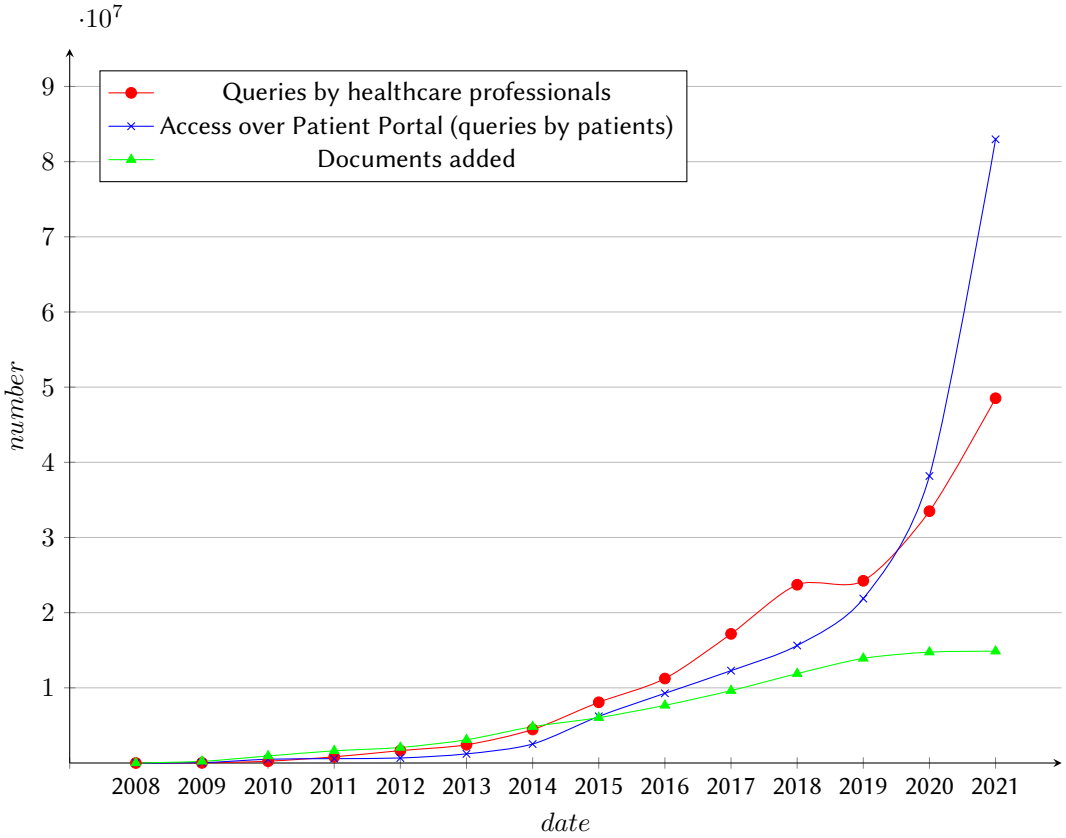

Figure 1: Data access of Estonian National Health Information System 2008-2021

3.2. Shifts in data quality occurring between documents

In the following subsections, we report on five illustrative findings of how data quality was assessed to shift between the analyzed documents according to the nine dimensions of the Data Quality Vector introduced above.

3.2.1. Shift between discharge summary and structured health declaration – the case of not smoking

In this example, we analyze two documents. The first document is the unstructured text of an anamnesis section of a discharge summary. The text states in one sentence both facts of non-smoking and alcohol-drinking behavior. The second document is a structured health declaration form that was filled in four months later. The structure of the health declaration form is such that data on smoking behavior and drinking behavior are in separate fields. Shifts in the following data quality dimensions can be observed (also Table 2):

Completeness does not shift, as the smoking status "not smoking" is semantically fully intact on both documents.
Consistency shifts in a technical sense as the first document is not machine-readable while the second is.
Duplicity of data does not occur, as smoking status changes over time.
Correctness of the data does not shift.
Timeliness does not apply, as smoking status changes over time.
Spatial stability shifts similar to the consistency dimension (text VS structured).
Contextualization of data does shift because the tight connection to alcohol-drinking behavior is not preserved in the health declaration.
Predictive value of the data does not shift. This dimension does not apply as nothing predicts a change in smoking behavior.
Reliability of the data can be said to shift, as health declarations are often filled by the patient.

Table 2
Shifts in data quality dimensions between discharge summary and structured health declaration – the case of not smoking (example 1, section 3.2.1).

Dimension	Shift
Completeness	No
Consistency	Yes
Duplicity	No
Correctness	No
Timeliness	-
Spatial stability	Yes
Contextualization	Yes
Predictive value	-
Reliability	Yes

3.2.2. Shift between discharge summary and structured health declaration – the case of smoking

In this example, we again have an anamnesis and a health declaration form filled in four months later. The first document states in anamnesis vitae the smoking longevity (“long-term

smoker"), temporal length ("30 years"), and the smoking amount ("one pack per day"). The second document is a health declaration form filled in four months later stating the patient is a smoker, the length in years ("30"), and the smoking amount in cigarettes per day ("2"). Shifts in the following data quality dimensions can be observed (also Table 3):

Completeness was unaffected, as all three semantic attributes (smoking status, length, and amount) remained intact. It could be argued that some interpretation of longevity (explicitly marked "long-term") is lost, although the length in years stays the same.

Consistency shifts as different units are used for the smoking amount (packs vs. cigarettes). Another shift could be argued with the loss of qualifier ("long-term").

Duplicity does not apply, as smoking status changes over time.

Correctness of the data can be analyzed both ways. If the smoking behavior has not changed, the number of cigarettes is a typo, as one pack (in Estonia) equals 20 cigarettes. In case of a change in smoking behavior, a decrease has occurred from 20 to 2 cigarettes per day.

Timeliness does not apply.

Spatial stability shifts similarly to consistency (text VS structured).

Contextualization did not change as both contexts can be interpreted as general knowledge about the patient's lifestyle.

Predictive value of the data does not shift. This dimension does not apply as nothing predicts a change in smoking behavior.

Reliability of the data can be said to shift, as health declarations are often filled by the patient.

Table 3
Shifts in data quality dimensions between discharge summary and structured health declaration – the case of smoking (example 2, section 3.2.2).

Dimension	Shift
Completeness	No
Consistency	Yes
Duplicity	No
Correctness	No/Yes
Timeliness	-
Spatial stability	Yes
Contextualization	No
Predictive value	-
Reliability	Yes

3.2.3. Shift between two inpatient discharge summaries

The first document states in the treatment synopsis of an inpatient rheumatology discharge summary a recommendation to stop smoking, among other recommendations. The second document is an inpatient cardiology discharge summary and states twice in the anamnesis the fact of being a smoker (first in the problem list and then in a separate smoking status field). Shifts in the following data quality dimensions can be observed (also Table 4):

Completeness was affected as the first document contained only the cessation recommendation, and the second document stated only the fact of being a smoker.

Consistency was breached as our hierarchy rules state smoking status should come before cessation recommendation.

Duplicity was found inside the second document without data reuse being evident; rather, the information was presented in two contextualizations.

Correctness of the data can be both ways: it can be seen as dependent on reasoning capabilities: if cessation recommendation implies being a smoker, then all is correct. Another view would be to allow non-linearity in that cessation recommendations can correctly be given to anyone, also non-smokers.

Timeliness was not evident in the data.

Spatial stability was analyzed similarly to correctness – it relies on reasoning capabilities, e.g., the rules specified by the consistency dimension.

Contextualization was different for each occurrence: implicitly in cessation recommendation (treatment), explicitly in the problem list, and separately as smoking status.

Predictive value of the data does not shift. This dimension does not apply as nothing predicts a change in smoking behavior.

Reliability of the data does not shift.

Table 4
Shifts in data quality dimensions Shift between two inpatient discharge summaries (in example 3, section 3.2.3).

Dimension	Shift
Completeness	Yes
Consistency	Yes
Duplicity	Yes
Correctness	-
Timeliness	-
Spatial stability	Yes
Contextualization	Yes
Predictive value	-
Reliability	-

3.2.4. Richness of the contextualization dimension

This example consists of only one document; therefore, no shift can be analyzed. Instead, our intention here is to highlight the value and richness of contextuality from the healthcare professional's point of view. We found in one document that the smoking behavior was stated in a comment next to the structured data fields with elevated blood pressure and pulse. The textual comment had a nomenclature code for a cardiovascular observable and an interpretation code for normal. The free text of the comment stated the patient not being a smoker.

3.2.5. Shifts occurring between multiple documents

In this example, we could trace smoking behavior across five different documents.

- The first document is the anamnesis section of a referral that states the longevity of smoking ("long-term smoker").
- The second document is the anamnesis section of an outpatient visit. It duplicates exactly the text from the referral and adds no more information. This we analyze as a reuse of timely available data.

- The third document is the anamnesis section of an inpatient discharge summary stating smoking status, adding an approximate numerical quantification of longevity ("more than 20 years") and the amount in packs per day as a span ("1–1.5").
- The fourth document is the anamnesis section of a pulmonology outpatient discharge summary. It duplicates the exact phrase from the previous document but adds to it the current trend of smoking amount ("has tried to cut down lately").
- The fifth document is a general practitioner outpatient discharge summary treatment regimen. It does not mention smoking status but states the importance that the patient stops smoking, e.g., is an instruction on smoking cessation.

Refer to Table 5 for our analysis of the shifts from document to document according to the DQV dimensions.

Table 5
Shifts in data quality dimensions between multiple documents (example 5, section 3.2.5).

Dimension	Shift I–II	Shift II–III	Shift III–IV	Shift IV–V	Shift V-VI
Completeness		No shift	Adds data	Adds data	
Consistency	Is consistent	No shift	Shift to more precise granularity	No shift	
Duplicity		Duplicates		Duplicates	
Correctness					
Timeliness		Was timely		Was timely	
Spatial stability					
Contextualization	Anamnesis in referral	Anamnesis in referral	Anamnesis in inpatient discharge summary	Anamnesis in outpatient discharge summary	Needed self-care activity, Treatment regime in GP discharge summary
Predictive value	-	-	-	-	-
Reliability	-	-	-	-	-

4. Discussion

Our data quality vector analyses show clearly that data quality shifts in several dimensions between documents. We observed shifts in the following dimensions: Completeness, Consistency, Duplicity, Correctness, Spatial stability, Contextualization, and Predictive value. It is evident that the granularity of information changes throughout the clinical process. If, for example, only the smoking status is needed for a decision, it is found in more documents than the more precise knowledge of how many cigarettes the patient smokes daily. Additionally, having information in both structured and unstructured forms creates redundancies, leading to inconsistency. This not only introduces challenges to data usage for decision support but also makes secondary use difficult because the inconsistently structured or even unstructured data is hard to aggregate (e.g., querying the average number of smoked cigarettes per age group).

The number of new documents added per year to the EHR system seems to be reaching its current maximum, with no sharp increases observed since 2019. In contrast, a sharp increase in

the number of queries can be observed for the same time span. This usage pattern indicates to us that both patients and healthcare professionals have been getting useful information from the system in recent years. Otherwise, the users would not query it increasingly. If the system did not bring a benefit, people would use it less and query rates should stagnate or even decrease. Such a trend is visible from 2008 to approximately 2014 in Figure 1. The EHIS was just launched back then and did not contain enough useful information for patients and healthcare professionals to yield high query counts. The trend of rising query rates, combined with our findings of inconsistent and potentially low-quality data, raises the question of why medical professionals still use the EHR system increasingly. It is a clinical routine that healthcare professionals have to use as many data sources as reasonably possible about the patient's health status to make medically relevant decisions. So far, medical professionals' education has emphasized the importance of reading previous patient files and test results. This means that in the Estonian case, healthcare professionals are approaching the patient data mainly in a conventional manner, not benefiting from the full spectrum of digital data-sharing opportunities. However, the latter (e.g., the use of DDSS in the clinical process) is possible only if the collected data is standardized and structured, making it available for computer processing. Also, it could be argued that humans can make better sense of the available low-quality textual data by intuitively determining which interpretation of the inconsistent data is most likely correct – a problem that is still hard for AI algorithms. Recent advantages in deep learning, especially deep natural language processing (NLP), do allow the use of unstructured data. However, whether those methods give the needed accuracy is still questionable. Using NLP methods to structure unstructured health data by extraction can also introduce additional inconsistencies and shifts in other data quality dimensions. As one example, the indicated practice of exchanging (unstructured) data within referral documents (see 3.2.5), where each receiving healthcare professional adds more detail and sends the elaborated data with a new referral. This practice leads to a data integration situation that is very hard to coordinate: the previously known data is duplicated, and the new data elements are rooted within their own contexts creating instability in both the spatial and timeliness dimensions. It is not the structuring of data that is hard, but instead, the coordination and interpretation. The shift in spatial stability leads to the question of which source should be accounted for, and the shift in the timeliness dimension leads to the question of when to account for what data. These shifts, in turn, affect the completeness and consistency dimensions. Therefore, using NLP technologies to structure textual health data would introduce additional risks for a DDSS since it potentially would work on tainted data.

One of the few pieces of information available in a quality-controlled format is demographic information linked from other e-government registered and the mandatory ICD-10-coded diagnosis, which needs to be recorded for billing purposes at each physician visit. The challenge for DDSSs is that much of the more granular information in the EHIS is still stored in a free text format instead of a machine-readable, structured form. Humans can interpret these textual descriptions, but they are not machine-understandable. This makes data reuse challenging. Structured data would be desirable to train AI algorithms for decision support, effectively query data, or perform statistical analyses. If we assume some mechanism that would make the free text of clinical documents machine-understandable, then in the case of smoking behavior, our results show the need not only to reuse but also to manipulate the data by later refining the semantics (pt is smoker > pt smokes for X years > pt smokes X cig/day > pt has cut down the

amount of cig/day > pt stopped smoking > pt has not smoked for X months). Our analysis supports the hypothesis that one document is not always enough for a granular understanding of smoking status, but rather a cumulative view should exist. But data aggregation presupposes structured data instead of free text. To maximize the usefulness of decision support, not only one axis, like smoking status (yes/no), needs to be queryable through structured data, but also at least a second, time-based dimension containing more granular data like the number of cigarettes/day. This would introduce more information for AI-based algorithms to train on and potentially allow more accurate predictions.

Generalizing this understanding to other health data, the rise in patient's document retrievals might also be due to difficulty finding the right information, resulting in multiple searches for the document containing the needed information. For example, more documents contain information answering yes/no questions, whereas few documents contain more granular information.

We want to highlight that this research only presents a preliminary analysis that probably does not cover all data quality issues in the current DHP. Our main goal was to show that there are severe data quality issues even in those small, random samples. Based on the methodology described, a more detailed data quality analysis of a larger cohort of patients will follow.

5. Conclusion

Our analysis shows that the use of nationwide electronic health records embedded in a digital health platform is well accepted and widely used by healthcare professionals and patients, despite the sometimes questionable quality of the data. The number of queries to the EHIS is rising, which shows increased use and indicates that people are finding helpful information. We discovered shifts in seven of nine data quality dimensions by analyzing individual documents in detail. The shifts express, among other, information being added upon and made more precise, and inconsistencies between the structured and unstructured (free text) parts of an entry to the EHIS. Humans can make sense of the shifting data quality and unstructured data by using abductive reasoning (intuitively using their knowledge to find the most likely interpretation of the available information). This is challenging for machines, making the data difficult for tasks like AI training, effectively querying data, or performing statistical analyses. For this, high-quality, structured data would be needed. Although specific mandatory structured data fields in the EHR, like ICD-10 coded diagnosis, can be utilized for DDSSs, structured data on more complex information is often still not available.

Acknowledgments

The Estonian Human Research Ethics Committee (TAIEK) of the Institute for Health Development (Decision No. 1.1-12/186) approved the research design and data usage for this study.

This work in the project 'ICT programme' was supported by the European Union through the European Social Fund and the Norway Grants Program "Green ICT" (Nmb. F21009).

References

[1] T. Habicht, K. Kahur, K. Kasekamp, K. Köhler, M. Reinap, A. Võrk, R. Sikkut, L. Aaben, E. Van Ginneken, E. Webb, et al., Estonia: health system summary, 2022 (2023).

[2] T. Lai, T. Habicht, M. Jesse, Monitoring and evaluating progress towards universal health coverage in estonia, PLoS Medicine 11 (2014) e1001677.

[3] R. Krimmer, T. Kalvet, M. Toots, A. Cepilovs, E. Tambouris, Exploring and demonstrating the once-only principle: a european perspective, in: Proceedings of the 18th annual international conference on digital government research, 2017, pp. 546–551.

[4] S. Lips, V. Tsap, N. Bharosa, R. Krimmer, T. Tammet, D. Draheim, Management of national eID infrastructure as a state-critical asset and public-private partnership: Learning from the case of Estonia, Information System Frontiers (2023). doi:10.1007/s10796-022-10363-5.

[5] M. Tiik, P. Ross, Patient opportunities in the estonian electronic health record system, in: Medical and Care Compunetics 6, IOS Press, 2010, pp. 171–177.

[6] J. Metsallik, P. Ross, D. Draheim, G. Piho, Ten years of the e-health system in estonia, in: A. Rutle, Y. Lamo, W. MacCaull, L. Iovino (Eds.), CEUR Workshop Proceedings, volume 2336, 3rd International Workshop on (Meta)Modelling for Healthcare Systems (MMHS), 2018, pp. 6–15. URL: ceur-ws.org/Vol-2336/MMHS2018_invited.pdf.

[7] V. Sauter, Decision support systems: an applied managerial approach, John Wiley & Sons, Inc., 1997.

[8] M. Bertl, M. Shahin, P. Ross, D. Draheim, Finding Indicator Diseases of Psychiatric Disorders in BigData Using Clustered Association Rule Mining, in: Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, SAC '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 826–833. URL: https://doi.org/10.1145/3555776.3577594. doi:10.1145/3555776.3577594.

[9] M. Bertl, P. Ross, D. Draheim, Predicting psychiatric diseases using autoai: A performance analysis based on health insurance billing data, in: Database and Expert Systems Applications, Springer International Publishing, Cham, 2021, pp. 104–111.

[10] M. Bertl, N. Bignoumba, P. Ross, S. B. Yahia, D. Draheim, Evaluation of deep learning-based depression detection using medical claims data, SSRN (2023).

[11] M. Bertl, P. Ross, D. Draheim, Systematic ai support for decision making in the healthcare sector: Obstacles and success factors, Health Policy and Technology (2023). doi:https://doi.org/10.1016/j.hlpt.2023.100748.

[12] M. Bertl, J. Metsallik, P. Ross, A systematic literature review of ai-based digital decision support systems for post-traumatic stress disorder, Frontiers in Psychiatry 13 (2022). URL: https://www.frontiersin.org/articles/10.3389/fpsyt.2022.923613. doi:10.3389/fpsyt.2022.923613.

[13] M. Bertl, P. Ross, D. Draheim, A survey on ai and decision support systems in psychiatry – uncovering a dilemma, Expert Systems with Applications 202 (2022) 117464. URL: https://www.sciencedirect.com/science/article/pii/S0957417422007965. doi:https://doi.org/10.1016/j.eswa.2022.117464.

[14] M. Bertl, T. Klementi, G. Piho, P. Ross, D. Draheim, How domain engineering can help to raise decision support system adoption rates in healthcare, 2023.

[15] K. Kõnd, A. Lilleväli, et al., E-prescription success in estonia: The journey from paper to phamacogenomics, Eurohealth 25 (2019) 18–20.
[16] F. Colombo, J. Oderkirk, L. Slawomirski, Health information systems, electronic medical records, and big data in global healthcare: Progress and challenges in oecd countries, Handbook of global health (2020) 1–31.
[17] S. M. Meystre, C. Lovis, T. Bürkle, G. Tognola, A. Budrionis, C. U. Lehmann, Clinical data reuse or secondary use: current status and potential future progress, Yearbook of medical informatics 26 (2017) 38–52.
[18] C. Sáez, J. Martínez-Miranda, M. Robles, J. M. García-Gómez, Organizing Data Quality Assessment of Shifting Biomedical Data, Quality of Life through Quality of Information (2012) 721–725. doi:10.3233/978-1-61499-101-4-721, publisher: IOS Press.