

# Towards a Verification Approach of a Smart Registry for Chronic Heart Failure Patients

Christoph EBERHARTER<sup>a,b</sup>, Martin BAUMGARTNER<sup>a,b</sup>, Karl KREINER<sup>a</sup>, Aaron LAUSCHENSKY<sup>a</sup>, Robert MODRE-OSPRIAN<sup>c</sup>, Sabrina NEURURER<sup>d,e</sup>, Dieter HAYN<sup>a,f</sup>, Luca BRUNELLI<sup>g</sup>, Gerald SLAMANIG<sup>h</sup>, Sarah PRANTL<sup>h</sup>, Bernhard PFEIFER<sup>d,e</sup>, Gerhard PÖLZL<sup>g</sup> and Günter SCHREIER<sup>a,b,1</sup>

<sup>a</sup> AIT Austrian Institute of Technology, Graz, Austria

<sup>b</sup> Technical University of Graz, Institute of Neural Engineering, Graz, Austria

<sup>c</sup> telbiomed Medizintechnik und IT Service GmbH, Graz, Austria

<sup>d</sup> Tyrolean Federal Institute for Integrated Care, Tirol Kliniken GmbH, Innsbruck, Austria

<sup>e</sup> Division for Digital Medicine and Telehealth, UMIT TIROL - Private University for Health Sciences and Health Technology, Hall (Tyrol), Austria

<sup>f</sup> Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria

<sup>g</sup> Medical University of Innsbruck, Innsbruck, Austria

<sup>h</sup> Tirol Kliniken GmbH, Innsbruck, Austria

**Abstract.** Background: The daily increasing amount of health data from different sources like electronic medical records and telehealth systems go hand in hand with the ongoing development of novel digital and data-driven analytics. Unifying this in a privacy-preserving data aggregation infrastructure can enable services for clinical decision support in personalized patient therapy. Objectives: The goal of this work was to consider such an infrastructure, implemented in a smart registry for heart failure, as a comparative method for the analysis of health data. Methods: We analyzed to what extent the dataset of a study on the telehealth program HerzMobil Tirol (HMT) can be reproduced with the data from the smart registry. Results: A table with 96 variables for 251 patients of the HMT publication could theoretically be replicated from the smart registry for 248 patients with 80 variables. The smart registry contained the tables to reproduce a large part of the information, especially the core statements of the HMT publication. Conclusion: Our results show how such an infrastructure can enable efficient analysis of health data, and thus take a further step towards personalized health care.

**Keywords.** Verification, smart registry, data-driven healthcare, clinical decision support, chronic heart failure

---

<sup>1</sup> Corresponding Author: Günter Schreier, AIT Austrian Institute of Technology GmbH, Graz, Austria, E-Mail: guenter.schreier@ait.ac.at

## 1. Introduction

The growing volume of health data from a wide variety of sources and the ongoing development of big data analytics are leading to new approaches to improving healthcare. In addition to clinical patient data - such as demographics, diagnoses, procedures, laboratory reports, or images - large data sets can also come from other clinical data sources, such as telehealth programs. Applying artificial intelligence to these comprehensive sets of patient data can provide clinically relevant information. Features from these data sources can be used to create models for predicting a patient's health status and assist doctors in their decision-making. This enables healthcare professionals to provide cost-effective and personalized therapy and intervene early when patients' health deteriorates. To support point-of-care and real-time patient care, an infrastructure is needed in which data from different sources can be aggregated and analyzed in real-time. [1-3]

Based on a previous established telehealth program called HerzMobil Tirol (HMT) [4], a post-discharge disease management program for heart failure (HF) patients, a so-called "smart registry" for HF patients was created in a first step. In addition to "normal" registries, the "smart registry" is automatically synchronized with various sources, it supports pseudonymization and privacy preserving record linkage, it is standardized, and it supports personalized healthcare through smart decision support services. Health data of HF patients within the Tirol Kliniken GmbH (largest hospital provider in Tyrol) from 2016 until today have been entered into this registry. In this process, patient data from the telehealth program HMT were linked with clinical data of HF patients from the health information system (HIS) of the Tirol Kliniken GmbH. In addition, data from an extract of the Austrian death registry were imported. In-house tools for pseudonymization and record linkage, as well as the Observational Medical Outcome Partnership Common Data Model (OMOP CDM) from the Observational Health Data Sciences and Informatics initiative for standardization of data were used [5,6]. Furthermore, the infrastructure supports deployment of features and models for the prediction of the patient's health status from these data.

In a recently published study, the effectiveness and feasibility of the HMT program was analyzed by Gerhard Pözl et al. [7], which is further referred to as the "HMT study". The study analyzed the outcome of acute heart failure patients after hospitalization who participated in the HMT program or received usual care. The combined primary endpoint of this study was death from any cause and readmission for acute HF at 6 months. Secondary endpoints were the 1- and 3-month readmission for acute HF as well as 1-year all-cause mortality. The authors collected data for this study manually in a time-consuming and laborious manner from the various sources such as the electronic medical record (EMR) and physician letters.

The aim of this paper was to consider the smart registry as a comparative method for the safe and efficient analysis of health data. To this end, we analyzed the extent to which the smart registry is capable of replicating the data from the HMT study, thus presenting a less labor-intensive automated pathway. Verifying whether the smart registry was populated correctly with data from the different sources improves its quality for future applications. This was also to show how such an infrastructure could make health data available at the push of a button and in real time in the future. This could potentially increase the efficiency for future studies significantly and open up new possibilities in patient care.

## 2. Methods

### 2.1. Data overview

Two datasets were used for the present paper, which are described in the following.

On the one hand side, for the HMT study, a table was created over the period between 1 April 2016 to 31 October 2019, which contained the baseline characteristics and outcome data of 251 HMT patients and 257 control group (usual care) patients summarized in 96 variables. Patient data came from a standardized assessment, which included medical history, 12-lead ECG, and echocardiography. The table comprised demographic data (e.g., age, sex), physiological parameters (e.g., blood pressure, heart, weight), laboratory measurements (e.g., creatinine, NT-proBNP, sodium), and previous illnesses (diagnoses) at the time of HMT initiation, as well as data regarding hospitalization and death during and after the HMT program. Some variables were derived features to calculate, for example, follow-up data or the Charlson Comorbidity Index.

On the other hand, the smart registry currently contains health data on 4,680 HF patients. Of these, 4,174 are patients from the HIS of the Tirol Kliniken GmbH, for which only patient records between June 2022 and January 2023 were available at the time of this study. This contains information on demographic data, physiological measurements, laboratory measurements, diagnoses and time of hospital admission and discharge, as well as date and reason for death of patients in the hospital. Data from 506 patients are from the HMT program over the period April 2016 to June 2021. These include data transmitted on a daily basis by patients over three months on blood pressure, heart rate, body weight, well-being, and medication, as well as patient baseline data and free-text clinical notes. Additionally, the registry includes the date of death of deceased patients from an extract of the Austrian death registry.

Considering the time period of the two datasets, there was no overlap of data from the HIS in the smart registry with that of the HMT study. Thus, a direct comparison was not possible here, but a statement about a possible replicability of variables regarding these data could be made based on the existing tables in the smart registry.

The studies were approved by the ethics committee of the Medical University Innsbruck (vote nr. 1035/2022).

### 2.2. Data processing

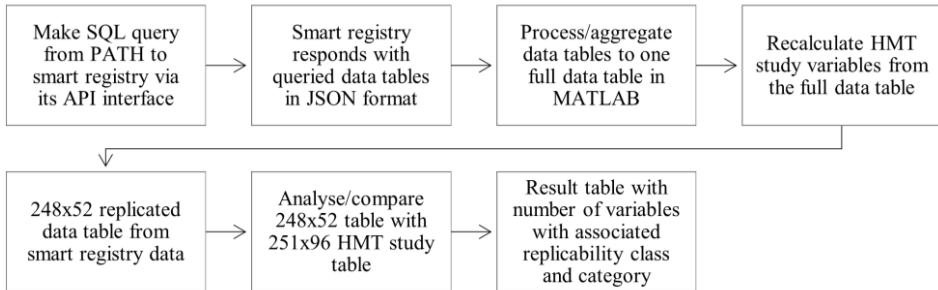
Data processing was primarily performed using the Predictive Analytics Toolset for Healthcare (PATH) developed by AIT, which supports data analysis and visualization based on MATLAB R2022a (The MathWorks, Inc., Natick, Massachusetts, United States). PATH accompanies the entire process of data processing, where each step can be specified using definition files in Microsoft Excel, and visualized in the associated PATH app with graphical user interface within MATLAB. [8]

Via an API interface from the smart registry's data store, SQL queries were used to load patient data tables in a JSON format. All relevant information for each patient could then be efficiently derived from these loaded data tables, as the data in the smart registry are managed in a standardized manner according to the OMOP CDM.

With the unique timestamp for the start of each patient's HMT program, 248 out of 251 patients in the HMT study could be matched in the smart registry. The 257 patients from the usual care group were not considered for this analysis. For a clear analysis, the

96 variables of the HMT study were divided into 9 categories as listed in Table 1. For each of the 248 HMT patients, all variables were recalculated as far as possible in MATLAB using the existing data from the smart registry.

An overview of the workflow of the methods is shown in Figure 1.



**Figure 1.** Overview of the workflow of the methods.

### 2.3. Data comparison

For the comparison of the two datasets, the data from the HMT study, which was available as an Excel-file, and the data replicated from the smart registry were converted into MATLAB tables.

According to Table 1, the quality of replicability of the 96 variables from the smart registry was graded into 4 classes, i.e.:

- Not replicable variables: Variables that could not be replicated from the smart registry tables because these tables did not exist.
- Theoretically replicable variables: Variables that would be replicable from the smart registry tables if these tables were appropriately populated with data and the time period of the data also matched that of the HMT study.
- Actually replicable variables: Variables that could be replicated from the smart registry tables, but the time period of the data did not match that of the HMT study.
- Correctly replicable variables: Variables that could be replicated from the smart registry tables, where the time period of the data matched that of the HMT study.

## 3. Results

As shown in Table 1, 80 out of the 96 variables in the HMT study could theoretically be replicated from the smart registry. For 16 variables in the categories doctor's letter and other sources, no tables existed in the smart registry at the time of this study, from which they could be replicated. 52 variables could actually be recalculated, but not for a matching period of time. 17 variables could be correctly recalculated according to the HMT study, which were found in the categories administrative, demographic and death data.

**Table 1.** Number of variables of each category of the HMT study, with the corresponding four classes of replicated number of variables from the smart registry.

Category	HMT study	Smart registry			
	Variables	Not replicable variables	Replicable variables (theoretical)	Replicable variables (actual)	Replicable variables (correct)
Administrative	4	0	4	3	3
Demographic	4	0	4	4	4
Physiologic	9	0	9	9	0
Laboratory	13	0	13	10	0
Hospitalization	12	0	12	6	0
Diagnosis	22	0	22	8	0
Doctor's letter	6	6	0	0	0
Other sources	13	10	3	0	0
Death	13	0	13	12	10
<b>Total</b>	<b>96</b>	<b>16</b>	<b>80</b>	<b>52</b>	<b>17</b>

#### 4. Discussion

In this paper, an attempt was made to reproduce data manually collected for a scientific publication (HMT study) via the path of a data aggregation infrastructure (smart registry). Theoretically, it was possible to replicate a large part (80 out of 96) of the variables of the paper using the smart registry. In practice, 52 variables with deviating data time period and 17 variables with correct data time period could be recalculated.

The 16 variables that could not (yet) be reproduced from the smart registry were, for example, individual reasons for disease, smoking behavior and left ventricular ejection fraction (LVEF). For their publication, this information was extracted manually from doctor's letters or from subsystems within the Tirol Kliniken GmbH EMR which are not yet linked to the smart registry. In order to approximately replicate these types of variables, unstructured information of the doctor's letters would probably have to be annotated and made available in the smart registry. Automated annotation of free texts is a very difficult task and needs a lot of further research in the future. Such natural language processing methods could be applied on free-text documents in the EMR or on freetext clinical notes from the HMT telehealth service. From those HMT notes, some variables, such as information on the patient's smoking behavior, could potentially also be traced in this way. The LVEF of patients is stored in a separate subsystem at Tirol Kliniken GmbH. A connection of these data with the smart registry is already being considered for the future.

In the direct comparison of the reproduced variables with those of the HMT study, significant differences were found, which has several reasons. This is mainly due to the fact that HIS data for the HMT study period were not available at the time of analysis, so most variables could not be directly compared. This concerns data in the categories physiological, laboratory, hospitalization, and diagnosis. It is also possible that the place of origin of the information of some variables from the smart registry differs from that of the HMT study (e.g., documented in a structured way vs. documented in doctor's letters only). As a result, a variable can be reproduced qualitatively, but differs quantitatively from the value of the HMT study. For example, certain diagnoses may have been entered at an earlier time in the HIS of the Tirol Kliniken GmbH and therefore not appear in the HIS export in the smart registry, or the diagnosis was basically never documented in the HIS, but only in the doctor's letter.

The infrastructure of the smart registry is theoretically capable of reproducing almost all information considered in the HMT publication. In particular, the study variables necessary for the primary and secondary outcome (hospitalization and death data) can be reproduced. The harmonization of medical data from different sources with the help of the OMOP Common Data Model will certainly make a significant contribution to this. Verification of the smart registry data by comparing with a quality-controlled data source, as the one used for the HMT publication, has the potential to identify a) missing relevant information and b) potential aberrations of the data. By that, the overall data quality of such smart registries can significantly be improved.

The need for such an infrastructure for the aggregation of health data, to whichever specific area it is applied, is almost universal, with obvious advantages. Data analysts can use such a smart registry to access and work with features and models of patient data at any time with just a few clicks. Eventually, insights and also predictive or classification models derived from this data can effectively support healthcare professionals in their treatment of patients. The implementation of such an infrastructure requires a great amount of resources but is indispensable for the future of data-driven healthcare.

## Acknowledgement

Parts of this work were supported by the state of Tyrol, in the framework of the project “d4Health Tirol”.

## References

- [1] Fei Jiang et al. “Artificial intelligence in healthcare: past, present and future.” In: *Stroke and Vascular Neurology* 2.4 (2017), pp. 230–243.
- [2] Pavel Hamet and Johanne Tremblay. “Artificial intelligence in medicine.” In: *Metabolism: Clinical and Experimental* 69 (Apr. 2017), S36–S40.
- [3] Kain Kordian Gontarska et al. “Predicting Medical Interventions from Vital Parameters: Towards a Decision Support System for Remote Patient Monitoring.” In: *CoRR* abs/2104.10085 (2021).
- [4] Elske Ammenwerth et al. “HerzMobil, an Integrated and Collaborative Telemonitoring-Based Disease Management Program for Patients With Heart Failure: A Feasibility Study Paving the Way to Routine Care.” In: *JMIR Cardio* 2.1 (2018), e11.
- [5] Michael Nitzlner and Günter Schreier. “Patient identity management for secondary use of biomedical research data in a distributed computing environment.” In: *Studies in health technology and informatics* 198 (2014), pp. 211–8.
- [6] OHDSI Observational Health Data Sciences and Informatics. *The Observational Medical Outcome Partnership (OMOP) Common Data Model (CDM)*. [Online; accessed 21-January-2023]. URL: <https://www.ohdsi.org/data-standardization/>.
- [7] G. Poelzl et al. “Feasibility and effectiveness of a multidimensional post-discharge disease management programme for heart failure patients in clinical practice: the HerzMobil Tirol programme.” In: *Clinical Research in Cardiology* 111 (2022), pp. 294–307.
- [8] Dieter Hayn et al. “Predictive analytics for data driven decision support in health and care.” In: *it - Information Technology* 60.4 (2018), pp. 183–194.