



OPEN ACCESS

Pharmacotherapy within a learning healthcare system: rationale for the Dutch Santeon Farmadatabase

Ewoudt M W van de Garde,^{1,2} Bram C Plouvier,² Hanneke W H A Fleuren,³ Eric A F Haak,⁴ Kris L L Movig,⁵ Maarten J Deenen,⁶ Marinus van Hulst^{2,7}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/ejpharm-2017-001329>).

¹Department of Clinical Pharmacy, St Antonius Hospital, Nieuwegein, The Netherlands

²Santeon, Utrecht, The Netherlands

³Department of Clinical Pharmacy, Canisius Wilhelmina Hospital, Nijmegen, The Netherlands

⁴Department of Clinical Pharmacy, OLVG, Amsterdam, The Netherlands

⁵Department of Clinical Pharmacy, Medisch Spectrum Twente, Enschede, The Netherlands

⁶Department of Clinical Pharmacy, Catharina Hospital, Eindhoven, The Netherlands

⁷Department of Clinical Pharmacy, Martini Hospital, Groningen, Netherlands

Correspondence to

Dr Ewoudt M W van de Garde, Department of Clinical Pharmacy, St. Antonius Hospital, 3435 CM Nieuwegein, The Netherlands; e.van.de.garde@antoniusziekenhuis.nl

Received 8 June 2017

Revised 16 August 2017

Accepted 22 August 2017

Published Online First

18 September 2017

EAHP Statement 1: Introductory Statements and Governance.
EAHP Statement 4: Clinical Pharmacy Services.



To cite: van de Garde EMW, Plouvier BC, Fleuren HWHA, et al. *Eur J Hosp Pharm* 2019;**26**:46–50.

ABSTRACT

Objectives The increasing number of available, often expensive, medicines asks for continuous assessment of rational prescribing. We aimed to develop a simple and robust data infrastructure in order to monitor hospital medicine utilisation in real time.

Methods Within a collaboration (Santeon) of large teaching hospitals in the Netherlands, we set up a process for extraction, transformation, anonymisation and load of individual medicine prescription data and major clinical outcomes from different hospital information systems into a central database. Quarterly reports were constructed to monitor and validate the quality of the uploaded data.

Results A central database has been developed that includes data from all patients from 2010 onwards and is refreshed on a weekly basis by an automated process. Beginning in 2017, the database holds data from almost 800 000 patients with prescriptions. All hospitals provide at least 18 mandatory data items per patient. Provided data include, among others, individual prescriptions, diagnosis data, and hospitalisation and survival data. The database is currently used to benchmark the level of biosimilar prescribing and to assess the impact of novel systemic treatments on survival rates in metastatic cancers.

Conclusion We showed that it is feasible for a group of hospitals to construct their own database that can serve as a tool to benchmark the positioning of medicines and to start with monitoring their impact on clinical outcomes.

INTRODUCTION

The increasing awareness that traditional clinical trials are more and more difficult and expensive to conduct, and the fact that external validity of clinical trial findings is limited, have paved the way for complementary strategies to find answers on what treatment is most likely to provide best clinical outcomes. One of those strategies is to integrate the collection of relevant key parameters within routine clinical practice, in order to obtain a useful context that allows to evaluate the added value of new treatment options relative to the past. This is often referred to as a learning healthcare system.¹

In early 2012, Santeon, a group of large teaching hospitals in the Netherlands, have joined forces to construct such a learning healthcare system. Because treatment with medicines is considered an important therapeutic intervention, the

development of a normalised database containing all prescribed medicines and major clinical outcomes of all their patients was one of the first parts to be completed within this context. With this information readily available, the introduction of new treatments can be monitored for impact on relevant outcomes, plus it allows to benchmark the positioning of medicines between the hospitals. Driven by the close and strong network of hospital pharmacists, Santeon succeeded in developing such database within 4 years.

This paper describes the characteristics and origin of the data available in the Santeon Farmadatabase, the challenges encountered to gather the data and how these data can be applied to improve pharmacotherapy within a learning healthcare system.

METHODS

Setting

Santeon (established in 2007) started as a group of six large teaching hospitals geographically spread across the Netherlands. In 2017, a seventh hospital became member of the group. Currently, the Santeon group of hospitals provides care to 13% of the Dutch population. The main reason for establishing the Santeon network was to trigger collaboration on quality of care not influenced by local market issues. The Santeon hospitals are Catharina Hospital Eindhoven, Canisius Wilhelmina Hospital Nijmegen, Martini Hospital Groningen, Medisch Spectrum Twente Enschede, OLVG (East and West) Amsterdam, Maasstad Hospital Rotterdam (member since 2017) and St Antonius Hospital Utrecht/Nieuwegein. [Figure 1](#) shows the geographical locations of the individual hospitals, and [table 1](#) provides an overview of activity data per hospital.

After a feasibility study conducted in two hospitals (Martini Hospital and St Antonius Hospital) in 2012 (supported by the Netherlands Organisation for Health Research and Development (grant number 836001003), Santeon decided to create the Farmadatabase with the aim to collect unequivocal prescription, diagnosis and outcome data from all participating hospitals on a continuous basis. The aim was to include in the database all medicines prescribed during hospital stay and in day-care setting and, for the outpatient clinics, all prescribed medicines that are part of the hospital financial budget. For the Netherlands, the latter includes all oncology medicines, fertility medicines, expensive immunosuppressants and growth hormones.

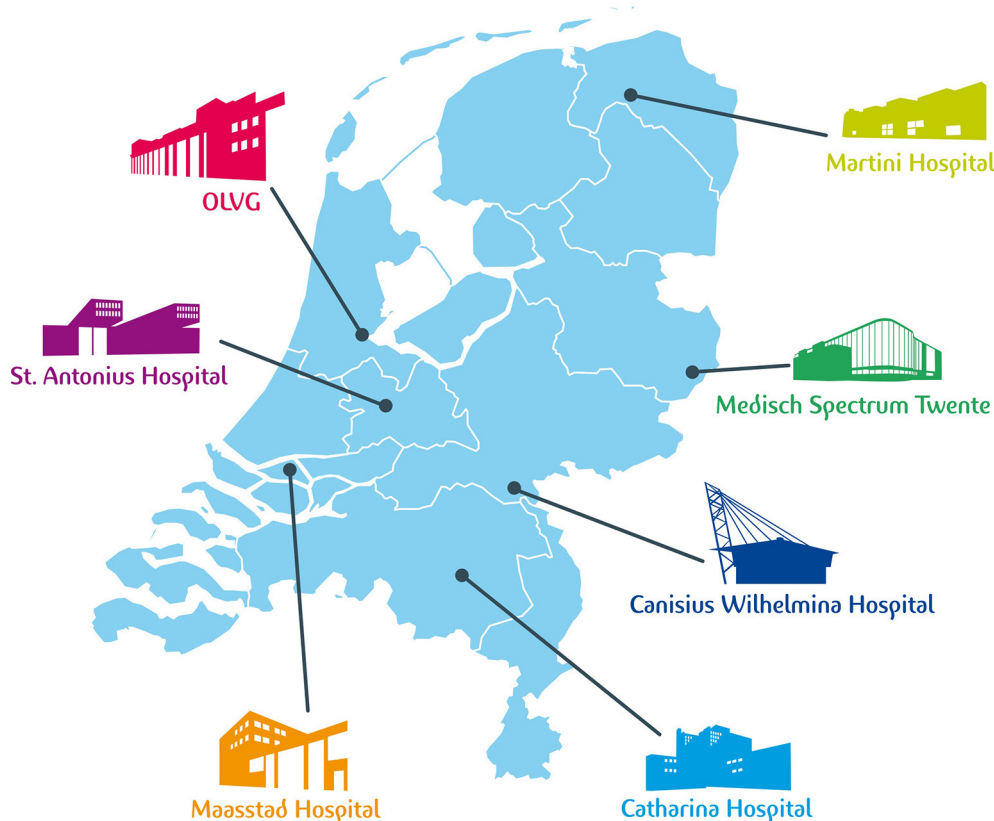


Figure 1 The locations of the Santeon hospitals in the Netherlands.

Data flow

From the start of the project it was aimed to extract the relevant data directly from electronic health record data sources used in day-to-day routine clinical care. This approach prevents double entry efforts and provides the unique possibility to evolve together with the information technology (IT) backbone of the hospitals. For each hospital, effort has been made to identify the applicable source applications to be able to extract, transform and load the requested individual patient data to a central data warehouse. The full data specification is included in online supplementary appendix 1. Data extraction is scheduled every week, encompassing all records changed within the past 3 weeks' period (new entry or modification) after a single loading of historical data back from 1 January 2010. In more detail, the data flow starts with extraction from the relevant data fields from the source data into a staging table. Next, several data items are normalised into categories agreed. An example is that all administration routes from the source data are mapped onto a standard list. The script doing this categorises values not

included in the mapping table as 'unknown'. Subsequently, data formats are transferred to the requested format (eg, MM-DD-JJJJ to ddmmyyyy). The individual hospitals use different tools to create their extracts; some use reporting tools such as SAP Crystal Reports, while others use tools such as Microsoft SQL Server Integration Services for this task. In the final phase, the extraction yield is sent to the Santeon server as a fixed-length text file. The fixed-length file format was chosen because it is a human readable format that can be produced by most data processing tools. As an alternative, the comma-separated values (CSV) file format was also considered. CSV was dropped in favour of the fixed-length format because different tools handle the escaping of the field separating character (delimiter) in different ways, which we expected to cause data parsing errors when loading the data into the central database. Subsequently, in the central data warehouse, all submitted text files are checked for integrity and added to a separate database for each individual hospital (see figure 2). The integrity checks ensure that the data conform to the data types as per the specification, that there is

Table 1 Hospital metadata for 2015

Hospital	Beds (n)	Admissions (n)	Nursing days (n)	Specialists (FTU) (n)
CWZ	455	29 040	139 153	177
Catharina	696	26 487	147 462	220
Martini	578	25 887	132 024	215*
MST	739	31 739	172 037	235
OLVG	1106	40 549	195 172	325
St Antonius	853	38 842	213 061	372*

Maastricht Hospital not mentioned because it is still in the process of connecting to the database.

*Absolute number of specialists, not full time units (FTU).

CWZ, Canisius Wilhelmina Hospital; MST, Medisch Spectrum Twente.

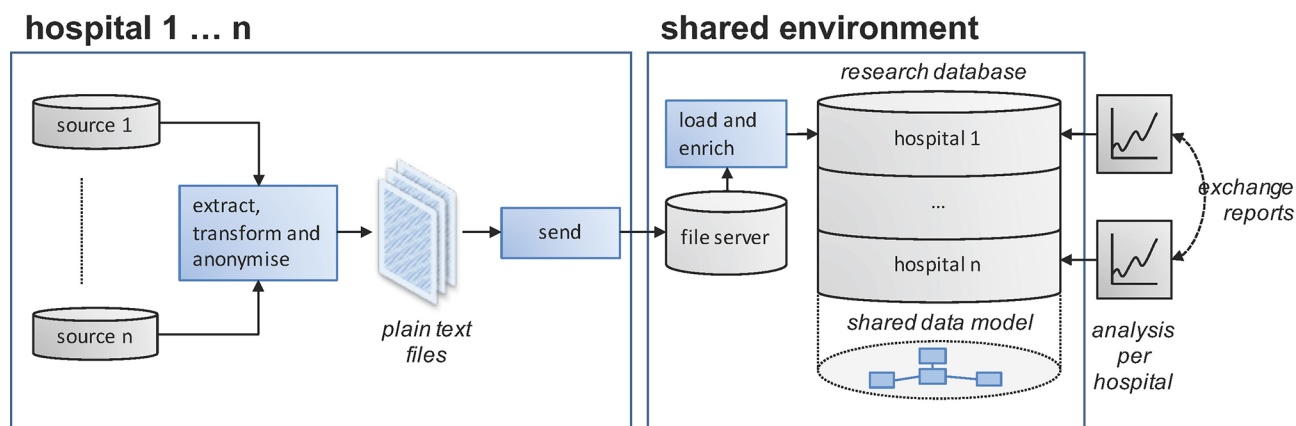


Figure 2 Data flow from individual hospitals towards shared environment.

only one record present in the files for each unique key and that patients being referred to exist in the demographic information table. One or more errors in the submitted text file will result in rejection of the total file and an email alert to the central data manager. Now the start-up phase is behind us, error rates are generally low. Last quarter (Q1 2017) a little over 2% of the submissions contained integrity errors, which were swiftly fixed by the hospitals involved. During the adding of data, any modified items within records (identified by their unique identifiers) are overwritten in the central database. In the final phase of data loading, the submitted records are enriched with data values from other sources (eg, information about the daily defined dose is added based on the Anatomical Therapeutic Chemical (ATC) code submitted and the current WHO ATC index).² Because not all source systems are equipped to record the deletion of records, all hospitals are requested to send in an additional file containing all unique record identifiers currently present in their source systems every 3 months. After submission of that specific file, a procedure will run that removes all records with a unique identifier not present in that additional file from the central database to a 'deleted records' table. This procedure will result in a central database representative for the source data with maximum validity at least from 3 months in the past backwards.

Validation of data

In the data management process, validity of the data is monitored by different means. First, during the construction phase, the historical data dump of each hospital has been extensively checked for inconsistencies. Per hospital, the total number of unique records in the source systems was counted to exclude the possibility of missing records in the first yield. Next, pivot tables were constructed that show what records are retained in the extraction and which are not and for what reason. An example is the exclusion of medication orders not yet electronically signed by a physician. Following this step, all variables were scanned for missing data (empty fields), and variables that require normalisation were judged for the amount of 'unknown' as category. Ideally, the 'unknown' category should be very rare. Because ATC is an important value during the exploration of the database, specific attention was made to have missing ATC values only for medication records that cannot have an ATC (eg, combination products and some extemporaneous preparations). The latter required that some data corrections have been made in the source systems (eg, add ATC to a product sheet).

Second, after the initial validation, every 3 months a standard feedback report is sent to the participating hospitals containing

information about the number of submitted records, descriptive data (eg, number of different values, min, max) and an overview of records that could be incorrect (eg, date of medication prescription after date of death) from the last 3 months. For comparison purposes, the report also contains data from the preceding quarter and from the same period 1 year earlier. An important part of the report is a count on the level of missing ATCs and the percentages of 'unknowns' in normalised fields. As an example, an increase in the percentage of medication orders with unknown administration route likely indicates that new administration routes are introduced in the source system that are not yet included in the mapping table. During the past years, we experienced that, with the system of quarterly feedback and repair in place, missing ATC and 'unknown' categories can remain below 1% of the data.

Third, besides the feedback reports, each hospital is also requested to enter into its source system a standard set of medication orders for three dummy patients every 6 months. These dummy patients are used to check the process of extracting, transforming and loading data into the central data warehouse for correctness.

RESULTS

On 12 April 2017, the Santeon Farmadatabase kept 18 741 766 medication orders, 3 366 162 hospital admissions and 15 930 414 diagnoses from 798 632 unique patients with prescriptions (the total number of patients in the database is over two million because patients can have only outpatient clinic visits without a prescription that is part of the hospital financial budget). Figure 3 shows an overview of the relative share of individual ATC groups (first level and second level). This information provides a fingerprint of the types of medicines used within the hospitals nowadays. As expected some obvious hospital-specific groups of medicines are present (eg, oncology medicines).

Since start of the project, several treatment benchmark reports have been constructed already. These include a benchmark report that shows per day of admission of patients with community-acquired pneumonia the level of patients who are switched from intravenous to oral antibiotics. Benchmarking the hospitals on this parameter showed a 1-day difference in median time to switch between some hospitals. Another example is a benchmark report on the level of biosimilar prescribing in newly diagnosed patients with colitis ulcerosa or Morbus Crohn. Linked to anti-tumour necrosis factor medication, we also build a report on the level of switching in these patients. This provided insight

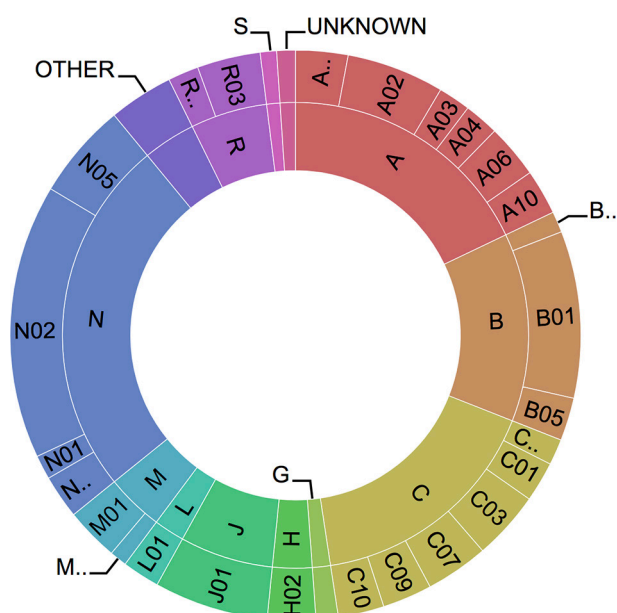


Figure 3 Sunburst Anatomical Therapeutic Chemical profile (based on all medication orders present in the database) with first level (inner circle) and second level (outer circle).

that switching due to ineffectiveness is a rare phenomenon, indicating that the first choice treatment decision has long-standing impact.

Besides the benchmark reports, an interactive medicines utilisation heatmap has been developed. This heatmap can show for every individual hospital the proportion of a chemical subgroup (ATC fourth level) within a pharmacological subgroup (ATC third level) compared with the proportion when all hospitals are grouped together. A colour reflects how much the proportion in an individual hospital deviates from the overall proportion (eg, bright yellow means that the proportion is much higher than average in the Santeon group, whereas dark blue indicates a much lower proportion). As an example, figure 4 shows that in hospital 1 the proportion of interleukin inhibitor (L04AC) users is 0.47% of the patients using immunosuppressants (L04A...) compared with 9.26% in hospital 3.



Figure 4 Snapshot of the interactive pharmacotherapy heatmap. The heatmap shows per hospital the relative share of a chemical subgroup (ATC fourth level) within a pharmacological subgroup (ATC third level) with colours changing depending on how the proportion deviates from the proportion in all hospitals grouped together (the legend shows how the colour corresponds with the ratio between the hospital proportion (%zkh) and the overall proportion (%tot)). Calculations are based on number of patients. Each column represents one hospital and sums up to 100%. ATC, anatomical therapeutic chemical.

The centralised Santeon Farmadatabase does not contain any patient-specific data that could pose a risk to patients' privacy. The only personal data included in the database are gender, year and month of birth. These items are allowed for storage without a formal approval from the Dutch Data Protection Authority and an informed consent from the patient is not required.^{3,4} The latter is under the condition that all hospitals have an operational opt-out procedure for patients who do not want their data to be used for scientific research. Patients who have opted out are actively filtered during the data uploading process. During loading to the central data warehouse, all original hospital patient identification numbers are replaced by a dummy number generated using through a 36-bit (or more) MD5 hash algorithm per hospital. With this procedure in place, there remains a possibility for an individual hospital to translate their hashed identifiers back to the original hospital patient number for validation purposes or with the purpose of collection of additional clinical data. For the latter, a separate procedure is in place that is in accordance with all current Dutch laws and regulations for medical research, including the need for an informed consent from the patient when possible and data transfer via a trusted third party. A drawback from our handling of patient identifiers can be that a patient who has visited two Santeon hospitals gets two entries in the database. Given the geographical locations of the hospitals, this scenario is expected to be very rare and it will not impact any analysis on in-hospital treatment planning because patients cannot be hospitalised at two locations simultaneously. When counting absolute numbers of patients on specific medication within the Santeon group, there remains a possibility for some overestimation that should be considered.

DISCUSSION

Starting 2017 the Santeon Farmadatabase is fully operational, meaning that individual prescription data from six hospitals are available in the central database from 2010 onwards, and are refreshed on a weekly basis. Although recently established, we believe that having these data timely and in a uniform format largely broadens the opportunities for direct observations on trends in in-hospital medicine utilisation and studying these trends in relation to relevant patient outcomes. The data include both all in-hospital prescriptions as well as all outpatient clinic prescriptions that are part of the hospital financial budgets in the

Netherlands. Outpatient prescriptions that are started by general practitioners are not available in the database.

The main challenge in this project was to implement the correct data extraction per hospital. All hospitals had different health record keeping and prescribing systems in place, and some of the hospitals migrated to another IT backbone during the period of observation. This required very close collaboration between business intelligence experts and hospital pharmacists to locate the correct data origins and to let it correctly move to the final data upload. Altogether, we think that the strong Santeon network, the support by the board of directors and medical staff, and the in-depth knowledge of hospital pharmacists about the prescribing process, together with skilled business intelligence experts, made us succeed to get the Santeon Farmadatabase fully operational beginning in 2017.

The learning healthcare system philosophy is built on continuously evaluating healthcare interventions in relation to patients' outcomes to identify what works best in what situation. Obviously, this is very relevant for pharmacotherapy with its increasing complexity and costs. We are convinced that by creating the Santeon Farmadatabase infrastructure, an important first step has been made towards implementing the learning healthcare system in the direction of pharmacotherapy, primarily because a complete and valid overview of all applied medicines initiated in the hospitals is now available together with granular data on diagnoses and hospital admissions. In the next phase, this information can be linked to more detailed outcome data collected in other Santeon databases. An example of such is a project that has been initiated recently, linking highly detailed chemotherapy data from the Farmadatabase to relevant patient characteristics and outcomes in lung oncology collected within

the Santeon Care for Outcome programme. These grouped data will be used to explore the impact of novel systemic treatment options on survival in patients with metastatic non-small cell lung cancer and to search for patient characteristics predictive for clinical effectiveness of the medicines. Another future direction is the incorporation of costs of medicines in the data analysis to proceed to value-based healthcare and pay for performance.

In conclusion, we showed that it is feasible for a group of hospitals to construct their own database that can serve as a tool to benchmark the positioning of medicines and to start with monitoring their impact on clinical outcomes.

Acknowledgements The authors greatly acknowledge Michiel Deerenberg and Kasper de Graaf for their valuable contribution in the starting phase of the project, and Erik van Delden for developing the quarterly validation reports.

Contributors EMWvdG and MvH designed the project, collected and interpreted data, and drafted the manuscript. HWHAF, EAFH, KLLM and MJD collected and interpreted data, and revised the manuscript for important intellectual content. BCP analysed data and drafted the manuscript. All authors approved the final version of the manuscript.

Funding This project received funding from the Netherlands Organisation for Health Research and Development (ZonMw) with grant number 836001003.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© European Association of Hospital Pharmacists (unless otherwise stated in the text of the article) 2019. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- 1 Institute of Medicine. In: Olsen L, Aisner D, McGinnis JM, eds. *The learning healthcare system: workshop summary*. Washington, DC: National Academies Press, 2007. <https://www.nap.edu/catalog/11903/the-learning-healthcare-system-workshop-summary> (accessed 4 Apr 2017).
- 2 World Health Organisation Collaborating Centre for Drug Statistics Methodology. Anatomical therapeutic chemical classification system. https://www.whocc.no/atc_ddd_index/
- 3 Dutch data protection exemptions decree. <http://wetten.overheid.nl/BWBR0012461/2017-01-01> (accessed 8 May 2017).
- 4 The Code of conduct for the use of data in health research. https://www.federa.org/sites/default/files/bijlagen/coreon/code_of_conduct_for_medical_research_1.pdf (accessed 8 May 2017).

What this paper adds

What is already known on this subject

Hospitals need real-time data to be able to follow up and benchmark real-world treatment medicine utilisation.

What this study adds

This study provides an example of how a group of hospitals constructed a database that is continuously filled with prescription practice data in order to be able to benchmark the positioning of medicines and their impact on clinical outcomes.