Cross-Border Challenges in Informatics with a Focus on Disease Surveillance and Utilising Big-Data
L. Stoicu-Tivadar et al. (Eds.)

doi:10.3233/978-1-61499-389-6-85

Belgian Primary Care EPR: Assessment of Nationwide Routine Data Extraction

Etienne DE CLERCQ[a1], Viviane van CASTEREN[b], Nathalie BOSSUYT[b], Geert GODERIS[c] and Sarah MOREELS[b]
[a] *Research Institute for Health and Society (IRSS), UCL, Belgium*
[b] *Scientific Institute of Public Health (IPH), Unit of Epidemiology, Brussels, Belgium*
[c] *Academisch Centrum voor Huisartsgeneeskunde, KU Leuven, Belgium*

Abstract. Starting in 2009, the first ever Belgian nationwide data collection network using routine data extracted from primary care EPR (upload method) has been built from scratch. The network also uses a manual web-based data collection method. This paper compares these two methods by analysing missing and most recent values for certain parameters. We collected data from 4954 practices, pertaining to 29,180 patients. Mean values for the most recent parameters were similar regardless of which data collection method was used. Many missing recent values (>46%) were found for all of the parameters when using the upload method. It seems that, in Belgium, uploading routine data from primary care EPR on a large scale is suitable and allows the collection of chronological retrospective data. However, the method still requires major, carefully controlled improvements.

Keywords: Computerised Patient Record, Primary Healthcare, Data Collection.

Introduction

In 2009, the Belgian National Institute for Health and Disability Insurance (NIHDI) launched a complex intervention in primary care based on the chronic care model [1], called “care trajectories”. As currently defined by the NIHDI, it organises and coordinates the treatment and follow-up for patients with Type 2 diabetes mellitus (DM-2) and chronic renal failure (CRF). The process starts with a contract being signed between the patient, the General Practitioner (GP), and a specialist.

Since 2009, the NIHDI has also funded the national ACHIL project (Ambulatory Care Health Information Laboratory), which assesses the quality of care delivered to patients covered by these care trajectories and monitors certain aspects of patients’ health status (for instance HbA1c and eGFR) at an aggregated level.

Data collected for this project would ideally have been primarily extracted from GPs’ EPR (Electronic Patient Records). However, using routinely collected general practice data for secondary usage (such as quality of care assessment), remains challenging [2-6]. Moreover, over 17 different software systems are currently being used by almost 10,000 practicing Belgian GPs, which hinders the rapid development of a secure primary care EPR-based research network.

---

[1] Corresponding Author.

This context has led the NIHDI to adopt a step-by-step approach. The first step (2009-2013) was to build a two-fold web-based data collection system that supports standardised forms for manual data capture as well as standardised data extraction from the EPR (upload). GPs can therefore freely choose their preferred method.

In this paper, we compare the two data collection methods and draw conclusions that should aid the future development of the national research network.

1. Methods

Recording patient data in the EPR and sending selected data to a research centre is foreseen by the care trajectory contract. The care trajectory program was started on 1 June 2009 for CRF patients and on 1 September 2009 for DM-2 patients.

Data was collected between May and September 2012 for all patients whose care trajectory began before 31 December 2011 and to cover a period starting the year prior to the start of their care trajectory and ending on 31 December 2011. Age and gender was collected for all patients. For DM-2 patients, additional data was collected relating to the following parameters: weight, height, systolic and diastolic blood pressure, LDL-cholesterol, and HbA1c. For CRF patients, additional data was collected relating to the following parameters: eGFR, systolic and diastolic blood pressure, PTH (parathormone), Hb (haemoglobinaemia), creatininaemia, and presence of Type 2 diabetes.

After identification and authentication, GPs connected to a specific web application hosted by a trusted third party (TTP) to send the data. GPs either entered patient data themselves or chose to extract the data from the EPR and upload it to the web application. The web application then performed validation tests (e.g. on data format and ranges). Following successful controls, an encrypted XML message was prepared for the research team that included only coded clinical data. Details of ACHIL privacy protection procedures have already been published [7].

In order to support the uploading procedure, each EPR software developer could decide to develop a specific extraction module, based on detailed specifications issued by the TTP in May 2012. Quality control of the extraction module had to be performed by the developers themselves. Additional data quality control procedures were applied at the research centre to control the entry of raw data (xml files) to the research database and to counter double entry of patients or parameters.

To compare the manual and upload data collection methods, we first report on the numbers of GPs, patients, contacts, and observations. We then analyse the number of missing values and the mean values for each parameter and outline some conclusions that may support the future development of the ACHIL network.

2. Results

We received usable data from 4954 practices, relating to 29,180 patients. This accounted for 77% of all patients involved in a care trajectory prior to 31 December 2011. Table 1 shows the basic figures for the research database. For technical reasons, in this paper, we discarded a small amount of data from a third data collection method called the “Réseau Santé Wallon”, which included 1353 care trajectory (CT) patients and 414 GPs. This did not affect our findings.

Table 1. Number of practices, patients, contacts, and observations by data collection method

Methods	Practices (N)	Patients (N)	Contacts (N)	Observations (N)
Manual	4600	26.444	40.203	199.057
Upload	354	2736	42.242	106.945
Total	4954	29.180	82.445	306.002

In Table 1, practices using both methods (1.03%) are counted twice. The percentages of patients that were registered twice, by different GPs or by different methods, were 0.77% and 0.20% respectively. In our database, 17,141 patients were enrolled in a DM-2 care trajectory (CT-DM2) and 12,629 patients were enrolled in a CRF care trajectory (CT-CRF). Two percent of these patients (590) were enrolled in both CTs. The median number of patients registered by practice (min: 1; max: 65) was four for the manual method and six for the upload method. For the manual method we found 1.31 observations by parameter and practice (mean value). This mean value rose to 7.4 for the upload method. Data was collected during a mean data collection period of 28 months. For both methods, the starting dates of the CTs were homogeneously spread over the entire observation period, with the exception of approximately 11% of the CTs using the upload method, which started on the first day of the official CT program. Figure 1 shows the contact registration dates by data collection method.

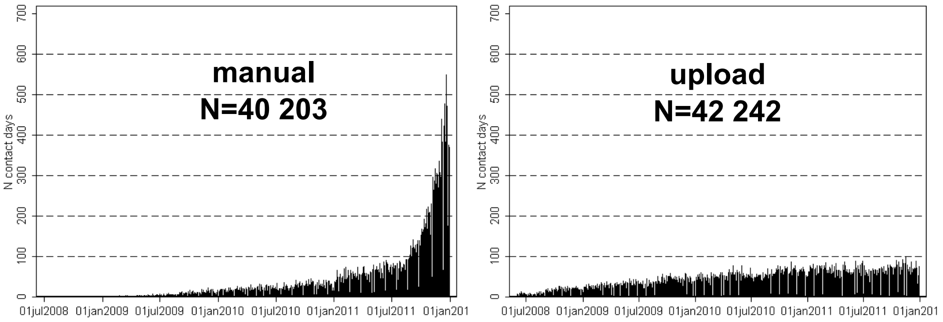

Figure 1. Number of contacts grouped by registration date and presented by registration method.

The mean age of the patients was 68 for CT-DM2 patients and 75 for CT-CRF patients, for both methods. Using the manual method, we found that 51.64% of CT-DM2 patients and 52.32% of CT-CRF patients were male and that 34.02% of CT-CRF patients were diabetic. Using the upload method, we found that 50.94% of CT-DM2 patients and 51.08% of CT-CRF patients were male and that only 5.31% of CT-CRF patients were diabetic.

Table 2 shows the numbers and percentages of missing most recent values for each parameter, by data collection method. Two time periods, one of four months and one of 12 months, were used to define recent values. For CT-DM2 patients, recent data (<4months) was missing for all the parameters for 25% of patients when using the manual method and for 12.94% of patients when using the upload method. These figures dropped to 4.90% and 2%, respectively, when a 12-month period was considered. For CT-CRF patients, recent data (<4months) was missing for all parameters for 27.90% of patients when using the manual method and for 19.15% of patients when using the upload method. These figures dropped to 9.10% and 5.56%, respectively, when a 12-month period was considered.

Table 2. Percentages of missing parameters, by data collection method and data collection period

DM-2 Care trajectory	Manual (N=15 541)		Upload (N=1600)		CRF Care trajectory	Manual (N=11 423)		Upload (N=1206)	
	missing < 4 month	missing < 12 month	missing < 4 month	missing < 12 month		missing < 4 month	missing < 12 month	missing < 4 month	missing < 12 month
HbA1c	29.19%	9.81%	57.81%	52.44%	**eGFR***	32.50%	14.75%	61.44%	53.98%
LDL-chol.	35.17%	17.30%	80.31%	76.75%	**Syst. BP**	29.82%	11.52%	54.73%	48.92%
Syst. BP	27.08%	7.53%	52.81%	47.06%	**Diast. BP**	29.87%	11.61%	54.81%	49.25%
Diast. BP	27.12%	7.58%	52.88%	47.19%	**PTH**	58.25%	48.47%	89.14%	87.15%
BMI	35.72%	18.67%	73.81%	70.50%	**Hb**	37.35%	20.89%	62.35%	55.14%

* Calculated or registered eGFR; LDL-chol.: LDL-cholesterol; Syst. BP: Systolic blood pressure; Diast. BP: Diastolic blood pressure; BMI: Body Mass Index; PTH: Parathormone; Hb: Hemoglobinemia

Mean values of most recent parameters were similar regardless of which data collection method and data collection period were used. Hereafter, we will refer only to the figures for the manual data collection method and the 4-month data collection period. For CT-DM2 patients, mean values were: HbA1c: 7.51% (+/- 1.19); LDL-Cholesterol: 89.48 mg/dl (+/-33.63); systolic BP (blood pressure): 133.69 mmHg (+/- 14.91); diastolic BP: 77.82 mmHg (+/- 8.63), and BMI: 30.54 Kg/m² (+/-5.63). For CT-CRF patients the mean values were: registered eGFR: 31.17 ml/min.1.73m² (+/- 11.4); systolic BP: 133.55 mmHg (+/- 17.2); diastolic BP: 76.6 mmHg (+/- 9.39); PTH: 99.07 pg/dl (+/- 83.25), and Hb: 12.35 g/dl (+/- 1.83).

3. Discussion

Starting in 2009, it took almost three years to come up with an accepted network architecture that met an acceptable level of privacy protection [7]. Following this, only four months were left for the development of software extraction modules and data capture. Time was therefore lacking to set up a standardised quality control procedure for the extraction modules. Some GPs reported problems with installing these modules. At the TTP, a strong quality control procedure rejected all the messages without the correct format or with non-authorised values. GPs who had been using EPR for many years may have been reluctant to use a manual data collection procedure instead. All of this may partly explain why we did not receive any data for over 20% of the patients involved in a care trajectory and why 11% of patients started their care trajectory on the official start date of the care trajectory program, which was probably used by the software developers as "default" start date. However, confirmation of these hypotheses would require further investigations that have not yet been planned.

Our results clearly show that more contacts per patient and more observations by parameter were obtained when the upload method was used. This was even more obvious when missing data was taken into account. The upload method therefore seems more suited to obtaining chronological retrospective data.

Only 5.31% of CT-CRF patients were recognised as diabetic patients by the upload method. Based on previous results (using strong quality control procedures), some underestimation related to an EPR under-recording issue was expected (expected value around 24%) [5]. The current underestimation might therefore be largely related to the quality level of the extraction modules.

As expected, the number of missing most recent values decreased considerably when a 12-month data collection period was used. However, the number of missing most recent values was still unexpectedly high when the manual method was used. For example, using data from a national billing database (2010), we found that 97% of the

CT-DM2 patients had had their HbA1c measured in 2010. We found no indication of lower percentages in 2011. For the manual method, missing data could be partly explained by data being no longer available at GPs' offices (lab results having been given to the patients) or by a lack of communication between healthcare actors (lab tests having been ordered by another physician and the results not communicated to the treating GP). This, however, requires further investigation.

For the upload method, missing values for all of the parameters are unexpectedly high [5]. In addition to a well-known under-registration issue in the EPR [3-5], this could also be partly explained by low quality of the extraction modules.

It is also worth pointing out that patient profiles were similar, in terms of age, sex, and mean values of the various parameters, regardless of the method that was used.

4. Conclusions

Building research networks for health service management at a national scale raises issues not encountered by smaller pilot studies. Gaining the trust of the many different stakeholders is of the utmost importance and requires time and effort. Meanwhile, setting up rigorous standardised and centralised control procedures remains essential if the expected benefits are to be achieved. Currently, correctness of the extraction is still an issue but the national Belgian labelling programme of EPR software systems may help the required quality level to be achieved within the next two or three years.

Uploading data from EPR allows chronological retrospective data to be collected at a national scale. The use of manual data capture is definitely not an option.

In future, the properties of the research information network will evolve. The quality of the extraction modules will improve. More GPs will use the upload method for more patients, providing more data. There may also be improved data recording in the EPR and improved data communication between healthcare actors. We strongly suggest that such improvements of the properties of the research information network be measured and monitored, so that future changes in the quality of documented care, primarily for process parameters, may be interpreted.

This study was funded by the NIHDI.

References

[1] Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. Milbank Q 1996; 74(4):511-44

[2] de Lusignan S, Van Weel C. The use of routinely collected computer data for research in primary care: Opportunities and challenges. Fam Pract 2006; 23:253–63.

[3] Chan KS, Fowles JB, Weiner JP. Electronic Health Records and the Reliability and Validity of Quality Measures: A Review of the Literature. Med. Care Res. Rev. 2010; 67(5):503-27

[4] De Clercq E et al. Quality assessment of automatically extracted data from GPs' EPR. Stud Health Technol Inform, 2012, 180:726-30

[5] De Clercq E et al. Routinely-collected general practice data from the Electronic Patient Record and general practitioner active electronic questioning method: a comparative study. Stud Health Technol Inform, 2013; 192:510-4

[6] Pearce CM et al. The Computerized Medical Record as a Tool for Clinical Governance in Australian Primary Care. Interact J Med Res. 2013 Jul-Dec; 2(2):e26.

[7] De Clercq E et al.. Nation-Wide Primary Healthcare Research Network: a Privacy Protection Assessment. Stud Health Technol Inform 2012; 174:23-8.