Collaborative Patient Centred eHealth
E. De Clercq et al. (Eds.)
IOS Press, 2008

doi:10.3233/978-1-58603-922-6-103

Electronic Patient Record data as proxy of GPs' thoughts

Etienne DE CLERCQ[1], Viviane VAN CASTEREN[2], Pascale JONCKHEER[3], Peter BURGGRAEVE[4] and Marie-France LAFONTAINE[2]

[1] *Health Systems Research, Université Catholique de Louvain, Brussels, Belgium*
[2] *Scientific Institute of Public Health (IPH), unit of Epidemiology, Brussels, Belgium,*
[3] *Société Scientifique de Médecine Générale (SSMG), Brussels, Belgium*
[4] *Domus Medica (Flemish Institute of General practitioners), Berchem, Belgium*

Abstract. Data currently available in primary care Electronic Patient Records (EPR) can potentially be used to study quality of care. In this paper we investigate to which extend these data can reflect GPs' "thoughts" that are an important issue when considering GPs' practice and quality improvement cycle. Within the Resoprim project, we mainly used the consolidated data of three software systems, 26 practices, 1 554 hypertensive patients and 1 977 contacts. Extracted data from the EPR were: some diagnoses, some drugs, referral events, marital status, some parameters (smoking status, height, weight, blood pressure). As "gold standard" of GPs' thoughts we used an electronic questionnaire at the end of each contact. Measures of missing and incoherent values were used to assess our "gold standard". Sensitivity, positive predictive values, correctness and global completeness were used to measure the quality of the automatic extracted data (our proxy). For the "gold standard", the global percentage of missing values is 1.88% and of incoherent values is 3.92%. For most of the practices, the PPV or the correctness of automatic extracted drugs and automatic extracted parameters is high (>95%). The PPV of automatic extracted diagnoses is variable (42.1% to 94.9%). The sensitivity of automatic extracted diagnoses and drugs is lower than 67%. For most of the practices the sensitivity of automatic extracted parameters (excl. smoking status) is higher than 95%. The global completeness of height and weight is lower than 76%. Referrals are badly recorded or extracted. Currently in Belgium, without additional investigations, databases built on data extracted from EPRs can hardly be considered as good proxies of what is thought or known by the GPs. To use them as proxies, we should at least develop tools such as electronic questionnaires to calibrate them. As priority, we suggest an improvement of the extraction procedure design, of the current software interfaces and of the quality control of the extraction modules in order to improve respectively the extracted drugs sensitivity, the global completeness of extracted parameters and the PPV of extracted diagnoses. Training GPs could also be helpful.

Keywords. Medical records, Primary health care, Data collection, Computerized patient records.

[1] Dr Etienne De Clercq, HSR – ESP – UCL, Clos Chapelle aux Champs 30.41, 1200 Brussels Belgium; Email: etienne.declercq@uclouvain.be

1. Introduction

Data currently available in primary care Electronic Patient Record (EPR) can be used for various research purposes. Opportunities and challenges have been described [1, 2]. In the scope of quality of care assessment, EPR may help to appraise the technical dimension of the quality, i.e. the conformance with specification or clinical guidelines, mainly for its process and outcome aspects [3]. Many studies related to quality of care critically reported on documented performance as measured by chart extraction [4-6].

In the practice, GP's acts are influenced by many factors such as patient's will, physician's own skill and knowledge, time constraints, organizational issues, but also by what the practitioner knows (or thinks) about patient's health status and about performed actions. Therefore, GP's thoughts are an important issue when considering GP's practice and quality improvement cycle.

A long time ago, Rector stated that "information in the medical record is not about what was 'true' of the patient, but what was observed and believed by clinicians"[7]. In this paper, we investigate to which extend extraction of routinely collected EPR data in primary care can reflect GPs' "thoughts". As far as we know, this issue has not yet been treated in the literature.

2. Material and Methods

Our research was performed within the Belgian ResoPrim project framework. This project involved 26 volunteer GPs' practices which between them used three different labelled software systems (out of the 19 currently used in Belgium). Thirteen GPs were using software1, five software 2 and 8 software 3. From these practices, data were prospectively collected over 6 weeks in early 2005 around the theme "hypertension and cardiovascular risk factors". Quality control and quality assessment procedures (using a dummy patients technique) were conducted for the extraction modules developed by each software package. More details are provided elsewhere[8].

For this study, we used data related to

- some specific diagnoses: hypertension, diabetes, hypercholesterolemia and cardiovascular event (myocardial infarction, angina pectoris, coronary revascularisation, stroke, transient ischaemic attack, carotid surgery, leg claudicatio, aorto-femoral revascularisation).
- some drugs: aspirin, statin, and drugs related to hypertension;
- referral "event";
- marital status;
- some parameters: height, weight, smoking status, systolic and diastolic blood pressure.

As a proxy for GPs' "thoughts", we used an electronic questionnaire (see Table 1). At the end of each contact with a patient the GP had to answer the first 4 questions. For hypertensive patients seen at GP's office (according to the first 2 questions), the GP answered to the whole questionnaire (14 questions) and, for the 5 parameters, the GP had also to validate, complete or correct the data extracted from the EPR. These validated parameters were used as "gold parameters". To assess various ways to build questionnaires (and to improve its acceptability for the participating GPs), only three

questions were mandatory (Q1, Q2 and Q4) and Q5.0 and Q5.x were mutually exclusive. The value "unknown" was foreseen for the questions Q3, Q4 and Q6 to Q14. Using (paper) questionnaires during (after) a contact is currently a well accepted technique for instance in the Belgian Sentinel Network (cf. http://www.iph.fgov.be/epidemio/epien/index10.htm).

After each contact data were automatically extracted from the EPR. For diagnoses we extracted ICPC2 and ICD10 codes, for drugs we extracted ATC codes, for the referrals, we extracted "events". To improve the completeness of data extracted, we searched for diagnoses codes in various places in the EPR (problem list, diary, personal past history, list of healthcare elements, family past history). We also extracted indirect codes (i.e. codes deduced by the software systems according to information available in the EPR). For the drugs we extracted drugs prescribed during the contact and active drugs including active chronic treatments. We only extracted a complete set of data for hypertensive patients seen at GP's office (according to the first 2 questions). We extracted from the EPR the various parameters and the marital status. No data were currently available in the EPR for the educational level.

Table 1. Electronic questionnaire

Q1	Contact at GP's office /home visit / other?	Q8	Is it a new case of hypertension?
Q2	Hypertensive patient? *(yes / no)*	Q9	Patient known with hypercholesterolemia?
Q3	Education level: superior?	Q10	Is the patient suffering from type 2 diabetes?
Q4	Marital status: married?	Q11	Patient with personal past cardiovascular event?
Q5	Drugs currently taken for hypertension Q5.0: none Q5.1to7: beta-blockers, diuretics, calcium antagonist, ACE-inhibitors, sartanes, alpha-blockers, central working agent	Q12	Patient with family past cardiovascular event?
Q6	Does your patient take low dosed aspirin?	Q13	Patient referred to a specialist for his hypertension during this contact?
Q7	Does your patient take a statin?	Q14	Patient referred to a cardiologist for his hypertension during the year 2004?
Extracted parameters presented in a table to be validated/corrected/completed: height, weight, smoking status, systolic and diastolic blood pressure.			

Note 1: Q1, Q2 and Q4 are mandatory questions
Note 2: : Possible value for Q3, Q4, Q6 to Q14: Yes / No / Unknown
Note 3: Q5.0 and Q5.1 to 7 are mutually exclusive

For the questions, to improve our confidence in the electronic questionnaire as "gold proxy" for the GPs' thoughts, we measured two indicators for each practice: the percentage of patients with missing value (for the 11 optional questions) and the percentage of patients with incoherent values (a kind of double entry method). For questions 3, 4 and 8 to 14, incoherent values are calculated for patients with at least two contacts during the registration period and some logically incompatible answers to a same question. For example for Q10: first answer 'yes', second answer 'no'. For Q5, we were also able to calculate incoherent values for software 3 that failed to implement properly the mutually exclusive property of Q5.0 and Q5.x (during a same contact, a positive answer to Q5.0 and to Q5.x is incoherent).

To compare automatic extracted (AE) data and the "gold standard" (answers to the questions) we used for each practice, sensitivity (proportion of patients with a positive AE data out of all the patients with a positive answer to the related question) and Positive Predicted Value (PPV, i.e. the proportion of the patients with a positive answer to a question out of all the patients with a positive AE data related to that question).

For all the parameters but the smoking status, we measured the sensitivity (proportion of patients with an AE parameter out of all the patients with a validated parameter), the correctness (proportion of patients with an AE parameter that has the same value as the validated parameter out of all the patients with the AE parameter) and the global completeness of the validated parameters (i.e. the proportion of all the patients with a validated parameter).

For the smoking status, we measured the sensitivity (proportion of patients with a positive AE smoking status out of all the patients with a positive validated smoking status), the correctness (the proportion of the patients with a positive validated smoking status out of all the patients with a positive AE smoking status) and the global completeness (the proportion of all the patient with a validated smoking status).

We applied these indicators to all the hypertensive patients attending GP's office during the six weeks period. This population was identified by the answers to the first two (mandatory) questions. For questions 1 and 2, sensitivity and PPV were calculated for the whole population of patients attending GP's office.

According to the results of the quality procedures (e.g. data not properly extracted from the EPR) and sometimes to the study design (e.g. data only related to one contact), some indicators were not applicable (n.a.) to some combinations of question and software system(s).

3. Results

We mainly used the consolidated data of all three software systems, 26 practices, 1 554 hypertensive patients (out of 7 831) and 1 977 contacts (out of 10 914).

3.1. Gold proxy (see table 2)

In the questionnaire, the global percentage of missing values is 1.88 %. It ranges by question from 0.97% to 4.76% and by practice from 0% to 8.33%. For 95% of the applicable combinations "Practice-question" (out of 265), the percentage of missing values is less than 5%. For question Q3, Q5 and Q14, respectively 2, 8 and 3 practices have a missing percentage higher than 5%. Six practices have one question with a high percentage of missing value (in-between 10% and 75%).

Table 2. Missing and incoherent values in the questionnaire ("gold standard")

	% of missing values				% of incoherent values			
Questions	Soft 1 (730 pat.)	Soft 2 (354 pat.)	Soft 3 (470 pat.)	**Total (1554 pat.)**	Soft 1 (244 pat.)	Soft 2 (178 pat.)	Soft 3 (207 pat.)	**Total (629 pat.)**
Q3	n.a.	0.00%	12.55%	**3.80%*** **	9.84%	3.37%	2.42%	**5.56%**
Q4	n.a.	n.a.	n.a.	**n.a.**	6.56%	3.37%	6.76%	**5.72%**
Q5	4.25%	3.39%	1.91%	**3.35%**	n.a.	n.a.	8.72%*	**n.a.**
Q6	2.19%	0.56%	0.43%	**1.29%**	n.a.	n.a.	n.a.	**n.a.**
Q7	1.78%	0.85%	0.21%	**1.09%**	n.a.	n.a.	n.a.	**n.a.**
Q8	1.78%	0.56%	0.43%	**1.09%**	2.46%	1.12%	0.97%	**1.59%**
Q9	1.78%	0.28%	0.21%	**0.97%**	6.15%	5.62%	2.90%	**4.93%**
Q10	1.64%	0.85%	0.64%	**1.16%**	1.64%	1.12%	0.00%	**0.95%**
Q11	1.64%	0.56%	n.a.	**1.54%****	2.87%	2.25%	n.a.	**1.75%*****
Q12	1.78%	1.13%	0.43%	**1.22%**	6.56%	2.81%	2.90%	**4.29%**
Q13.	1.78%	0.56%	0.43%	**1.09%**	n.a.	n.a.	n.a.	**n.a.**
Q14	9.59%	0.56%	0.43%	**4.76%**	9.59%	6.74%	5.31%	**5.25%**

n.a. = not applicable for various design of technical reasons.
* denominator = 470 patients; ** denominator = 1084 patients; *** denominator = 824 patients;**** denominator = 422 patients

The global percentage of incoherent values (excluding Q5) is 3.92%, ranging by question from 0.95% to 5.72% and by practice from 0% to 10.42% (1 practice with only 3 patients excluded). The higher incoherent percentage for one combination "Practice-question" is 29.17% (1 practice excluded). For 77% of the applicable combinations "Practice-question" (out of 192, 1 practice excluded), the percentage of incoherent values is less than 5%. Twenty "Practice-questions" have a high percentage of incoherent values (in-between 10% and 29.17%).

For Q5, most of the incoherences (39/41) are related to only one practice (out of 8).

3.2. Sensitivity and PPV of automatic extracted data (see table 3)

For the drugs (Q5 – Q7) the PPV is ranging from 91.5% to 94.1% and the sensitivity from 34.1 to 59.34%.

For the diagnoses (Q2, Q9 and Q10), the PPV is ranging from 42.1% to 94.90% and the sensitivity from 53.6% to 67.1%.

For the referrals (Q13 and Q14), the PPV and the sensitivity are low (<37%)

Table 3. Sensitivity and Positive Predictive Value of Automatic Extraction vs questionnaire

	Soft 1 (3.367 pat.)			Soft 2 (1.326 pat.)			Soft 3 (2.741 pat.)			Total (7.434 pat.)		
	N**	**Sens.**	**PPV**	**N****	**Sens.**	**PPV**	**N****	**Sens.**	**PPV**	**N****	**Sens.**	**PPV**
Q2	869	69.30%	96.20%	516	38.80%	92.60%	595	43.70%	93.90%	1980	53.60%	94.90%
Q3	991	n.a.	n.a.	235	n.a.	n.a.	471	n.a.	n.a.	1697	n.a.	n.a.
Q4	1746	27.70%	92.40%	814	46.20%	93.30%	1354	18.30%	91.90%	3914	28.30%	92.60%
	Soft 1 (730 pat.)			**Soft 2 (354 pat.)**			**Soft 3 (470 pat.)**			**Total (1554 pat.)**		
	N**	**Sens.**	**PPV**	**N****	**Sens.**	**PPV**	**N****	**Sens.**	**PPV**	**N****	**Sens.**	**PPV**
Q5 (1 to 7)	671	76.30%	96.79%	325	38.46%	98.43%	396	47.73%	80.43%	1392	59.34%	92.70%
Q6	259	49.40%	93.40%	130	20.00%	96.30%	174	21.80%	95.00%	563	34.10%	94.10%
Q7	183	54.60%	93.50%	120	25.00%	100.00%	143	37.10%	84.10%	446	41.00%	91.50%
Q8	30	43.30%	14.40%	27	18.50%	33.30%	44	38.60%	17.00%	101	34.70%	17.10%
Q9	235	85.50%	39.90%	177	13.60%	77.40%	173	n.a.	n.a.	412	54.61*%	42.1*%
Q10	105	89.50%	59.50%	51	47.10%	68.60%	84	51.20%	89.60%	240	67.10%	66.80%
Q11	142	12.70%	75.00%	59	32.20%	52.80%	n.a.	n.a.	n.a.	201	18.4*%	61.7*%
Q12	101	n.a.	n.a.	42	n.a.	n.a.	64	n.a.	n.a.	207	n.a.	n.a.
Q13.	130	7.70%	31.30%	12	8.30%	4.80%	36	22.20%	38.10%	178	10.70%	25.70%
Q14	180	17.80%	33.30%	128	15.60%	43.50%	155	n.a.	n.a.	308	16.88*%	36.60*%

*: related to 1084 patients
N** number of patients with a positive answer to the question
Sensitivity and PPV were calculated for the automatic extraction procedure
n.a.: not applicable for various technical reasons

3.3. Quality of automatic extracted (AE) and validated parameters (see table 4)

In the software systems, 'never' was used as default value for the smoking status. It was thus impossible to calculate the completeness. The percentage of smokers and past smokers are respectively 11.5% and 11.3% (soft. 1 and 3). For all the other parameters, the completeness of the validated parameters varies between 68% and 97.30%.

For all the five parameters, the correctness is very high (from 97.87% to 99.80%).

The sensitivity for the smoking status (value 'smokes') is 66.67% (78.85% for the value 'stopped smoking' for the software 1). For the other parameters, the sensitivity ranges from 77.20% to 92.30%. 58% of the practices (15/26) have sensitivity higher than 95% for all the parameters, but the smoking status.

Table 4. Quality of automatic extracted and validated parameters

Validated parameters		Soft 1 (730 pat.)	Soft 2 (354 pat.)	Soft 3 (470 pat.)	Total (1554 pat.)
Smoking status	completeness	n.a.	n.a.	n.a.	n.a.
	Sensitivity (AE)	91.50%	n.a.	13.60%	66.67%*
	Correctness (AE)	97.70%	n.a.	100.00%	97.87%*
Height	completeness	57.90%	79.90%	74.50%	68.00%
	Sensitivity (AE)	97.60%	55.10%	70.30%	77.20%
	Correctness (AE)	100.00%	99.40%	99.60%	99.80%
Weight	completeness	68.10%	81.10%	83.40%	75.70%
	Sensitivity (AE)	98.00%	61.70%	78.80%	82.70%
	Correctness (AE)	99.60%	97.30%	98.40%	98.80%
Systolic blood pressure	completeness	98.50%	92.70%	98.90%	97.30%
	Sensitivity (AE)	98.30%	68.90%	99.60%	92.30%
	Correctness (AE)	98.30%	97.40%	99.80%	98.70%
Diastolic blood pressure	completeness	98.20%	92.70%	99.40%	97.30%
	Sensitivity (AE)	98.50%	68.60%	99.40%	92.30%
	Correctness (AE)	98.50%	98.30%	99.80%	98.90%

Global completeness is measured for the validated parameters
Correctness and sensibility for smoking status are calculated for "smokers"
n.a.: not applicable for technical reasons
* related to 1200 patients

4. Discussion

For the questions, the percentage of missing values for simple questions (answer: yes/no/unknown) is very low. Therefore more complex question design, such as Q5 with mutually exclusive parts, or mandatory questions (Q1, Q2 and Q4 were mandatory with a strong technological control, all the other questions were optional) does not seem very useful or required (see table 2).

Missing and incoherent values (for Q3 and Q6 to Q14) may be considered as a partial measure of the gap between what is thought by the GP and what is recorded in our "gold standard" (unknown is a possible value). For these questions, incoherent values were only measurable for patients attending GPs' office at least twice (40% of the hypertensive patients) Given the quality of our "gold standard", we think that a PPV and a sensitivity of 90% for the automatic extraction (AE) is quite acceptable. To assess this quality, alternative methods such as think aloud techniques could have been considered but they can hardly be implemented in a GP's running environment.

For drugs we found a high PPV but an unexpectedly [9, 10] low sensitivity. Even if we took into account all the active drugs as registered in the EPR (plus all the drugs prescribed during the pilot phase), the sensitivity was not higher than the diagnoses sensitivity. This is valid for drugs with (Q5, Q7) or without (Q6) compulsory prescription. Means to improve it could be: to take into account all the drugs prescribed during a longer period, e.g. the 3 last month, to improve the 'active drugs' management by the software systems, to stimulate GPs to use their prescription software modules.

For the diagnoses the sensitivity is not very high, even if we took into account all the rubrics of the various EPR systems (problem list, past medical history, diary, etc.), including codes derived by the software systems. For each of the three software systems, diagnoses codes are effectively spread over several rubrics. More challenging is the PPV, which is variable by practice and by code. This hampers the ability to identify homogeneous groups of patients. This could be related to the quality of the software extraction modules (confusion between personal past event and family history, calculation of the derived codes by the software systems) or to the ways questions are presented to the GPs (Q2 = mandatory, Q9 and 10 = optional and included in a broader list of questions). In the latter case, it raises the issue of using questionnaires as "gold standard" for diagnoses. This will be investigated later using data of the second phase of the Resoprim project.

For the referrals, sensitivity and PPV are low. This could be progressively improved by stimulating the exchange of computerized referrals thanks to the development of secured messaging systems, electronic signatures and unique patient identification systems.

Smoking status is badly encoded and extracted. For all the other 4 parameters, the correctness and, for most of the practices, the sensitivity are rather good. Therefore we can wonder if the semi-automatic data extraction procedure (i.e.: secondary validation by the GP after extraction) is very useful. Only for six practices (out of 26), the number of patients with documented parameters was improved by more than 10%. To increase the sensitivity, an alternative strategy could be applied: first to improve the software interfaces and second, to persuade GPs to write down in their EPR what they have measured. To improve the completeness of some parameters (height, weight) is still challenging. It requires a change in GPs' behaviour to measure them.

5. Conclusions

Currently in Belgium and without additional investigations, databases built on data extracted from EPRs can hardly be considered as good proxies of what is thought or known by the GPs. Therefore, to make them usable in well defined contexts [11], we strongly advise to develop tools, such as an electronic questionnaire, in order to calibrate these proxies. As technical priorities, we suggest an improvement of the extraction procedure design, of the current software interfaces and of the quality control of the extraction modules in order to improve respectively the extracted drugs sensitivity, the completeness of extracted parameters and the PPV of extracted diagnoses. Training GPs could also be helpful to improve the quality of data recording.

Acknowledgments

We thank all the GPs and the industrial partners involved in the data collection. This work benefits from grants of the Belgian Federal Public Planning Service Science Policy (Nr I2/AE/207 and Nr I2/2F/207).

References

[1] de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. Family Practice 2006; 23:253-63.
[2] De Clercq E, Van Casteren V, Jonckheer P, Burggraeve P et al. Research networks: can we use data from GPs' Electronic health Records. Stud.Health Technol. Inform. 2006; 124:181-6.
[3] Campbell SM, Roland MO, Buetow SA. Defining quality of care. Soc Sci Med 2000; 51(11):1611-25
[4] Dresselhaus TR, Peabody JW, et al.. Measuring compliance with preventive care guidelines: standardized patients, clinical vignettes, and the medical record. J Gen Intern Med 2000; 12(11):782-8.
[5] Kerr EA, Smith DM, Hogan MM, Krein SL, et al.. Comparing clinical automated medical record, and hybrid data sources for diabetes quality measures. Jt Comm J Qual Improv 2002; 28(10):555-65.
[6] Kirk SA, Campbell SM, et al. Assessing the quality of care of multiple conditions in general practice: practical and methodological problems. Qual Saf Health Care 2003; 12(6):421-7
[7] Rector AL et al. Foundations for an electronic medical record. Meth Inf Med 1991;30:179-86.
[8] De Clercq E, Van Casteren V, et al. Resoprim – Are GPs' Electronic Health Records suitable for use in Public health Research?, Scientific Institute of Public Health ed., Brussels, January 2008, 64 pp. Report D/2008/2505/03. http://www.iph.fgov.be/epidemio/epien/re_pr_en/D_2008_2505_03.pdf
[9] Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *BMJ* 2003;326:1070.
[10] Luck J, Peabody JW, Dresselhaus TR, Lee M, Glassman P. How Well Does Chart Abstraction measure Quality? A Prospective Comparison of Standardized Patients with the Medical Record. The American Journal of Medicine 2000; 108:642-49.
[11] Nicole Boffin, Van Casteren V, et al. P. Elektronische dossiergegevens en kwaliteit van zorg: ervaring uit Resoprim over hypertensiepatiënten. Huisarts Nu. 2008;37(3):131-38.