MEDINFO 2013
C.U. Lehmann et al. (Eds.)

doi:10.3233/978-1-61499-289-9-510

Routinely-collected General Practice Data from the Electronic Patient Record and General Practitioner Active Electronic Questioning Method: A Comparative Study

Etienne De Clercq[a], Sarah Moreels[b], Nathalie Bossuyt[b], Katrien Vanthomme[b], Geert Goderis[c] , Viviane Van Casteren[b]

[a] *Institut de Recherche Santé et Société (IRSS), Université Catholique de Louvain, Brussels, Belgium*
[b] *Scientific Institute of Public Health (IPH), Unit of Epidemiology, Brussels, Belgium.*
[c] *Academisch Centrum voor Huisartsgeneeskunde, KU Leuven, Belgium*

Abstract

The numerous existing primary care-based research networks currently use various data collection methods. In this paper, we compared routine data extracted from general practitioners' (GPs') electronic patient records (EPRs) and GPs' answers to an electronic questionnaire.

We investigated for 10,307 Belgian patients 10 healthcare conditions using clinical and biological parameters (cholesterol, blood pressure, and body mass index), diagnoses (hypertension, diabetes, and personal past cardiovascular event(s)), and drug prescriptions (antidiabetic drugs, aspirin, statins, and antihypertensive drugs).

We found a relatively fair agreement (Kappa ≥ 0.40) between the two data collection methods for 7 healthcare conditions, but no agreement for the biological parameters. When EPR data was used and compared with the questioning method, the prevalence of diagnoses and drug prescriptions was relatively lower and the prevalence of clinical and biological parameters was relatively higher (all missing data excluded) in the EPR data than in the data collected using the questioning method. Using EPR data, we calculated an acceptable proxy for the prevalence as observed using the questioning method.

The comparison of the two data collection methods was a worthwhile approach, in that it could highlight potential ways to improve both care quality and information systems.

Keywords:

Computerized patient record, primary health care, data collection.

Introduction

Many benefits are expected from the secondary use of routine primary care data, such as auditing and improving care quality, health service planning, and epidemiological research. The opportunities and challenges presented by these secondary uses have already been described [1]. For many years, studies have been conducted to assess the quality of data extracted from general practitioners' (GPs') Electronic Patient Records (EPRs) [2-5]. Currently, drawing valid conclusions from this type of data remains difficult [2, 6].

Primary care-based research networks have been set up and described in many countries. Various data collection methods are used: standardized paper forms, standardized Internet forms, and standardized extraction from the EPR [1, 7].

In Belgium, several primary care research networks are currently running. These use different data collection methods, including standardized paper forms (Belgian network of sentinel general practitioners [8]) and standardized extraction from the EPR (Intego network [9]). Moreover, some of these networks are considering moving from paper-based data collection towards electronic data collection methods (which would partly include data extraction from routine primary care EPRs) [8].

In the meantime, the Belgian National Institute for Health and Disability Insurance (NIHDI) has set up a large national health research network called ACHIL (Ambulatory Care Health Information Laboratory), involving nearly all 10,000 practising Belgian GPs and the 17 different software systems they use. In its first phase (2009-2013), ACHIL focused on assessing the quality of care for some patients with type 2 diabetes and/or chronic renal failure. Only a small amount of structured data was collected in 2012 (no free text). Two methods of data collection with acceptable data privacy protection were developed at the national level: standardized Internet forms and standardized data extraction from the EPR [10].

Using data extracted from EPR to draw conclusions, identify patients, or define the proportion of people undergoing specific treatments or suffering from specific chronic conditions (such as diabetes, hypertension, chronic renal failure, hypercholesterolemia, being overweight) is still challenging [1, 6, 11-15]. For many years it has been accepted that a disparity could exist between clinical notes and the actual care provided [16].

In this context of various data collection methods that still include controversial EPR data extraction methods and numerous software systems, our study investigated whether, at the public health level (aggregated data), patients' healthcare conditions as perceived by GPs (from active questioning methods) can be compared with or deduced from research databases built using routinely collected EPR data.

Methods

This study used data from the ResoPrim project (phase 2: 2006-2008), a Belgian primary care research network involving 43 volunteer GPs, 10,307 patients, and 4 EPR systems. To build the ResoPrim research database, coded and structured

data was searched for in various places within the EPR and automatically extracted at the end of each contact (Summer 2007, mean extraction period: 3.8 weeks). GPs were actively questioned at the end of each contact (electronic questionnaire) about various healthcare conditions of patients, with questions such as: "Is your patient suffering from type 2 diabetes?", "Does the patient take a statin?", or "Is your patient overweight (Body Mass Index > 25)?" (See Table 1). A more detailed description of the ResoPrim research network has already been published [17].

We used three types of data to investigate 10 healthcare conditions:

- Clinical and biological parameters: blood pressure (BP), body mass index (BMI), and cholesterol;
- Diagnoses: hypertension, diabetes, and personal past cardiovascular event(s);
- Drug prescriptions: antidiabetic drugs, aspirin, statins, antihypertensive drugs.

For clinical and biological extracted parameters, we used the most recent values extracted and the most recent values that were not more than 4 months old. For diagnoses, we extracted ICPC2 codes, ICD10 codes, and codes from the national Belgian Thesaurus[1], whatever their entry dates to the EPR. For current drug treatments, we extracted national codes mapped with ATC for the drugs prescribed during the current contact or previously recorded in an "active medication list".

Table 1 – Healthcare conditions

Patient known to have a history of hypercholesterolemia? (total cholesterol > 190 mg/dl and/or LDL cholesterol > 115 mg/dl)
Is the patient's blood pressure currently higher than 140/90?
Is the patient overweight (BMI > 25)?
Is the patient suffering from hypertension?
Is the patient suffering from type 2 diabetes?
Does the patient have a history of cardiovascular event(s)?
Does the patient currently take any antihypertensive drugs?
Does the patient currently take any antidiabetic drugs?
Does the patient currently take low-dose aspirin?
Does the patient currently take a statin?

To investigate the level of agreement between the two methods (automatic extraction and active questioning), we built 2x2 tables and calculated the Kappa statistic (missing data excluded). In Belgium, however, it is currently impossible to differentiate between missing and negative values for extracted diagnoses and drug prescriptions. There are no "negative" codes in the EPR to indicate that a patient is not suffering from a disease or not taking a drug. We therefore also calculated the Kappa statistic treating all missing extracted data values as negative for the clinical and biological parameters, in order to be able to apply one standard method to all 10 healthcare conditions included in the study.

We calculated the prevalence, as observed by the GPs (missing data excluded), for the 10 healthcare conditions, using the answers to the electronic questionnaire (further called the Q Prevalence). We also calculated the observed prevalence using the AE (automatically extracted) data, once excluding missing AE values and once treating all missing AE values as negative.

In order to deduce the Q Prevalence from the AE data, we used and assessed the Equation (1):

$$Estimated\_Q\_\Pr ev. \equiv AE\_\Pr ev. \times {}^{PPV}\!/_{Sensitivity} \quad (1)$$

where Estimated_Q_Prev. is a proxy for the Q prevalence; AE_Prev. is the prevalence observed using AE data (tested with and without missing AE data); PPV is the positive predictive value of the AE data (i.e. the proportion of patients who gave a positive answer to the relevant question, of those with positive AE data), and Sensitivity is defined as the proportion of patients with positive AE data, of those who gave a positive answer to the relevant question.

Results

For the diagnoses, we obtained usable data for all 10,307 patients that attended the GPs' offices during the data collection period. For drugs and for clinical and biological parameters, we obtained data from 3261 patients (all of the patients with hypertension, diabetes, or personal past cardiovascular events, of the 10,307 patients attending GPs' offices).

Table 2 – Clinical and biological parameters

3261 patients		**Relevant questions**		
Extracted Parameters		+	-	missing
Cholesterol	+	673	897	140
	-	466	369	79
	missing	210	350	77
Cholesterol < 4 month	+	316	303	51
	-	218	165	47
	missing	815	1148	198
Blood Pressure	+	713	737	36
	-	116	1422	38
	missing	57	123	19
Blood Pressure < 4 month	+	674	590	26
	-	84	1222	25
	missing	128	470	42
BMI	+	1120	398	32
	-	102	393	12
	missing	571	596	37
BMI < 4 month	+	569	137	15
	-	45	142	6
	missing	1179	1108	60
BMI: Body Mass Index				

Raw data for the clinical and biological parameters is shown in Table 2. Answers to the questions regarding these parameters were missing in 4.8% of cases. This percentage rose to 5.7% for all 10 healthcare conditions. For the AE data (most recent

1 National Thesaurus for diagnoses and symptoms funded by the Belgian Ministry of Public Health and mapped with ICPC2 and ICD-10.

values), the percentages of missing data were: 19.5% for cholesterol, 6.1% for blood pressure, and 36.9% for BMI. For AE data no more than 4 months old, these percentages rose to 66.3%, 19.6% and 72.0%, respectively.

The Kappa statistics and the various prevalences for the 10 healthcare conditions are shown in Table 3.

Discussion

As shown in Table 3, we found a relatively fair agreement (Kappa between 0.42 and 0.48) between the two methods (AE and active questioning) for clinical parameters (blood pressure and BMI). No agreement was found for biological parameters (cholesterol, Kappa ≈ 0). As shown in Table 2, there were large numbers of false negative (466) and false positive (897) values for cholesterol, which explains this lack of agreement.

The number of patients taking statins (figures not shown in Table 2) could explain these false negative values. Of the 466 patients, 410 (88%) were taking statins. These 410 patients could have been suffering from hypercholesterolemia (as mentioned by the GPs). If treated with statins, their cholesterol levels could have been normalized. However, we cannot consider all of the patients being treated with statins and with normal cholesterol to be patients suffering from hypercholesterolemia. Indeed, 50 patients being treated with statins had normal cholesterol (AE data) and were not considered to be suffering from hypercholesterolemia (negative answer to the question). These 50 patients (out of 369 true negative patients) could be receiving preventive treatment with statins to reduce their cardiovascular risk. Receiving preventive treatment was not surprising given that all 3261 patients had a moderate or high cardiovascular risk.

Explaining the high level of false positive values for cholesterol (897, i.e. 27.5% of all the patients) is more difficult. This figure decreased to 303 (9.3% of all the patients) when recently extracted data no more than 4 months old was used. This suggests that GPs were possibly not aware (negative answer to the question) that these 303 patients could have been suffering from hypercholesterolemia (positive AE data). This highlights a potential way to improve quality of care (these patients could perhaps benefit from a prescription for a statin).

We also observed a large number of false positive values (737) for the blood pressure parameter. Even when only the most recent blood pressure values (< 4 months) were used, we still found 18% of patients (590) with high blood pressure values that were considered "normal" by the GPs. This merits further investigation.

We noticed that many patients (N=1179) considered to be overweight by the GPs (N=1793) had no recent (< 4 month) BMI measurements (at least no data extracted from the EPR) and 60.4% (N=815) of the patients with hypercholesterolemia (based on answers to the questions) had no recent cholesterol value (< 4 months) extracted from the EPR. Most of these patients already had previous extracted data for BMI or cholesterol. This suggests that their care could be improved.

As expected, using more recent values (< 4 months) slightly increased the agreement between the methods when missing AE data was excluded (see Table 3) but also (drastically) increased the numbers of missing AE data values (more than 2/3 of the AE data was missing for cholesterol and BMI, see Table 2).

Table 3 – Observed and estimated prevalence

		Kappa (AE vs. Questions)		**Observed AE Prevalence**		**Q Prevalence**	**Estimated Q Prevalence**	
Healthcare conditions	**Extracted data**	Missing excluded	Missing AE = "-"	Missing excluded	Missing AE = "-"	Missing excluded	Missing excluded	Missing AE = "-"
Hypercholesterolemia	Cholesterol	-0.12	-0.06	65.17%	52.4%	45.5%	47.3%	45.1%
	Cholest. < 4 months	-0.06	0.05	60.91%	20.5%	45.5%	52.5%	44.8%
Blood Pressure > 140/90	Blood Pressure	0.42	0.40	48.5%	45.6%	28.0%	27.7%	27.8%
	BP < 4 months	0.47	0.44	49.2%	39.6%	28.0%	29.5%	27.7%
Overweight (BMI > 25)	BMI	0.44	0.33	75.35%	47.5%	56.4%	60.7%	56.1%
	BMI < 4 months	0.48	0.20	78.88%	22.1%	56.4%	68.6%	56.2%
Hypertension	HT diag. code	N/A	0.47	N/A	17.4%	30.6%	N/A	31.5%
Diabetes	Diab. Diag. code	N/A	0.55	N/A	5.3%	7.5%	N/A	8.0%
PCVE	PCVE diag. code	N/A	0.36	N/A	4.7%	8.6%	N/A	9.1%
HT Drugs	HT drug code	N/A	0.24	N/A	68.1%	91.9%	N/A	90.5%
Diab. Drugs	Diab. Drug code	N/A	0.75	N/A	14.0%	26.7%	N/A	19.5%
Aspirin	Aspirin drug code	N/A	0.44	N/A	20.6%	42.8%	N/A	42.6%
Statin	Statin drug code	N/A	0.54	N/A	22.4%	38.3%	N/A	38.3%

Q Prevalence: prevalence observed by using the answers to the Questionnaire (missing answers excluded); PCVE: personal history of cardio-vascular event, N/A: not applicable

Treating missing data as negative decreased the agreement between the methods (see Table 3). This decrease was greater for most recent parameters (< 4 months) with more missing AE data. For most recent BMI (< 4 months) data, the agreement decreased to 0.20 when the missing AE data values were considered to be negative values. This is related to the high percentage of missing AE data (72%) but also to their spreading in the table (see Table 2).

A relatively fair agreement was found for hypertension (0.47), diabetes (0.55), aspirin (0.44), and statins (0.54). The good agreement (0.75) found for the diabetic drug prescriptions could be explained by the very high level of true negatives (1652 patients out of 3261, not shown in the results). Most of the patients in the group are not diabetic. When we restricted the analysis to the diabetic patients, the agreement between the methods decreased to 0.37 (not shown in Table 3). The slight agreement observed for the antihypertensive drug prescriptions (0.24) could be related to the high level of false negative values (768 patients out of 3261, not shown in the results). However, it must be remembered that negative and missing AE drug values cannot be differentiated. Therefore, this low level of agreement could be explained by missing extracted drug codes. This suggests that improving the EPR data extraction modules may be necessary.

We observed (see Table 3) large variations between the prevalence observed using AE data (AE Prevalence) and the prevalence observed using the answers to the electronic questionnaire (Q Prevalence). AE Prevalence (missing data excluded) for all 3 parameters was higher than Q prevalence. Using only the most recent values (< 4 months) had an unpredictable effect on the AE Prevalence (missing AE data excluded), which increased for BMI and decreased for cholesterol.

For the parameters, AE Prevalence (missing AE = "-") was lower than AE Prevalence (missing AE data excluded), but AE Prevalence (missing AE = "-") was higher (blood pressure) or lower (cholesterol > 4 months) than the Q Prevalence. Treating missing AE data as negative AE data also had a variable effect on the AE Prevalence, which decreased closer to the Q Prevalence in the case of blood pressure, but did not in the case of cholesterol (< 4 months).

For diagnoses and drug prescriptions, the AE Prevalence was always lower than the Q Prevalence. This was not unexpected, based on the current literature [2, 3].

As expected, the best estimated Q prevalence (as proxy for Q prevalence) was obtained when missing AE data were treated as negative AE values (see Table 3). The slight variations between estimated Q prevalence and Q Prevalence can be explained by the number of missing answers to the questions and their spreading in the 2X2 table (see Table 1). An unexpectedly large variation was observed for the antidiabetic drug prescriptions: estimated Q prevalence was 19.5% and Q Prevalence was 26.7%. This can be explained by the large number of missing answers in the questionnaire (30.2%) for that question (not shown in the results). Once a GP had mentioned that a patient was not suffering from diabetes he/she often neglected to mention that the patient was not taking any antidiabetic drugs. Restricting the analysis to diabetic patients (713) drastically reduced the percentage of missing answers (4.6%), which resolved the discrepancy between estimated Q prevalence and Q Prevalence (not shown in the results).

Despite our efforts, the ResoPrim GP sample is not representative. We also observed large variations between GPs and between software systems. Our results cannot, therefore, be generalized. Further studies are required, involving more patients by practice, more participating GPs, and more data types. Large variations between practices and between systems were also reported in the literature [2].

For most of the healthcare conditions (7), we obtained a relatively fair agreement (≥ 0.40). For the biological parameter (cholesterol) we obtained no agreement (Kappa ≈ 0). As was once stated by Rector [18] "information in the medical record is not about what was 'true' of the patient, but what was observed and believed by clinicians". In Belgium, most EPR are managed by GPs themselves, which means that most of the clinical parameters, diagnoses, and drug prescriptions are encoded in the EPR by the GPs. Therefore, reaching a fair agreement between the two methods was not unexpected. However, most of the laboratory results are automatically integrated into the patient record after global acceptance by the GP (through electronic data transfers using secure medical messaging systems). This could perhaps partly explain the unexpectedly low agreement we found for the biological parameter. GPs would perhaps not have been fully aware of all the laboratory results included in the patient records, at least for frequently ordered laboratory tests such as cholesterol tests. This suggests potential improvements to care quality and information systems, such as clinical reminder systems based on laboratory results that could be included in the GPs' EPR systems. Our findings need to be confirmed by investigating other biological parameters and including laboratory tests that are carried out less frequently.

Conclusion

We found a relatively fair agreement between data extracted from GP's EPRs and GPs' answers to the electronic questionnaire for most of the healthcare conditions (7 out of 10), including diagnoses, drug prescriptions, and clinical parameters. There was no agreement at all for the biological parameter (cholesterol).

The comparison of the two data collection methods was a worthwhile approach in that it could highlight potential ways to improve both care quality and information systems. These could include improving preventive and curative treatments using statins, improving the treatment of hypertensive patients, improving the follow-up processes related to weight or cholesterol, and improving drug prescription extraction modules.

When comparing the AE Prevalence and the Q Prevalence, we found that diagnoses and drug prescriptions related AE Prevalences were lower and that parameters related AE Prevalences (missing AE data excluded) were higher. For the parameters, using missing AE data in the denominator had an unpredictable effect (AE Prevalence lower or higher than Q Prevalence).

Using the properties of the information system (PPV and Sensitivity of the AE data) and the AE Prevalence (missing AE data included), we calculated an acceptable proxy for the Q Prevalence (prevalence as observed by the GPs through an active electronic questioning method).

The fact that the GP sample was not representative and that data was not comparable between practices and software systems means that our findings could not be generalized. Further research is required to strengthen our results.

Acknowledgments

We would like to express our gratitude to all of the participating general practitioners, for their deep involvement in the

ResoPrim project, and to all of the researchers and collaborators who have supported and contributed to this work since 2003.

This research was funded by the Belgian Federal Public Planning Service Science Policy (grants No. I2/AE/207 and I2/2F/207) and by the Belgian National Institute for Health and Disability Insurance.

References

[1] 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.

[2] 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.

[3] Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *BMJ* 2003: 326 (7398):1070.

[4] Jordan K, Porcheret M, Croft P. Quality of morbidity coding in general practice computerized medical records: a systematic review. *Fam Pract* 2004: 21: 396-412.

[5] Häyrinen K, Saranto K, Nykänen P. Definition, structure, content use and impacts of electronic health records: A review of the research literature. *Int J Med Inf* 2008: 77: 291-304.

[6] de Lusignan S, Hague N, van Vlyumen J, Kumarapeli P. Routinely-collected general practice data are complex, but with systematic processing can be used for quality improvement and research. *Inform Prim Care* 2006: 14: 59-66.

[7] Deckers JG, Paget WJ, Schellevis FG, Fleming DM. European primary care surveillance networks: their structure and operation. *Fam Pract* 2006: 23: 151-8

[8] Boffin N, Bossuyt N, Vanthomme K, Van Casteren V. Readiness of the Belgian network of sentinel general practitioners to deliver electronic health record data for surveillance purposes: results of survey study. *BMC Fam Pract* 2010: 11:50.

[9] Bartholomeeusen S, Chang-Yeon K, Faes C, Mertens R, Buntinx F. The denominator in general practice, a new approach from Intego database. *Fam Pract* 2005: 22: 442-7.

[10] De Clercq E, Van Casteren V, Bossuyt N, Moreels S, Goderis G, Bartholomeeusen S, Bonte P, Bangels M. Nation-Wide Primary Healthcare Research Network: a Privacy Protection Assessment. *Stud Health Technol Inform* 2012: 174: 23-8.

[11] Anandarajah S, Tai T, de Lusignan S, Stevens P, O'Donoghue D, Walker M, Hilton S. The validity of searching routinely collected general practice computer data to identify patients with chronic kidney disease (CKD): a manual review of 500 medical records. *Nephrol Dial Transplant* 2005: 20: 2089-96.

[12] Wong MCS, Jiang JY, Tang JL, Lam A, Fung H, Mercer SW. Health services research in the public healthcare system in Hong Kong: An analysis of over 1 million antihypertensive prescriptions between 2004-2007 as an example of the potential and pitfalls of using routinely collected electronic patient data. *BMC Health Serv Res* 2008: 8: 138.

[13] Vandenberghe HEE, Van Casteren V, Jonckheer P, Bastiaens H, Van der Heyden J, Lafontaine MF, De Clercq E. Collecting information on the quality of prescribing in primary care using semi-automatic data extraction from GPs' electronic medical records. *Int J Med Inf* 2005: 74(5): 367-76.

[14] Johansen MA, Scholl J, Hasvold P, Ellingsen G, Bellika JG. "Garbage In, Garbage Out" – Extracting Disease Surveillance Data from EPR Systems in Primary Care. In: Proceedings of the ACM 2008 conference on CSCW. San Diego: 2008; pp. 525-34.

[15] Chisholm CD, Weaver CS, Whenmouth LF, Giles B, Brizendine EJ. A Comparison of Observed Versus Documented Physician Assessment and Treatment of Pain: The Physician Record Does Not Reflect the Reality. *Ann Emerg Med* 2008: 52(4): 383-9.

[16] Rethans JJ, Martin E, Metsemakers J. To what extent do clinical notes by general practitioners reflect actual medical performance? A study using simulated patients. *Br J Gen Pract* 1994: 44: 153-6.

[17] De Clercq E, Moreels S, Van Casteren V, Bossuyt N, Goderis G, Bartholomeeusen S. Quality assessment of automatically extracted data from GPs' EPR. *Stud Health Technol Inform* 2012: 180: 726-30

[18] Rector AL, Nowlan WA, Kay S. Foundations for an Electronic Medical Record. *Methods Inf Med* 1991: 30: 179-86.

Address for correspondence

Prof. Etienne De Clercq
Université Catholique de Louvain, Institut de Recherche Santé et Société (IRSS), Clos Chapelle aux Champs 30, bte B1.30.13, 1200 Brussels, Belgium.
etienne.declercq@uclouvain.be