Reuse and Sharing of Electronic Health Record Data

Sollie, J.W.

2017

document version
Publisher's PDF, also known as Version of record

Summarizing Discussion

Reflection

When we, as patients, visit our General Practitioner (GP) for a persistent cough, when we are worried about a lump we seem to feel in our breast, or about a little blood in our stool, we assume our GP will record this. We expect he or she will also register measurements and findings as well as a hypothesis or diagnosis for the cause of our complaints and a plan to reach a diagnosis and treatment somewhere in the Electronic Health Record (EHR). We also assume the GP will know about our worries when we visit her the next time. We are surprised when we visit an out-of-ours clinic or emergency department in the local hospital during the weekend and the doctors do Not seem to have access to our EHR or don't know anything about our medical history. We are annoyed when we have to repeat the serious diagnoses we have had in the past to every next doctor we meet. Diagnoses such as asthma, myocardial infarction or even cancer should be available to the locum or the surgeon when our sprained ankle turns out to be broken.

From a patient perspective, despite the anxieties about privacy issues that are often expressed, electronic health record (EHR) data reuse and sharing for purposes of care is rational and desirable. In fact nowadays patients do expect high quality data recording and sharing between doctors. Most patients do not object and are even happy to contribute when researchers want to re-use their data (even genomic data) for research, as we know from studies in the field of rare diseases[1]. Many patients also welcome early detection of (genetic) risk at serious disease as is illustrated by studies but also by the popularity of new e-health tools such as www.yourdiseaserisk.wustl.edu or www.testuwrisico.nl We do not know how patients feel about the fact that their EHR records are being used to assess quality of care but we know patients do expect high quality healthcare[2].

As we have shown in the introduction (chapter 1) data reuse and sharing is highly desirable not only from a patient's perspective but also from the perspective of the researcher, the quality assessor and the GP. We know that reuse and sharing of EHR data is already becoming commonplace despite serious concerns about data-quality and subsequent reusability. We felt there was a need to quantify this problem in Primary Care, with a special focus on diagnosis registry and diagnosis coding and on exploring novel ways for obtaining complete data. Furthermore we wanted to explore ways to enable EHR data reuse and sharing in a sensible way. We decided to broaden our horizons by working with rare diseases as well

as common diseases (cancer) but also by searching cooperation with medical specialists (hospital EHRs) and bio-informaticians. We chose to study disease coding by actually developing a coding system in the field of rare diseases and participating in the development of a coding tool. Because we discovered a lack of application of available genetics-knowledge in Primary Care, which is partly due to limitations in the EHR, we decided to develop a roadmap on this subject.

We set ourselves the following aims which in our opinion, we have successfully fulfilled.

Aims of the thesis

1. Assess (aspects of) data quality in (parts of) the Primary Care EHR, focussing on diagnosis registry as a central item and diagnosis coding as an important tool;
2. Find strategies and solutions to improve quality of (Primary Care) EHR data and to contribute to the enabling of reuse and sharing of EHR data.

Summary of Results

Part One - Quality of data: literature review & hands on identification of bottlenecks and areas for improvement

Literature on data quality in primary care is scarce; our review shows that available studies focus mostly on completeness and some also on correctness of registry of a small number of data items. Quality of data varies per GP centre and per data category. In general, coded data such as diagnoses, medication prescriptions and laboratory test results is registered fairly accurate and complete but there is room for improvement. Registry of vital parameters, risk factors and allergies & intolerances is often incomplete and incorrect (chapter 2).

For the studies described in chapters 3, 4 and 5 we used the routine EHR data extracted from practice centres in the Utrecht area, the Netherlands, that are a member of the Julius General Practitioners' Network (JGPN; 120 GPs, 50 practice centres, 290,000 patients). Coded and free-text primary care data from individual patients enlisted with these centres is periodically extracted to the central anonymized EHR. From the two studies (chapters 3 and 4) we performed to assess quality (completeness and correctness) of diagnosis registry in the primary care EHR we learned that the quality of coded data, as demonstrated for patients with cancer or suspected of having cancer, is suboptimal. GPs do know their cancer patients but

this does not mean that re-users of data can find these cancer cases using anonymized, coded EHR data easily. In both studies we compared cancer cases found in the EHR with the Netherlands Cancer Registry (NCR), a reference standard which is considered reliable. We found that when re-users of data try to select cancer cases using only coded data on a population level (chapter 3) as well as on an individual level (chapter 4) a large number of cancer cases is due to be missed (up to 40% false-negatives[#]) and a large number of cancer cases will be wrongly classified (up to 50%, false-positives[#]).

We conclude that the quality of coded EHR data improves over the years and that the type of EHR system used influences data quality. More specifically we found that in recent years diagnosis registry is more complete but as a drawback also the number of false-positives increases. In our linkage study (chapter 4) we discovered that for 77% of the missing (false-negative) cancer cases information about the cancer is available in the EHR elsewhere, merely in un-coded plain text. Also for 38% of seemingly wrong (false-positive) cases the GP appeared to have correctly registered the cancer diagnosis, including 31% (of 38%) where the diagnosis is not or not yet retrievable from the NCR.

In our study described in chapter 5 we reused coded as well as free-text EHR data for a research study to gain experience and by doing so also assessed GP management of women with breast cancer related concerns. We selected for the period under study all women from the EHR that presented with physical signs and symptoms of the breast (for instance pain in the breast or a lump) but also women that presented with fear of breast cancer or a family history of breast cancer. We found that concerns relating to breast cancer are presented to an average GP frequently (incidence rate 25.9 per 1,000 women per year), the larger part consisting of women experiencing physical signs and symptoms of the breast (85.3% or 23.2 per 1,000 per year). Symptomatic and asymptomatic women are referred for further investigation equally often (50%), so the GPs diagnostic workup phase does not seem to be paramount in the decision process. Referral practice for annual screening and genetic counselling is suboptimal and relevant information concerning family history of cancer is often missing in the EHR. Identification and management of women with an increased risk of breast cancer by GPs can be improved as well as identification and reassurance of women without an increased risk or relevant symptoms.

In this study we presented incidence rates based on extracted EHR data, taking into account the limitations of routine care data (see recommendations) but without applying

corrections to results because of lack of information on data quality of symptoms and family history registry in Primary Care. Furthermore, considering the dimensions of data quality (see table 1 introduction), in all three studies we experienced that data in the EHR can be incomplete, incorrectly coded and not up to date (current). We also found examples of lacking concordance and plausibility of data but these two dimensions were not structurally assessed in our studies.

Part two: Strategies & Solutions for improving data-quality and enabling reuse and sharing of EHR data

Improving disease coding systems and the development of tools mapping codes between those systems can help to increase EHR data quality, not just within primary care, but in health care in general. We have studied the quality of coding systems for EHR use in the field of rare diseases, in particular of metabolic disorders. Collectively, this group of diseases is large (>6.000) and growing steadily due to the identification of new diseases or variants of known diseases and improved clinician awareness. We know from experience that the annotation of rare diseases by means of adequate coding systems, and thus the possibility to accurately code patients with these disorders and identify them in EHRs, has been left behind. This has recently been confirmed by other researchers[3]. Our study, as presented in chapter 6, demonstrates that there are large gaps in the widely used existing international coding systems ICD-10 (International Classification of Diseases) (76% missing codes) and SNOMED-CT (Sytematized Nomenclature of Medicine Clinical Terms) of (54% missing codes) for metabolic disorders. Based on our clinical experience, we suspect that there may be similar gaps for other types of rare disorders. This has also been recognized by the SNOMED and Orphanet organisations which have joined efforts recently to improve coding for rare diseases. Existing gaps are a barrier to database- and data-sharing efforts especially for rare disorders, where the disease code is often used as a key to communication. We have shown that with the help of dedicated clinicians and code development agencies, the problem of coding gaps for rare disorders can be successfully addressed and that rich and up to date coding systems actually contribute to the quality of annotation for rare diseases, and thus to healthcare for patients with these diseases. Although this study was performed in a hospital setting, patients with a rare disease diagnosis should be recognizable as such also in Primary Care. Furthermore this study provided insight in the extensive process of developing a high-quality, usable, up-to-date coding system which will actually be adopted by prospective users.

Another barrier to data sharing for various purposes is the need to standardize semantics of data values such as diagnoses and other phenotypical codes. Ideally coding systems are aligned before data entry but often retrospective standardization will be required. In chapter 7 we describe the development of SORTA, a software tool to ease data (re-)coding and mapping between coding systems. We participated in this study by using SORTA for a pilot project to map an existing Dutch coding system for phenotype coding of physical symptoms to the international Human Phenotype Ontology (HPO) and demonstrated that existing coding systems can be harmonized with significant speed and quality improvement compared to earlier manual procedures.

Coding of disease diagnoses and symptoms is pivotal in EHR performance, but not all types of relevant medical information, e.g. family history, may be captured well in this manner. The EHR data structure design should allow for storing basically all relevant information and matching codes should be available to capture that information. In addition, the quality of the user interface itself is one of the other factors contributing to EHR performance. These aspects were studied in the context of delivery of genetic services, which despite readily available genetic knowledge is reported to be inadequate in Primary Care, among other medical specialties. We have confirmed this problem in chapter 5 where we found suboptimal referral practices for annual screening/genetic counseling but also information missing in the EHR concerning family history of women with breast cancer related concerns. In chapter 8 we identify existing barriers to implementation of genetic services in Primary Care such as shortcomings in design and interface of EHR systems to register genetic information. We propose a step-by-step roadmap including adjustments to the EHR and to existing coding systems to integrate genetics in General Practice and clinical research. This roadmap can be used as an example for introducing other complicated additions or adjustments to the EHR or to coding systems driven by needs in daily medical practice.

Lessons Learned

Summarizing and interpreting the results of the studies performed we conclude there are a number of lessons to learn from this thesis:

1. Data-quality in primary care currently is suboptimal, even for a key-item such as the coded diagnosis and for a serious disease such as cancer; relevant information regarding important risk factors such as family history is either frequently missing, incorrectly coded or cannot be found easily;
2. Despite suboptimal quality and subsequent reusability with clear limitations, the primary care EHR is a rich and voluminous source of (mostly uncoded) medical data, often comprising many years of follow-up
3. Because of suboptimal quality, primary care data should only be reused by people that fully understand the context of routine care data-entry and take explicitly into account the limitations of this data which can be assessed using the checklist in appendix 1 of this thesis;
4. There is a need to improve data-quality since reuse and sharing are desirable and expanding, ideally at the source (at data entry) and supported by adequate coding options. GPs can and should be facilitated and supported to achieve this in a number of ways (see recommendations below);
5. Adequate and up-to-date coding systems are pivotal for data reuse and sharing, not only for common but also for rare diseases and can successfully be developed using not only coding agencies but also dedicated clinicians and can be facilitated by software tools;
6. These coding systems are only valuable if they are continually be maintained, provide adequate synonyms and relevant crosslinks to other systems and are equipped with a guideline for use and an extensive fool-proof thesaurus;
7. Obligatory coding in EHR systems results in more complete registry but also leads to (over-) registration errors;
8. Linkage of EHR records to other data-sources can be useful to validate diagnoses but is currently complex and time-consuming;
9. The Primary Care EHR can be complementary to other data sources, even to a known reliable reference standard such as the Netherlands Cancer Registry;
10. Concerning reuse of Primary Care EHR data, there are many stakeholders involved, all interested in data reuse but from different perspectives: patients, GPs, the Dutch Association of General Practitioners (NHG), EHR suppliers, health inspection/insurance companies, (quality assessors), hospitals and out-of-hours clinics, researchers, Academic Practice Based Research Networks, departments of Vocational

Training for GPs, Coding agencies/organizations, and owners of External data sources such as the NCR.

Recommendations

The lessons learned can be translated into recommendations for the various stakeholders involved.

Patients:

Patients should be made aware of anonymous and non-anonymous EHR data reuse which requires their consent. Patients should be made aware that there is a difference between reuse of data for purposes of care, which cannot be done anonymously, and re-use for purposes such as research, which can be done with anonymized patient data. Also, there are ways to let re-users work with anonymized data and still provide options to contact the patient if necessary through a third party. This means there are options to avoid privacy issues.

Patients should be stimulated to take responsibility and make sure all important diagnoses and information regarding allergies and intolerances is known and registered in their medical file. The Primary Care EHR is central for registry of a patient's medical condition over the years and the GP in particular has an overview of this data. GPs receive results from laboratory and diagnostic tests and medical reports whenever their patients visit a medical specialist or a paramedical professional. Patients have the right to read and check their own medical files including their EHR record at the GP and should do so at least once to suggest possible corrections and additions. They should take responsibility and make sure all important diagnoses are recorded as Episodes in their EHR, as well as information regarding allergies and intolerances. If patients are convinced their GP is adequately keeping records and they trust their GP to keep doing this, it is safe and better to let him/her do the file-keeping. In recent years a number of companies have introduced software to maintain your own online medical file as a patient, beside or instead of the doctors' file, such as www.zorgdoc.nl, www.patient1.nl or www.healthvault.com. It is however not easy to assemble and interpret all the right information to keep your own medical file, as is illustrated by the premature exit of Google Health in 2011 (http://www.medischcontact.nl/Nieuws/Laatste-nieuws/Nieuwsbericht/125705/Google-health-stopt.htm). However, for patients with chronic or rare diseases the keeping of personal

records with certain measurements and symptoms or complaints can be very useful, especially when these personal records could be combined with medical files in the future. When the EHR record is complete and correct, only privacy issues could be barriers to data reuse and sharing, for instance when "opting-in" to share data through the LSP[4], and these issues should be individually weighed by every patient.

General Practitioner (GPs):

GPs should improve data-quality by investing in updating and coding key-items in their EHR files and by optimizing working processes at the GP practice. Our research shows that GPs *do know* their patients but relevant information is missing or hidden, because it is uncoded, within the EHR text and this is barring and confounding data reuse and sharing. Data quality in many GP EHRs can be improved. Despite the fact that we acknowledge the need to lessen the administrative burden and we therefore support like almost 70% of Dutch GPs the movement "het roer moet om"[5] we do feel data reuse will only expand and GPs can benefit if they improve their data quality. They should however be supported and facilitated much more than they are today (see recommendations below to Dutch College of General Practitioners and EHR suppliers). There are several ways to improve data quality from a GP perspective as we will explain.

GPs should invest in updating key-items (such as the important diagnoses, allergies & intolerances and risk factors) in their EHR files starting by making sure there are coded Episodes, correctly dated, for every disease the patient suffered or suffers that could have medical consequences or could be important regarding future medical decisions for the patient and/or his family. It is important to add the code for a disease only after the diagnosis has been determined and for suspected disease to code the main symptom, in line with the Guideline for Adequate EHR registry (ADEPD) Guideline ([6]). Also, the date of the Episode should actually be the date the final diagnosis was made, not the date of data entry.

GPs should evaluate and update working processes at the GP practice to integrate diagnosis registry after a letter from a hospital or diagnostic laboratory is received. Although most EHR systems do not adequately support registry of a positive family history for genetic disease such as cancer, it is useful to record this information since it can have major consequences (suggestion: create a separate Episode and code this with one of the available ICPC-1 codes for positive family history A29.01 – A29.07)..

We expect data reuse and sharing to expand in the coming years since this is the key to transition of care, such as the follow- up care for cancer patients from hospitals to Primary Care. We also expect, like the NHS in England (https://www.gov.uk/government/publications/personalised-health-and-care-2020), that it will not be long before patient empowerment will be taken a step forward and patients will seek and gain the right to access their full EHR record online and will even be able to add notes to their EHR (but not edit medical entries made by the GP). From a GP perspective it is undesirable to reuse and share low quality data for any purpose but certainly not with the patient since this can have many negative consequences (see discussion chapter 4), besides the risk of the patient losing confidence in their GP.

We recommend GPs to participate in relevant projects which require data reuse and sharing. We recommend GPs, that have improved the quality of their data, to participate in projects, for instance research, by sharing patient data. Sharing patient data for other purposes will further stimulate to improve its quality. Privacy barriers can be addressed and suboptimal coding could potentially be solved by arising text mining techniques[7]. Participating in a Practice Based Research Network at a local Academic centre can be feasible for GPs for instance to obtain benchmarking information ("spiegelinformatie") which can be valuable for management purposes. Last but not least, many EHR systems are provided with often unrecognized but useful ICT functions, for instance to build selection queries which can also be used to assess patient mix and assemble relevant management information.

Dutch College of General Practitioners (NHG):

The Dutch College of General practitioners (NHG) should on the one hand support and facilitate improvement of data quality, on the other restrict the unlimited reuse and sharing of EHR data. This is especially true for uncoded plain text in the EHR which is more prone to misinterpretation; reuse should be restricted to those that fully understand the Primary Care context in which the data was registered. The focus should be on improving, if possible (re-) coding, and subsequently sharing and reusing only a limited set of key data items, including the diagnosis. The quality of these data might be improved by the implementation of tools that can assist in recoding text (f.i. text mining and optimizing thesauri). This is in line with national developments such as the Continuity of Care record developed in the project Registry at the Source fom the NFU (Netherlands Federation of University Medical Centres) (http://www.nfu.nl/thema/registratie-aan-de-bron/). These key-

items should be chosen as a subset of the seventeen data-items that are part of the Continuity of Care record for hospitals, coded with SNOMED.

The Dutch College of General Practice should study and test techniques such as voice-recognition and text-mining to facilitate recording of high quality data at the source.
Supporting the improvement of data quality can be done in a number of ways, without losing sight of the actual goal of the EHR: supporting the primary process in every-day General Practice. For the GP this means facilitating easy, fast and user-friendly recording of consultations. In most EHRs this is suboptimal now and actions taken to improve data quality should not add to this burden but rather enhance functionality. This means that existing and upcoming techniques like voice-recognition and text-mining should be studied and tested.

Furthermore the implementation of the EHR Reference Model by EHR suppliers should be prioritized and stimulated and amendments should be made to the ADEPD and the Reference Model.
Also, the publication of the EHR reference model for EHR suppliers provides a firm basis but the implementation of the standard needs immediate attention. Many EHR suppliers have not yet implemented recent versions of the model. Furthermore the Reference Model and ADEPD guidelines should be amended on, among other things, the following items:

1. The Episode date should be the date the diagnosis was made (not the date of the first attached consultation) and hence the system should propose this date and furthermore it should be possible to alter this date. For instance for assessing familial risk of cancer, but also for research purposes it is important to know the age at diagnosis;
2. Enable attaching consultations to more than one Episode which will hugely simplify en fasten registration of consultations with multiple symptoms. Investigate the options and consequences of selecting more than 1 ICPC code per consultation;
3. Enabling the registry of a family history within the context of the EMR is necessary and should be facilitated for instance following the roadmap we presented in chapter 8;
4. Develop and add to the Reference Model a list of integrity checks for the EMR (for instance: it should not be possible to register prostate cancer with a female patient);
5. Design ways that support easy registration of suspected disease/differential diagnoses, pathological-proven disease, recurrent disease, etcetera;

6. Enable registry of a suspected or proven rare disease including relevant coding, by making use of existing and supporting connected coding systems, and develop a way to make these patients "visible" for the GP.

Investigate the digital integration of guidelines in the EHR and monitor the further development of the ICPC coding system.

Furthermore it would be useful to investigate ways to integrate guidelines in the EHR, for instance to support GPs in risk assessment of familial cancer and subsequent referral. Last but not least, the monitoring and further development of the ICPC-1 coding system maintains to be an important issue. This could be improved by extending the ICPC-1 with useful codes (for instance such as suggested in chapter 8), but also by providing a fool-proof thesaurus that would suggest coding options during registration of consultations. This thesaurus should comprise adequate synonyms and relevant crosslinks (f.i. with SNOMED) to facilitate data-sharing between sources.

EHR suppliers:

Investigate how user-interface and system design can be adjusted to support high-quality data entry at the source. We demonstrated that the type of EHR system used influences data quality in the EHR. EHR suppliers should investigate (in cooperation with an academic research team) how their user-interfaces and system design can be adjusted to actually improve data quality. This thesis provides a number of directions: facilitate user-friendly and accurately coded diagnosis registry without encouraging false-positives, add options to directly suggest and register the correct date of diagnosis, suspected, recurring and metastasizing disease, treatment, increased markers (eg for Prostate Specific Antigen/ or PSA) and a positive family history within the context of the EHR. Extending integrity checks at data entry would improve data quality as well.

Integrate referral guidelines into the EMR and facilitate feedback by easy-to-use selection queries.

Furthermore EHR suppliers should think about optimizing the availability of [online] up-to-date and easy to use referral guidelines by integrating them into the EMR. Also, facilitating feedback on a practice level by providing easy-to use selection queries for the GP would be worthwhile. Last but not least, the feasibility of voice-recognition and text-mining to facilitate structured data entry and retrieval should be investigated.

Quality Assessors (health inspection/insurance companies)

Stop measuring the quality of registration and find ways to adequately measure quality of care in dialogue with GPs.

Quality Assessors should be aware that data quality in Primary Care is suboptimal, even for a key item such as a cancer diagnosis. The current list of indicators (www.nhg.org/themas/publicaties/download-indicatoren) in Primary Care that can be calculated by retrieving information from the EHR relies heavily on adequate disease coding, for instance because the total number of patients with a certain disorder is used as a denominator. Taking into account the patient mix of a practice is justly becoming more important, which is another reason to aim for a reliable denominator. Also, by identifying patients with a certain disorder such as asthma or diabetes, information registered within those patients records such as smoking or blood pressure measurements are counted. This means that incorrect or incomplete diagnosis registry will bias results. Furthermore, we suspect that the quality of data for items in the EHR such as risk factors will also be suboptimal. This means that the "paper tiger" that is being created before our eyes, measures, inaccurately, the quality of registration instead of the quality of care and should be stopped, the sooner the better.

Education:

Add "adequate health file recording" to the MD curriculum and teach GPs necessary skills. Since digital health file recording is standard procedure in General Practice and more recently also in hospitals, there is a need for a new subject which should be added to the MD curriculum at University called " adequate health file recording". Doctors and GPs (to be) should be made more aware of consequences of recording choices and be taught necessary skills such as correct coding. This subject should also be integrated more widely into courses for practising GPs and other medical professionals too.

Researchers & Practice Based Research Networks (PBRNs):

When working with routine care data, validate diagnoses. Using EHR data means understanding and taking into account the limitations of routine care data. If researchers choose to work with EHR data, the diagnoses should be validated, either by linkage to external sources or other means. Linkage to other sources could decrease the number of false-negative records and hence more cases could be traced and included. False-positive records

can only (partly) be identified by studying the full EHR text, which is time-consuming and may be undesirable considering privacy issues.

PBRNs should seize the opportunities to support participating General Practitioners in improving the data quality in their EHRs for instance through providing benchmarking information. In this way they could provide additional advantages to practices to participate in their Networks. By providing bench marking and management information GPs could actually assess their patient mix and find EHR records that need quality improvement.

Future Research crossing boundaries

In this thesis we have been able to successfully assess data quality in Primary Care for certain data items and have also been able to identify strategies and solutions to improve data quality to actually enable reuse and sharing. We realize these are pieces of a large jigsaw that has to be completed in the coming years.

We believe we have only been able to obtain results and devise recommendations because we have crossed boundaries between academic disciplines: primary care, clinical genetics, medical informatics, computer science and bio-informatics. Working with scientists from other disciplines provides new insights and solutions to research questions.

A number of research challenges remain to be studied in the near future, all of them interdisciplinary. First of all, beside completeness and correctness, other dimensions of data quality should be evaluated: concordance, plausibility and currency, not only for diagnosis registry but also for other key data items, for instance risk factors, treatments and allergies.

Secondly, user-interface designers should be involved in these studies: what aspects of user interfaces in the different EHRs lead to differences in data quality and how can user interfaces be improved to enhance data quality?

Thirdly, the design, implementation and evaluation of actual interventions in the GP practice to improve data quality could provide the effective interventions needed to improve daily practice.

In the fourth place, possibilities of natural language recognition to suggest coding alternatives during data entry should be investigated.

Last but not least, it is necessary to experiment with patient entered data to assess usability of this data for various purposes, first of all care. This can be done for instance by developing an app or software tool, that patients can use to enter family history information.

About our patient

January 2020: a 48-year-old male visits his General Practitioner (GP) for a persistent mucus producing cough. A few days ago he made the appointment online and entered the reason for consultation and his complaints in the text box. Also he answered a few multiple choice questions presented by the EHR system triggered by the reason for encounter, about his complaints. Just before the allotted time the GP reads this information and looks at the patients' personal health data which includes data from various apps the patient uses such as exercise-apps. She notices that the patient has lost some weight but also that the training frequency and duration of this running-enthusiast have decreased substantially in the last 4 weeks. The GP asks some additional questions and performs a physical examination that turns out to be normal. She summarizes her findings orally using speech recognition software and along the way selects relevant codes, prompted by the system, for symptoms, signs and differential diagnosis. On her screen a pop-up appears (based on the guideline "Acute cough") asking her if a request for a chest X-ray should be send to the nearest hospital selected on diagnostic quality, reimbursement by the patient's health care insurance company and shortest waiting list. She clicks "yes" and schedules an e-consultation for follow-up a week later. The GP has a few minutes to spare and chats with her patient about his wife, children and his new job.

This thesis hopefully contributes to the improvement of EHR data in general and to the exposure of the true goldmine these data can become, with the ultimate goal to improve care for patients with common and rare diseases.

Notes

[#] False-negatives are cancer cases that are present in the NCR but not in the Primary Care EHR
False-positives are cases that are registered in the Primary Care EHR as having cancer but are not present in the NCR

Checklist before EHR data reuse and/or sharing

Nr	Item	Check
	Relevance	
1	List the data items you need	
2	Critically assess the items you just listed on necessity to answer your (research) question. Delete every item that is not absolutely necessary	
	Data Quality (for each data item)	
3	Gather existing information on the quality of data for each item (dimensions: completeness, correctness, concordance, plausibility, currency)	
	Origin (for each data item)	
4	Who entered this data?	
5	For what purpose was this data entered?	
6	What information is captured with this data?	
7	What information is NOT captured with this data?	
8	Could entry of this data be biased in any way?	
9	Are there other ways to enter the same data in this system?	
10	Could another user with the same role, decide to enter this data differently or not at all?	
	Condition (for each data item)	
11	When was the data entered (relative to disease process)?	
12	Is there any metadata available?	
13	Was the data changed since entry, why and by whom?	
14	Was there financial benefit for registering this data at all or in a certain way?	
15	List possible errors that could have occurred at data entry	
	Format (for each data item)	
16	What format was used entering the data? If coded: what coding system?	
17	If a coding system was used: what version, using which instructions? Check out the alternative codes in the system for registry of this item.	
18	Were there any restriction rules in the EHR system for entry of this data?	
	Assessment	
19	Critically assess every data item using the information gathered and determine usefulness for answering (research) question.	

NB: privacy /policy issues are not included in this list

1 Burstein MD, Robinson JO, Hilsenbeck SG, *et al.* Pediatric data sharing in genomic research: attitudes and preferences of parents. *Pediatrics* 2014;**133**:690–7. doi:10.1542/peds.2013-1592

2 Lateef F. Patient expectations and the paradigm shift of care in emergency medicine. *J Emerg Trauma Shock* 2011;**4**:163. doi:10.4103/0974-2700.82199

3 Fung KW, Richesson R, Bodenreider O. Coverage of rare disease names in standard terminologies and implications for patients, providers, and research. *AMIA Annu Symp Proc* 2014;**2014**:564–72.http://www.ncbi.nlm.nih.gov/pubmed/25954361 (accessed 16 Jun2016).

4 Zorgcommunicatie) V (Vereniging Z voor. Sharing your medical file and the LSP (Brochure: Uw medische gegevens elektronisch delen?). https://www.vzvz.nl/page/Zorgconsument/Links/Informatie/Informatiemateriaal

5 Het manifest van de bezorgde huisarts. Het roer moet om (free translation: 'We need a radical change'. www.hetroermoetom.nu; www.hetroergaatom.lhv.nl

6 The Dutch College of General Practitoners. Guideline adequate EHR registry. Revised version 2013. Available at: https://www.nhg.org/themas/publicaties/richtlijn-adequate-dossiervorming-met-het-epd.

7 Hoogendoorn M, Szolovits P, Moons LMG, *et al.* Utilizing uncoded consultation notes from electronic medical records for predictive modeling of colorectal cancer. *Artif Intell Med* 2016;**69**:53–61. doi:10.1016/j.artmed.2016.03.003