Secondary use of electronic medical records for clinical research: Challenges and Opportunities

Wen-wai Yim[1,2], **Amanda J. Wheeler**[3], **Catherine Curtin**[1,3], **Todd H. Wagner**[1,3], and **Tina Hernandez-Boussard**[2,3,4]

[1]VA Palo Alto Health Care System, 3801 Miranda Ave, Palo Alto, CA 94304

[2]Department of Medicine, Biomedical Informatics Research, School of Medicine, Stanford University, 1265 Welch Road, Stanford, CA 94305

[3]Department of Surgery, School of Medicine, Stanford University, 300 Pasteur Dr. Stanford CA 94305

[4]Department of Biomedical Data Science, School of Medicine, Stanford University, 365 Lasuen Street, Stanford CA 94305

Abstract

With increasingly ubiquitous electronic medical record (EMR) implementation accelerated by the adoption of the HITECH Act, there is much interest in the secondary use of collected data to improve outcomes and promote personalized medicine. A plethora of research has emerged using EMRs to investigate clinical research questions and assess variations in both treatments and outcomes. However, whether because of genuine complexities of modeling disease physiology or because of practical problems regarding data capture, data accuracy, and data completeness, the state of current EMR research is challenging and gives rise to concerns regarding study accuracy and reproducibility. This work explores challenges in how different experimental design decisions can influence results using a specific example of breast cancer patients undergoing excision and reconstruction surgeries from EMRs in an academic hospital and the Veterans Health Administration (VHA) We discuss emerging strategies that will mitigate these limitations, including data sharing, application of natural language processing, and improved EMR user design.

Keywords

Electronic medical record; breast cancer; mastectomy; statistical modeling

1. Introduction

With the passage of the HITECH Act as part of the 2009 American Recovery and Reinvestment Act [1], [2], government-led incentives for electronic medical record (EMR) adoption and demonstration of its meaningful use has propelled its ubiquity, as well as

spurred interest in leveraging EMR data for scientific research. Parallel to this, President Barack Obama's promotion of the Precision Medicine Initiative [3], [4] and the Cancer Moonshot Initiative [5] have raised increased awareness—as well as much influx in funding—towards leveraging large-scale computable EMR data for driving healthcare innovation and improving patient care.

However, despite the excitement, the resources and the availability of these data, there remains immense hurdles, clinically and technically, to effectively apply EMR data for secondary use. It is part of these many difficulties that have led to issues in the scientific community in terms of contradictory and non-reproducible studies emerging from these new data [6]–[8].

In this work, we examine the challenges associated with EMR research by stepping through the life cycle in designing, developing, and implementing a clinical research study using retrospective EMR data, with a study of post-mastectomy pain as an example. Some challenges discussed are the role of multi-institutional data availability and study variable choice on study design, the differences inherent in datasets, the challenges of measuring outcomes, the effects of modeling choice and data sparsity on results, and the difficulties of handling longitudinal data. The goal of this paper is not to give clinical recommendations, but rather to illustrate that typical measures of "statistical significance" in many clinical EMR studies can be undercut by the realities of the data collection, patient population biases, and limitations of data analytic tools, which are imperative to understand if we are to make meaningful use of EMR data.

2. Case study on pain outcomes following mastectomy

The study of pain following mastectomy and reconstructive surgeries is ideal to understand hurdles associated with EMR research. On one hand, breast cancer has a clear acute consequence without intervention (death); on the other hand, it is treatable with several combinations of treatment options, with some element of elective ability, and it is associated with many different symptoms. It is estimated that in the US one in eight women will develop breast cancer during their lifetime. [9] Following a diagnosis of breast cancer, women have many different treatment options, and variation in treatment pathways is well documented [10]. Often, treatment may include the surgical removal of cancerous tissue (e.g. mastectomy, lumpectomy) and a following breast reconstruction surgery. Unfortunately, chronic pain after breast surgery does occur, estimated to be between 20 to 40% [11]. However, little evidence exists regarding the association between adverse pain outcomes and different treatment choices and therefore both clinicians and patients lack necessary evidence to guide treatment decisions.

3. Methods

3.1 Data

The motivation of our case study is to characterize postoperative pain following different breast cancer treatments using EMRs from two healthcare systems: Stanford Healthcare Medical Center, a tertiary academic hospital, and the Veterans Health Administration

(VHA). In the Stanford Healthcare Medical Center, further referred to as AH, patient records were identified from Epic's Clarity relational database. VHA clinical data were extracted from the VA Corporate Data Warehouse (CDW), which includes national data from multiple medical centers [12].

3.2 Patient Population

Patients with breast cancer excision surgeries with a previous documented diagnosis of breast cancer, taken from 2008/01/01 to 2016/06/30, were extracted from the EMRs. Breast cancer diagnosis was identified by ICD-9-CM codes "174*" and "175*" and ICD-10-CM code "C50*". When multiple surgeries occurred, only the first excision surgery was included. Only patients with body mass index (BMI), preoperative pain, postoperative pain, pain at 30 days (+14/−7 days), and 3-months (+/− 14 days) were included. The number of patients with excision surgeries filtered by diagnosis codes were at 3387 and 3308 patients for AH and VHA respectively; after all exclusionary criteria was applied, 735 and 1210 patients remained.

3.3 Outcomes

To understand different aspects of patient pain after surgery, we assess multiple outcomes: postoperative length-of-stay in days (LOS), discharge day pain (a 0–10 numeric score), prescription of analgesics, physical therapy visits, and pain management visits (identified through institution-specific visit department names or classifications) at 30-days (+14/−7 days), 3-months (+/− 14 days), and 1 year (+/−84 days) outcome time points. Each outcome is meant to be a proxy measure for postoperative pain in different ways. For example, a patient may have an extended LOS if their postoperative pain is not appropriately controlled. Similarly, pain management, analgesics prescription, and physical therapy may be prescribed for patients with high pain.

3.4 Study Variables

Patient demographics variables included age, race/ethnicity, gender, Charlson score, and body mass index (BMI) at the time of surgery. Race/ethnicity was merged as follows: if hispanic was marked, they were considered hispanic, if there were more than one race marked, they were considered "other", otherwise the race was given as the final race/ethnicity value. We only took female and male genders in this study. The last BMI prior to surgery was taken. We also included diagnosis of anxiety (Supplemental Table 1) and diagnosis of depression (Supplemental Table 2), as they have been shown to have longer LOS and postsurgical complications [13]. In AH, medication names are mapped to RxNorm and in the VHA, drugs are mapped to the national drug formulary (Supplemental Table 3). We developed categories to rate the invasiveness of a breast cancer excision surgery. The list of codes using Current Procedural Terminology (CPT) were identified by examining the CPT ontology as well as including frequent billing codes by the department. A domain expert reviewed the list of codes and classified the excision surgeries as mild, moderate and severe (Supplemental Table 4). Reconstructions were similarly categorized as one of 3 categories: autologous, implant, and oncoplastic (Supplemental Table 5).

To control for confounding variables, we added information about other types of patient treatments that may occur from breast cancer patients measured at multiple time ranges. Included are additional breast excision and reconstruction surgeries, as well as common breast cancer treatments such as chemotherapy (Supplemental Tables 6), radiation therapy (Supplemental Tables 7), hormone therapy (Supplemental Tables 8) [10]; in addition to physical therapy (PT), pain management (PM) and other surgery (identified through HCUP surgery flag excluding excision/reconstruction surgery [14]). Time periods included several time ranges prior to surgery {(anytime up to 1 year prior to surgery), (between 1 year and 3 months prior to surgery), (3 months to surgery date)}, as well as several time periods between surgery date and an outcome time point {(surgery day to 3 months prior to outcome), (3 months to 1 month prior to outcome), (less than 1 month to outcome)}. These are not factored in simple bivariate analysis, however are included as variables in multivariate analysis.

3.5 Statistical Analysis

We measured bivariate association for each variable against each outcome using spearman correlation, fisher's exact test, chi-squared test, and Kruskal-Wallis ANOVA test as appropriate. Specifically, for two numeric variables, we used Spearman correlation; for two categorical variables, fisher's exact test was used if any values were less than five, otherwise chi-squared test was used; finally, for one categorical and one numeric variable, we used Kruskal-Wallis ANOVA. We developed regression models to test the association between our dependent and independent variables of interest. We applied linear regression for numeric outcomes and binary logistic regression for binary outcomes. For example, dependent variables pain and LOS are numeric variables so a linear regression model was developed; meanwhile, whether or not patients had an analgesic prescription or had a PM visit or a PT visit are binary dependent outcomes and therefore a logistic regression was used. Processing of EMR data was done using SQL and java. Statistical analysis was performed using python packages pandas and scipy and R packages glm and lm. This study was approved by the IRB at both the AH and the VHA.

3.6 Qualitative analysis

In addition to the standard statistical analysis, we use our case study to exemplify challenges of clinical studies using EMR. This includes the effects of using multiple institutions with different data organization, the influence of EMR definitions on results, the bias of disease populations on outcomes, the choice of modeling on results, and the difficulty of incorporating long-term information in EMR studies. When possible, we include quantitative analysis in addition to qualitative analysis.

4. Case study quantitative study results

Table 1 shows the populations between AH and VHA. As expected, the AH population includes more minority groups as well as a relatively younger population. The VHA population included more males, higher BMI, and higher reported rates of depression and anxiety status based on ICD-9-CM coding. A majority of patients in both institutions were outpatient (discharged the same day) or were discharged the next day. In our study

population, a minority of AH patients (20%) had a LOS of 2 nights or longer. AH did not include any severe class excision surgeries.

Table 2 and 3 show factors significant (as related to p-values for these models) for bivariate and regression analysis. While there are reoccurring themes for each outcome between AH and VHA, there was no major consensus regarding the role of each study variable for overall outcomes. Specifically, for each outcome, "clinically significant factors" change: (1) across different time points for the same outcome measurement (horizontal rows give different significance factors, e.g. pain at discharge, 30-day, and 1 year), (2) across different outcome measure proxies (vertical columns give different significance factors), (3) across different institutions (the same horizontal, vertical location in AH may have different factors in VHA) and (4) compared to the different analysis types (the same corresponding boxes in Table 2 and 3). Differences could either be due to genuine differences in the population however could also be due to fitting to data idiosyncrasies.

5. Challenges in secondary use of retrospective EMR studies

In this section, we identify challenges that are part of the life cycle for engaging in any EMR study. We describe the problems both in general and with respect to our specific post-mastectomy pain case study described previously.

5.1 Discrepancies in data availability directly impact experimental design decisions

AH is a tertiary care center, and therefore more likely to have patients come for treatment and receive postoperative care elsewhere. This results in a scenario in which many patients may be lost to follow up. On the other hand, VA data represents a national system, including approximately 153 hospital facilities and 788 community-based outpatient clinics [15], in which patients going to multiple VA hospital clinics will still be captured by the system. Furthermore, pain (one of our defining filters) is not always uniformly recorded across visits (pain score were not always collected, and if collected they were collected at different frequencies). In an attempt to capture long-term postoperative pain scores (i.e. 1 year follow-up), the time-window of assessment had to be larger, ranging from 281–449. In AH, the drop-off discharge (2277 patients in the beginning), went to 64, 41 and 32 percent at the 30 day, 3 month, and 1 year follow-up; in contrast, in the VHA, the drop off from discharge (2429 patients) was 74, 54, and 50 percent respectively. Thus, despite such a lenient window, the rates of loss-to follow up were substantial and only 32% had pain information recorded during the 1-year follow-up period.

Table 4 shows the raw number of patients that visit after surgery, prior to filtering by pain information available at different time points. The percentage of patients that drop off is steeper for AH compared to VHA; meanwhile, the percent of patients with pain information is higher in the VHA. The return of patients may signal a bias in the resulting population for people who had worse problems for both locations. VHA pain data because of its larger care system and mandated collection as a vital sign had substantially less of dropout using the same time specifications.

5.2 Multi-institutional involvement drives study design to the lowest common (and less ideal) granularity

While AH is a private single institution, the VHA CDW data is national data already transformed from individual centers around the country, some data granularity may be lost due to data processing. Biases arise from accommodating for multi-institutions' differences in practice. For example, while 0–10 numeric pain scores were available both in AH and VHA, there were differences in the collection methods. AH pain score data had much more granularity including location, descriptions, alleviating, aggravating factors, etc.

In order to normalize for the substantially less informative VHA CDW data, anatomic locations of pain scores were dropped from the AH data. Therefore, it would be possible (especially for time points further out) that the pain measured from the EMR was not in fact due to the breast excision or reconstruction surgery. The implications of these various constraints led to two-fold results: (1) noisier time windows for further time points and (2) unspecified pain measurements.

5.3 Methods of billing will affect data completeness

Furthermore, the difference in payment can be reflected in records. For example, AH is paid for care it provides (thus it has incentives to maintain coding systems that support its revenue). VHA is capitated and there is no incentive to code effectively. VHA physicians, therefore, tend to rely on more data in notes than in structured codes. For our cohort, the AH population patients were found to have a max, minimum, and average of 552, 16, and 141 ICD-9 codes for their entire record; in contrast it was 355, 1, and 96 in the VHA.

5.4 Study variables are defined on inherently imperfect EMR element proxies which may be designed arbitrarily and may be institution-dependent

Despite using common vocabularies such as ICD9, ICD10, CPT codes, there remains many different ways in which the same information, e.g. chemotherapy, can be identified. The chief reason for this is that EMR data are imperfect proxies for events we wish to study. An event can be identified using different methods which have different accuracies depending on institution. Even for the same concept, e.g. diagnosis of anxiety, there may be different definitions available [13], [17]. In this study, chemotherapy was defined using CPT codes alone, however it could also be defined using chemotherapy medications alone, or both in conjunction. In Table 5, we show that different definitions can lead to different frequency counts. These minor differences may propagate to differences in analytical conclusions. For example, in our bivariate analysis, depending on the identified outcome, e.g. LOS, discharge pain, 30-day, 3 month or 1 year pain, different chemotherapy-related confounding variables were found to be significant and at different levels. For discharge pain, whether or not if the patient received chemotherapy within 1 year to 3 months prior to surgery and whether or not if the patient received chemotherapy within 1 month prior to surgery were significant in using CPT and medication prescription dates, however the significance level for the latter for the former was for an alpha level of 0.1 in contrast to the latter which was at 0.05. For 30-day pain, whether or not the patient received chemotherapy immediately after surgery was shown to be significant within 0.05 alpha levels for CPT codes, but only within 0.15 alpha

levels for medications; and only medications showed a 0.05 significance for concurrent chemotherapy within the 30-day window.

Another challenge is that some variables may be institution-dependent which introduces room for data mapping issues. For example, for this study we identified use of physical therapy and pain management services as a visit to a particular department(s) within the healthcare setting. This required manual review of institutional-dependent coding. Identification of the closest corresponding data in the VHA were stop-codes, which are classifications of visit type that are assigned after the time of the visit. In another example, for identification of medication related variables such as hormone therapy and analgesic, required using two separate vocabularies (RxNorm in AH and NDF-RT in VHA) which are not easily mapable [18].

5.5 Missingness and scarcity of observable adverse outcomes biases findings

In order to compare how certain processes are correlated with good and bad outcomes, various procedures and outcomes must be recorded for fair comparison. However, patients may obtain their health care services from multiple locations, e.g. a patient that receives surgery in one institution but continues care back at their local unconnected clinics. Therefore, the information recorded in one EMR may not be complete or accurate. Furthermore, health care is a service that is typically sought out for when there are health concerns which leads to data representation problems. For example, we can only capture pain scores or use of physical therapy if patients request care—healthy patients are unlikely to return to the hospital for unnecessary care. For our outcomes, the percent of patients with PT or PM visits were lower than 25 percent and average pain scores were lower than pain score of 3 (Supplemental Table 9). Thus, there is an inherent bias in conclusions derived from EMR studies as they often represented sicker patients.

Besides challenges due to problems in follow-up, outcomes are difficult to measure for several reasons. First, patients do not always report their issues (patients may not seek help for pain). Second, some outcomes may be subjective and between patient variation can be significant. Third, even if reported by the patient, problems may not be recorded in the EMR. If reported, it may not be in structured form, and therefore sophisticated data mining algorithms may be necessary to utilize these data. Finally, another challenge is that typically the number of patients with severe issues is much smaller compared to the number of patients with "no problems". In our study, in all time periods, pain score measurements were concentrated towards the low pain scores as indicated by the percentile values (Supplemental Table 8). For all time periods, the 75 percentiles are at most at a pain score of 4. Meanwhile, use of PM and PT are relatively less populated, especially the pain management usage criteria, all less than 23% for both institutions.

5.6 Model choice and data idiosyncrasies can create arbitrary constructs of clinical significance

The differences in the predictive results for bivariate versus regression analysis and AH and VA populations is indicative of some very challenging issues faced by clinical sciences. The first is the modeling choices and defining which variables need to be controlled for in the

population. For example, bivariate analysis gives a very simple view. However, they do not take into account a host of influencing variables available in the electronic medical record, such as additional treatments after initial operation. On the other hand, adding many other variables makes statistical inference more difficult—thus without sufficient representation of various combinations of treatments, true relationships may be masked by idiosyncrasies in the dataset. The second issue is related to the latter point, which is, the influence in sample populations in influencing results. We see that in the VHA population diagnosed depression and anxiety is much more common, meanwhile the representation of breast cancer among men is higher as a result of a large male population in the VA. These can be true indicators of some connection of the disease or they can be spurious relations to the dataset.

The use of multivariate analysis allows for the inclusion of confounding effects such as other treatments and operations. Our choice of simple linear regression for numeric outcomes and binary logistic regression for binary outcomes was based on simplicity and interpretability. While this offers only a first order approximate picture of how individual study variables may have some effect on outcomes, the complexity of human biology guarantees there are many non-linear interaction amongst input variables that can affect outcomes that are difficult to capture in such a simplistic way. Restated, the assumptions of each model (as well as its significance measures) may not be acceptable for every question – and since we do not know the answer, it is impossible to know when they are or are not appropriate. Use of other models that fit to the data, for example decision trees, higher order regression models, or regression models with interaction terms can do better to model nonlinearities in data. Unfortunately, again, it is impossible to know when you are doing the "right" amount of fitting when you do not know the truth.

5.7 Longitudinal studies require controls for confounding events however then create problems of data sparsity

Although understanding of long-term treatments outcomes is crucial, longitudinal studies have more confounding variables to control for as patients will undergo various physiological changes and additional treatments. Indeed, we found that many variables that were significantly related to the outcomes, were confounding variables such as recent or current (within the time frame) treatments for chemotherapy, radiation therapy, hormone therapy, reconstruction surgery, and past opioid prescription. However, some variables indicate a more long-term pain problems (e.g. Chemotherapy, PM, or pain prior to surgery). Effectively, longitudinal studies are riddled with a practical problem: if confounding variables are not included, the study is not realistic, if they are included the data becomes sparse, making it even harder to draw conclusions. Table 6 show the amount of patients who have different other treatments, e.g. chemotherapy, within the time-frame of the study for AH, broken down relative to the operation day. For example, from the first row, 12 people had chemotherapy more than 1 year prior to the operative day Models not only have to generalize all the different types of treatments available over time (e.g. each box in the tables), but also have to deal with all permutations of them! Taking a logistic regression model as an example, a plain logistic regression with non-processed inputs will miss many non-linear relationships if no interaction terms are allowed. However, if interaction terms are

allowed, then there becomes combinatorially more variables to reason over, for which it is easy to overfit, and for which there is not enough data to provide enough examples.

To get a sense of patients in our selected cohort, we randomly sampled 10 patients from AH and manually reviewed all their medical notes within their 1 year time periods. Only one patient during their 1 year time period was not receiving additional revision surgeries, biopsies, or undergoing chemo- or hormone therapy. This manual review suggests that we are selecting patients who will tend to have additional medical problems whether or not they may be related to the original breast cancer, a metastatic condition, additional medical procedures, or continuing health issues. While one can attempt to control for this in the multivariate analysis, this would be a serious issue of bivariate analysis of long-term outcomes.

6. Discussion

In this study that compared data on breast cancer surgeries from two healthcare systems, we found that throughout the process of formulating a retrospective clinical experiment, there are a myriad of intertwining factors which provide deep challenges to unearthing clinically meaningful insights.

From conception, data availability and institutional differences in practice already exert limitations in experimental design. Specifically, outcomes at further time points can may be more biased because of a decrease in visits. Moreover, measurable outcomes and variables are limited by institution data capturing practices, often determined by the lowest granularity of data available across multiple institutions, as shown in the identification of pain information. There are practical challenges regarding the transferability of study definitions, even within the same institution, though this is incensed with multiple institutions which are organized under dissimilar data models. This was demonstrated by our experiment of using different chemotherapy identification techniques as well as our discussion of RxNorm versus NDF used for medication identification. Furthermore, generalizability of studies across disparate populations may confound experiments; our AH and VHA populations are good examples of this.

Finally, there are limitations in the ability of analytic methods to accurately model real-world effects. The nature of medicine and the complexity of biology leaves deep impressions on the data, which affects the practicality of experimental designs. Some of this was exhibited by the choice of modeling technique between bivariate and univariate analysis, revealing different significant variables per different outcomes across time. However, our experiments only showed the surface, even within bivariate analysis, there are different analytical methods, e.g. fisher's exact, chi-squared tests. For multivariate analysis, we chose linear regression, however we could have easily chosen higher-order fitting (e.g. polynomial regressions) or kernel methods. With many different variables, it is difficult for any algorithm to reason over and give clear results. Though it is difficult to reduce the number of variables as real patient clinical histories are in fact riddled with many confounding clinically relevant events.

Although many of the challenges discussed would be difficult to eliminate, there are several trends that can mitigate these problems related to population differences and data sparsity: which is the movement towards shared data repositories, the application of natural language processing, and increasing attention towards user-centered design for software development. All three developments target a strategy of increasing data points (patient numbers) and adding more variety in the sampling. The last item has an additional potential to increase the quality of data capture.

6.1 Moving towards common data repositories and normalized data representation

Efforts are underway to develop shared data repositories as well as data normalization standards [19–22]. This includes virtual warehouses that can query multiple institutions using the same data model for larger cohorts. These efforts will be critical for obtaining greater cohort population numbers which can be used in retrospective clinical studies. With a greater number of cases, many combinations of patient treatment and characteristic nuances may be taken account of whilst not exasperating the data sparsity problem (small numbers can give arbitrary attribution of clinical significance). Of course, pooling data for meaningful research is nontrivial, and before such efforts can take place, problems such as data security and privacy needs careful attention.

6.2 Application of natural language processing into the clinical domain

Use of natural language processing applied to biomedicine is increasing in prevalence [23]. As free text is a critical mode of communication among clinical teams, this will mitigate the difficulties of data incompleteness. Moreover, this area offers rich new information as a majority of clinical patient information is locked in text. For example, a clinical note may include information about patients' histories at other institutions or provide detailed and nuanced medical histories. In fact, for several patients who had reconstruction surgeries prior to their first excision surgery, manual review of clinical notes revealed that 4/6 had a previously unrecorded structured data signally an excision surgery. Though careful quantification of errors associated with NLP necessitates additional quality assessments, if applied judiciously, this may lead to an increase in the quality of data, as well as the quantity of data captured.

6.3 Push towards user-centered design for EMR software

Another strategy of maximizing information from data, is to decrease noise by increasing the quality of data at the point of capture. However, the burden of entering data generally falls on overworked clinical staff, whose first priority is to care for their patients. In fact, identified barriers for EMR adoptions includes perceived disruption to clinical workflow, software adaptability and complexity, as well as training and leadership engagement [24]. As the primary purpose of EMRs in a healthcare organization is for billing and record keeping, The result is uneven data capture that is sensitive to local institution protocol, software usability, and changes to billing practices. There is increased awareness towards usability problems and frustrations in EMR software [24–25]. This is leading to proposals for more user-centric design of EMR software that can shift the burden of making minute, nuanced data input decisions away from clinicians [26], [27]. With design of EMR software with emphasis on usability, clinicians can focus on care of patients with the burden of sifting

through correct documentation eased through better human interfaces. Cleaner and more consistent datasets would reduce data definition transferability problems, and with a greater number of data points, would help mitigate large variances in modeling to identify clinically significant variables.

7. Conclusions

This study highlights the importance of secondary use of electronic medical record data for clinical research. The adoption of EMRs has significantly increased the amount of detailed treatment information available today, that would be difficult to get using survey data or costly and laborious manual chart review. However, coupled with these data are important variability issues, that are magnified in multi-center studies. For example, the multiple-center studies can promote generalizable solutions and address population generalizability, it also carries with it limitations such as the problem of defaulting to a common lowest denominator in variable definitions, population differences, and information capture problems. Fortunately, movement towards collective data repositories and common data models will help mitigate these problems. Moreover, application of NLP, aptly used, will increase the quality and quantity of available data to build larger cohorts. Finally, better user-centric design of EMR software may increase the quality of data at the point of capture.

In conclusion, though EMRs offer opportunities for providing very granular and detailed items regarding a patient's clinical history and care, studies must account for many potential data modeling challenges, as we have highlighted in this work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. [Accessed: 26-Jan-2015] American Recovery and Reinvestment Act of 2009 (2009 - H.R. 1). GovTrack.us. [Online]. Available: https://www.govtrack.us/congress/bills/111/hr1

2. Adler-Milstein J, Jha AK. HITECH Act Drove Large Gains In Hospital Electronic Health Record Adoption. Health Aff. (Millwood). Aug.2017 36(8):1416–1422. [PubMed: 28784734]

3. [Accessed: 31-Aug-2017] Program Overview - All of Us | National Institutes of Health. [Online]. Available: https://allofus.nih.gov/about/about-all-us-research-program

4. [Accessed: 01-May-2016] Precision Medicine Initiative. National Institutes of Health (NIH). [Online]. Available: https://www.nih.gov/precision-medicine-initiative-cohort-program

5. [Accessed: 31-Aug-2017] Cancer Moonshot&#8480. National Cancer Institute. [Online]. Available: https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative

6. Chanock SJ, et al. Replicating genotype–phenotype associations. Nature. Jun.2007 447(7145):655–660. [PubMed: 17554299]

7. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? Nat. Rev. Drug Discov. Sep.2011 10(9):712–712. [PubMed: 21892149]

8. Ioannidis JPA. Why Most Published Research Findings Are False. PLoS Med. Aug.2005 2(8)

9. Pusic AL, et al. Patient-Reported Outcomes 1 Year After Immediate Breast Reconstruction: Results of the Mastectomy Reconstruction Outcomes Consortium Study. J Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. Aug.2017 35(22):2499–2506.

10. [Accessed: 31-Aug-2017] Cancer Facts & Figures 2017. [Online]. Available: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2017.html

11. Wilson GC, Quillin RC, Hanseman DJ, Lewis JD, Edwards MJ, Shaughnessy EA. Incidence and predictors of neuropathic pain following breast surgery. Ann. Surg. Oncol. Oct.2013 20(10):3330–3334. [PubMed: 23975300]

12. [Accessed: 31-Aug-2017] Corporate Data Warehouse (CDW). [Online]. Available: https://www.hsrd.research.va.gov/for_researchers/vinci/cdw.cfm

13. De Groef A, et al. Pain characteristics as important contributing factors to upper limb dysfunctions in breast cancer survivors at long term. Musculoskelet. Sci. Pract. Jun.2017 29:52–59. [PubMed: 28319882]

14. Weiser TG, et al. In-hospital death following inpatient surgical procedures in the United States, 1996–2006. World J. Surg. Sep.2011 35(9):1950–1956. [PubMed: 21732207]

15. Zullig LL, et al. Cancer Incidence among Patients of the United States Veterans Affairs (VA) Healthcare System. Mil. Med. Jun.2012 177(6):693–701. [PubMed: 22730846]

17. Barrera TL, et al. Diagnostic specificity and mental health service utilization among veterans with newly diagnosed anxiety disorders. Gen. Hosp. Psychiatry. Apr.2014 36(2):192–198. [PubMed: 24268573]

18. Pathak J, et al. Using RxNorm and NDF-RT to Classify Medication Data Extracted from Electronic Health Records: Experiences from the Rochester Epidemiology Project. AMIA. Annu. Symp. Proc. 2011; 2011:1089–1098. [PubMed: 22195170]

19. Klann JG, McCoy AB, Wright A, Wattanasin N, Sittig DF, Murphy SN. Health care transformation through collaboration on open-source informatics projects: integrating a medical applications platform, research data repository, and patient summarization. Interact. J. Med. Res. May.2013 2(1):e11. [PubMed: 23722634]

20. McMurry AJ, et al. SHRINE: enabling nationally scalable multi-site disease studies. PloS One. 2013; 8(3):e55811. [PubMed: 23533569]

21. Murphy SN, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am. Med. Inform. Assoc. Mar.2010 17(2):124–130. [PubMed: 20190053]

22. Makadia R, Ryan PB. Transforming the Premier Perspective Hospital Database into the Observational Medical Outcomes Partnership (OMOP) Common Data Model. EGEMS Wash. DC. 2014; 2(1):1110. [PubMed: 25848597]

23. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed. Inform. Oct.2009 42(5):760–772. [PubMed: 19683066]

24. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement. Sci. IS. Oct.2016 11

25. Kaipio J, et al. Usability problems do not heal by themselves: National survey on physicians' experiences with EHRs in Finland. Int. J. Med. Inf. Jan.2017 97:266–281.

26. Guo U, Chen L, Mehta PH. Electronic health record innovations: Helping physicians - One less click at a time. HIM J. Sep.2017 46(3):140–144.

27. Ratwani RM, Fairbanks RJ, Hettinger AZ, Benda NC. Electronic health record usability: analysis of the user-centered design processes of eleven electronic health record vendors. J Am. Med. Inform. Assoc. Nov.2015 22(6):1179–1182. [PubMed: 26049532]

Table 1

Patient demographics stratified by study site

Variable		AH	VHA
N		735	1210
Age (yr), mean (SD)		52 ± 12	59 ± 11
Gender, n(%)	Female	729 (99.2)	986 (81.5)
	Male	*	224 (18.5)
BMI, mean (SD)		26 ± 6	31 ± 7
Race, n(%)	Asian	182 (24.8)	0 (0)
	Black	18 (2.4)	330 (27.3)
	Hispanic	84 (11.4)	70 (5.8)
	White	385 (52.4)	724(59.8)
	Other◻	66 (9.0)	86 (11.7)
Charlson score, mean (SD)		5 ± 3	5 ± 3
Diagnosed anxiety, n(%)	No	632 (86.0)	774 (64.0)
	Yes	103 (14.0)	436 (36.0)
Diagnosed depression, n(%)	No	661 (89.9)	703 (58.1)
	Yes	74 (10.1)	507 (42.9)
LOS (days), mean (SD)	0	309 (42.0)	940 (77.7)
	1	265 (36.1)	148 (12.2)
	2+	161 (21.9)	122 (10.1)
Excision surgery class, n(%)	Mild	451 (61.4)	533 (44.0)
	Moderate	284 (38.6)	613 (50.7)
	Severe	0 (0)	64 (5.3)
Reconstruction type, n(%)	Autologous	40 (5.4)	16 (1.3)
	Implant	155 (21.1)	112 (9.3)
	Oncoplastic	19 (2.6)	*

*
Value less than 10 at time of first surgery,

◻Other is an aggregate of Native American, Pacific Islander, and Unknown categories, as well as any mixed combinations.

Table 2

Bivariate analysis significant indicators

Bivariate association for each variable against each outcome using spearman correlation, fisher's exact test, chi-squared test, and Kruskal-Wallis ANOVA test as appropriate.

	Discharge day	30 day	3 month	1 year
		AH		
LOS	Surgery category *** Reconstruction (autologous) *** Reconstruction (implant) ***	--	--	--
Pain	Surgery category *** Race * Reconstruction (implant) ***	Surgery category * Diagnosed Anxiety * Reconstruction (autologous) * Reconstruction (oncoplastic) ***	Race *	Surgery category * Diagnosed Anxiety * Reconstruction (implant) ***
Analgesic	--	--	--	Age * Charleston score *** Surgery category *** Reconstruction (implant) ***
PM	--	--	--	--
PT	--	Age * Charson_score * Depression *** Diagnosed Anxiety ***	Age * Charleston score ** Deppress_stat ** Diagnosed Anxiety ***	Age ** Charleston score ** Diagnosed Anxiety *
		VHA		
LOS	Surgery category *** Race ** Reconstruction (autologous) * Reconstruction (implant) ***	--	--	--
Pain	Depression *** Gender *** Surgery category *** Diagnosed Anxiety *** Reconstruction (implant) ***	Depression *** Gender *** Diagnosed Anxiety ***	Depression *** Gender *** Diagnosed Anxiety *** Race *	Depression *** Gender *** Diagnosed Anxiety *** Race **
Analgesic	--	Age *	Age-Charleston score **	Age ***

	Discharge day	30 day	3 month	1 year
				BMI[*]
		Depression[***]		Charleston score[**]
			Diagnosed Anxiety-Race[*]	Depression[***]
				Surgery category[***]
			Reconstruction (oncoplastic)[*]	Reconstruction (implant)[***]
PM	--	Diagnosed Anxiety[*]		BMI[**]
				Depression[***]
				Diagnosed Anxiety[*]
PT	--	Charleston score[***]	Charleston score[**]	BMI[*]
		Surgery category[**]	Sugery class[**]	Depression[**]
				Gender[*]
				Diagnosed Anxiety[*]

Signif. codes: 0.001 "***" 0.01 "**" 0.05 "*" For two numeric variables, correlations larger than 0.6 are shown with "=".

Table 3

Significant study variables for both AH and VHA for regression analysis

Treatments such as breast cancer excision surgery, reconstruction surgery, radiation therapy, hormone therapy, other surgery, at different time points relative to outcome points need to be controlled for, as they could be associated with our outcome of interest. Other surgery-related factors such as post-operative pain and pre-operative pain were also included in the models. Here surgery class 1, 2, 3 refer to mild, moderate, and severe respectively. Linear and logistic regression were used for numeric and binary outcomes respectively.

	Discharge day	30 day	90 days	365 days
		AH		
LOS	Charleston score*	--	--	--
	Surgery category(2)***			
	Reconstruction (autologous)***			
Pain	Age**	Age*	Race (Asian)*	Race (Black)*
	Race (Unknown)*	Reconstruction (implant)*	Race (Black)***	Race (Other)*
	Reconstruction (implant)***			
Analgesic	--		Race*	Charleston score**
PM	--	N/A	N/A	N/A
PT	--	Age***	Race*	Age**
		Anxiety**		Surgery category(2)*
		Reconstruction (implant)*		
		VHA		
LOS	BMI*	--	--	--
	Gender*			
	Surgery category(2)***			
	Surgery category(3)***			
	Race (Black)**			
	Race (Pacific Islander)*			
	Reconstruction (autologous)**			
	Reconstruction (implant)*			
Pain	Age***		Age*	Age***
	Charleston score-		Charleston score*	Depression*
	Depression-		Depression-	Race*
	Gender**		Surgery category (2)*	
	Surgery category (2)***		Race***	

	Discharge day	30 day	90 days	365 days
	Surgery category (3) **		Reconstruction (autologous) *	
	Reconstruction (implant) ***			
Analgesic	--		Depression status **	
			Diagnosed Anxiety *	
PM	--	N/A	N/A	BMI **
PT	--		Depression_stat *	BMI *
			Surgery category(2) *	Gender **

Signif. codes: 0.001 "***" 0.01 "**" 0.05

Table 4

Frequency of patients with visits and pain scores, stratified by days 90-day intervals following surgery

| | Days Following Surgery | | | | | | | |
	1–90	91–180	181–270	271–360	361–450	451–540	541–630	631–720
AH (N=2378) *								
Patients with visits, n(%)	2378 (100)	2107 (89)	2040 (86)	1892 (80)	1771 (74)	1538 (65)	1389 (58)	1229 (52)
Patients with a pain score, n(%)	2323 (98)	1719 (82)	1563 (77)	1278 (68)	1088 (61)	886 (58)	803 (58)	715 (58)
VHA (N=3255) *								
Patients with visits, n(%)	3255 (100)	3130 (96)	3031 (93)	2973 (91)	2896 (89)	2776 (85)	2675 (82)	2514 (77)
Patients with a pain score, n(%)	3168 (97)	2783 (89)	2614 (86)	2538 (85)	2405 (83)	2264 (82)	2139 (80)	2011 (80)

*
Visits are measures with respect to the first time period (between 1 day after surgery and 90 days); Patients with pain score measured are in respect to the number of patients with a visit within the given time period. Population numbers represent numbers prior to filtering by requiring a pain score at discharge and at subsequent time points.

Table 5

Frequency of patients receiving chemotherapy using CPT versus medication prescription for chemotherapy identification in AH

	CPT	Medication Prescription
>1 year prior to surgery	12	18
1 year to 3 months prior to operation day	169	173
3 months prior to and up to operation day	172	172
operation day		2
operation day to 3 months prior to time-point	375	374
3 to 1 month prior to time-point [*]	139	126
1 month prior to time-point [*]	109	88
time-point [*] after surgery	119	107

[*] time-point was set to 365 (+/−84) days

Table 6

Number of patients with confounding events relative to the operative day

	Prior to operative day			Operative day	Relative to 365 day time window			
	More than 1 year	1 year to 3 months	3 months to operative day		Operative day to 3 months prior to time point	3 months to 1 month prior to time point	1 month prior to and up to time point	Time point
Chemotherapy	12	169	172		375	139	109	119
Excision					184	#	#	#
Hormone therapy	12	24	13	#	384	155	64	268
Analgesic prescription	124	198	144	735	635	140	96	261
PT**	33	26	25	#	128	55	39	88
RT*	#	#	#	10	373	91	20	#
Reconstruction (autologous)	#			40	19	#	#	23
Reconstruction (implant)	#			155	102	34	21	57
Reconstruction (oncoplastic)	#	#		19	47	21	11	69
Other surgery	46	26	13		173	59	39	111

#Less than 10 patients in cell;

*RT: Radiation Therapy;

**PT: Physical Therapy