Biases in electronic health record data due to processes within the healthcare system: retrospective observational study

Denis Agniel,[1] Isaac S Kohane,[1,2] Griffin M Weber[1,3]

[1]Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck St, Boston, MA 02115, USA

[2]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

[3]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

Correspondence to: G M Weber weber@hms.harvard.edu

ABSTRACT

OBJECTIVE
To evaluate on a large scale, across 272 common types of laboratory tests, the impact of healthcare processes on the predictive value of electronic health record (EHR) data.

DESIGN
Retrospective observational study.

SETTING
Two large hospitals in Boston, Massachusetts, with inpatient, emergency, and ambulatory care.

PARTICIPANTS
All 669 452 patients treated at the two hospitals over one year between 2005 and 2006.

MAIN OUTCOME MEASURES
The relative predictive accuracy of each laboratory test for three year survival, using the time of the day, day of the week, and ordering frequency of the test, compared to the value of the test result.

RESULTS
The presence of a laboratory test order, regardless of any other information about the test result, has a significant association (P<0.001) with the odds of survival in 233 of 272 (86%) tests. Data about the timing of when laboratory tests were ordered were more accurate than the test results in predicting survival in 118 of 174 tests (68%).

CONCLUSIONS
Healthcare processes must be addressed and accounted for in analysis of observational health data. Without careful consideration to context, EHR data are unsuitable for many research questions. However, if explicitly modeled, the same processes that make EHR data complex can be leveraged to gain insight into patients' state of health.

Introduction

Rapid progress is being made towards the adoption and use of electronic health record (EHR) systems, resulting in massive amounts of data being generated through the routine delivery of healthcare.[1-3] This, in turn, is transforming biomedical research as investigators now have access to information on millions of patients through informatics tools that can query and analyze EHRs,[4-7] link to genomic and other types of biomedical data,[8 9] and scale to a national level and beyond.[10-14] However, there is a serious and increasing risk that naive use of Big Data analytical techniques without a full understanding of the complexities and limitations of EHR data is resulting in biased or incorrect medical findings.

An easily overlooked aspect of EHRs is that they are observational databases—the data reflect not only the health of the patients, but also patients' interactions with the healthcare system. For example, the date associated with a code for diabetes is when the physician made the diagnosis, not when the patient first developed the disease. Furthermore, the billing code used for that office visit might be influenced more by reimbursement policies than the original reason for the visit. Similarly, a patient might have an elevated white blood cell count; however, it will never be known unless a physician orders the laboratory test. Hripcsak and Albers describe this as a healthcare process model, where EHR data must be viewed as an indirect measure of a patient's true state due to the recording process.[15]

The recording process itself is affected by many factors, such as clinicians' decisions to order diagnostic tests and treatments and policies and workflows of provider and payor organizations. These are dynamic in that they vary over time as a result of evolving standards of care, changes in demand for care, and changing population demographics.[16] For example, separate studies, each examining routinely recorded patient data from at least 100 clinical practices, found the following: organizations were inconsistent in how they reported patient falls;[17] opioid prescribing increased from 2005-12, but at rates that differed by practice and patient population;[18] and financial incentives to screen for depression greatly increased the number of new depression related diagnoses.[19] The interactions between healthcare processes can be complex, as evident from the conflicting literature seeking to explain why patients admitted to the hospital during weekends have worse outcomes (known as the weekend effect).[20] Healthcare processes also vary by country. For example, the use of prostate specific antigen testing is generally higher in Western countries than in Asia,[21] and more than a dozen countries have implemented a Choosing Wisely campaign to reduce the use of unnecessary medical tests.[22] Distance matters too. Dozens of studies have shown that patients with cancer who live far from treatment centers are screened

WHAT IS ALREADY KNOWN ON THIS TOPIC

Dynamic processes within the healthcare system, such as the hours when clinics are open and when patients are scheduled to be seen, leave an imprint on electronic health record data

WHAT THIS STUDY ADDS

An evaluation of using the effects of healthcare processes on 272 laboratory tests to predict three year survival in the full patient populations seen over a year at two large hospitals

The hour of the day the test was ordered, the day of the week, and the amount of time between consecutive tests is more predictive of three year survival than the actual value of the test result, for most tests

less frequently, more likely to receive surgery than chemotherapy, and have worse outcomes.[23] Practical issues, such as how long it takes a clinician to enter a laboratory test order into an EHR,[24] the availability of certain tests in evenings or on weekends, and the level of automation in laboratories,[25] also affect the timing of EHR data.

The effects of healthcare processes on EHR data should not be viewed as data quality problems or noise.[26] This incorrectly suggests that these effects have no information value. In fact, they generate a signal, which can be used to identify subpopulations of patients and improve predictive models. This is especially true for laboratory tests, since they provide insight into a clinician's decision making process. For example, through analysis of EHR data, Hripcsak and Albers found the following: patients with kidney failure are more likely to have a creatinine measurement between 10 pm and 6 am than healthier patients;[27] the timing of glucose measurements can be used to stratify patients into health states;[28] and laboratory tests are ordered more frequently for sick patients.[29] In a study of 24 laboratory tests, they found that ordering patterns differ by clinical context, such as an inpatient admission compared with an ambulatory surgery event;[30] and, Levine evaluated methods for addressing this effect with four laboratory tests and five clinical contexts.[31] Lasko used an alternative approach based on unsupervised machine learning to identify temporal patterns of uric acid measurements associated with different diseases.[32] In an analysis of 70 laboratory tests and 14 000 patients, Pivovarov showed that the time interval between consecutive measurements adds information beyond just the test result value,[33] and we previously used these time intervals to derive normal ranges for 97 different tests.[34] Other research has shown that models predicting diagnoses can be improved by considering whether or not certain tests had been ordered;[35 36] and, acute care patients whose nurses recorded vital signs more frequently were more likely to experience a cardiac arrest.[37] In contrast to these studies, Dahlem found that the timing of diagnosis codes in EHR had relatively little predictive value;[38] however, the presence and timing of laboratory test data might reveal more about the thoughts and concerns of clinicians than the final diagnoses they record in the EHR.

In this study, we build on previous research into the healthcare process model, but on a larger scale. Specifically, we systematically evaluate the ability of 272 laboratory tests to predict three year survival across the full patient populations seen over a year at two large hospitals. We treat laboratory test data in the EHR as having two distinct dimensions. One dimension is the value of the test result, which is a measure of the patient's pathophysiology. The other is the timing of when the test was ordered, which is a marker of the underlying healthcare processes. For each laboratory test, we compare the predictive value of the patient pathophysiology and healthcare process dimensions first independently and then together.

Our hypothesis is that in a simplistic model of three year survival, healthcare process variables will have stronger predictive value than patient pathophysiology variables. Note that our outcome measure is not the absolute accuracy of the models, but rather the relative importance of healthcare processes when using raw EHR data. We make our entire dataset freely available to allow others to expand on this research in the future.

We chose to focus on the timing of laboratory tests, as opposed to many other potential measures of healthcare processes, for several reasons. First, as previously noted, other studies have found associations between the healthcare processes in laboratory tests and patient outcomes. Second, a large amount of laboratory test data are present in many EHRs. Third, the date of a laboratory test is usually recorded in EHR data, whereas other healthcare process variables, such as doctor experience, clinic operating hours, and hospital policies are more difficult to quantify or obtain. Fourth, there are hundreds of types of laboratory tests that are affected by different healthcare processes,[33] which enables us to detect variability in the predictive value of healthcare processes. Fifth, both the result value and time of a laboratory test can be expressed on a numeric scale, which enables us to create similarly structured patient pathophysiology and healthcare processes models. There are also natural groupings of both dimensions (eg, normal v abnormal test result values, and weekday v weekend timing), which we incorporate in our models.

Methods
Data source
This study is a retrospective analysis of patients with at least one clinical encounter over one year (28 July 2005 to 27 July 2006) at two large hospitals in Boston, Massachusetts: Brigham and Women's Hospital and Massachusetts General Hospital. Patients with unknown age or sex and patients older than 89 were excluded from the study, leaving 669 452 patients in the final cohort. Figure 1 shows that five years of observational electronic health record (EHR) data (28 July 2001 to 27 July 2006) for these patients were extracted from a single clinical data repository, the Partners Healthcare Research Patient Data Registry, which combines data from the two hospitals.

Three year survival was based on mortality data recorded on 27 July 2009—three years after the primary data collection period ended. Unfortunately, the actual date of death for deceased patients was not available in the source data. As a result, the follow-up time for patients whose last clinical encounter was near the start of the cohort period (28 July 2005) is close to four years. Also, the two hospitals determine patient deaths primarily by matching patient demographics to the Social Security Administration's Death Master File. However, missing and incorrect demographic information in both the Death Master File and EHR data can affect the accuracy of the matches and the resulting estimated survival rates. To circumvent these
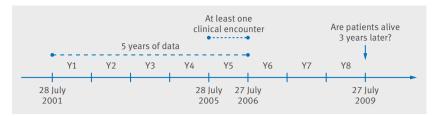
Fig 1 | Study design

limitations, our outcome was literally whether the EHR indicates that the patient is alive three years after our cohort period ended. We were not modeling time until death or conducting a traditional survival analysis.

We coded tests using the Logical Observation Identifiers Names and Codes (LOINC) terminology. A total of 272 distinct LOINC codes were used in this study, corresponding to all tests with numeric results that were ordered for at least 1000 patients in the final year of the data collection period, except for HIV related tests, which were removed for privacy reasons. Table S1 in the supplementary material lists the test codes, test names, and the abbreviations used in the other tables and figures.

Experiments

Two experiments were conducted. The first used the existence of a laboratory test in the patient's record to predict three year survival. In the second experiment, the patient pathophysiology and healthcare process dimensions of a single laboratory test observation were used to predict three year survival. The patient pathophysiology variables were the value of the test result and any high or low flag that was assigned to the test based on the reference range of the test. The healthcare process variables were the hour of the day the test was ordered and the day of the week it was ordered. We also considered whether that same test had previously been ordered for the patient. When two consecutive tests of the same type were present in the patient's record, we repeated both experiments, including the patient pathophysiology and healthcare process variables of both the main test and the previous test in the new models. An additional healthcare process variable—the number of hours between the two tests—was also included in the new models.

For each patient, one observation in the final year of the data collection period for each distinct LOINC code was randomly selected. For example, if a patient had three white blood cell count tests and two calcium tests between 28 July 2005 and 27 July 2006, one white blood cell count and one calcium test were selected. The dates of those two tests could be different. For each LOINC code, the most recent test previous to the randomly selected one was also recorded. The date of the previous test could go as far back as the start of the data collection period (28 July 2001). Not all selected tests had a previous test. A total of 8 867 400 observations of 272 laboratory tests were used in the experiments.

Predictive models

Logistic regression was used in the first experiment to model three year survival based only on the presence of a test and the age, sex, and race (ASR) of the patients. Generalized additive models with a logistic link were used in the second and third experiments to predict three year survival using only the ASR; ASR and a single patient pathophysiology or healthcare process variable; ASR and the combined patient pathophysiology variables; ASR and the combined healthcare process variables; and ASR and both the combined patient pathophysiology and healthcare process variables. Generalized additive models allow us the flexibility to model the effect of continuous variables, such as the test result value, without imposing restrictive assumptions like linearity. For example, having very high or very low white blood cell count is associated with decreased survival. Generalized additive models allow us to detect this type of nonlinear pattern. Additional details about the predictive models are presented in the supplementary material.

Patient involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Results

We first present a detailed analysis of a single laboratory test type, while blood cell count, to illustrate our approach. Then, we summarize the findings across all 272 test types.

White blood cell count

The full cohort of 669 452 patients had a three year survival rate of 95.0% (see supplementary materials, table S2). Of these patients, the 227 505 (34.0%) who had a white blood cell count test during the final data collection year had a three year survival rate of 92.9%. Thus, the presence of a white blood cell count test order is associated with a 2.1% lower survival rate (P<0.001). This is partially related to the demographics of patients who are more likely to receive a white blood cell count test. For example, the mean age (47.7 years) of patients with a white blood cell count test, is older than the mean age (43.8) of all patients (P<0.001). However, even when controlling for age, sex, and race (ASR), the conditional odds ratio of death for patients with a white blood cell count test was 1.45 (P<0.001). The ASR adjusted conditional odds ratio of death increases to 1.53 (P<0.001) for patients who had a pair of white blood cell count observations in the dataset.

White blood cell count is measured in thousands of cells per microliter, with a normal value between approximately 4 and 10. Causes for a low white blood cell count include autoimmune disorders, bone marrow
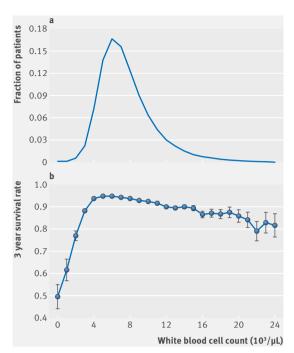
Fig 2 | The patient pathophysiology dimension of white blood cell count laboratory tests and survival

failure, and various cancer therapies. Causes for a high white blood cell count include bacterial infections, inflammatory disease, and leukemia. Figure 2a shows that the one randomly selected white blood cell count observation per patient was mostly likely to have a value within the normal range. The three year survival (fig 2b) for patients with a normal white blood cell count value is 94.3%. Not surprisingly, patients with a white blood cell count value that was flagged as abnormally low or high have lower survival rates of 86.7% (P<0.001) and 87.9% (P<0.001), respectively.

The value of the white blood cell count test only describes part of the picture—the patient pathophysiology dimension. Figure 3b shows that patients tested at 4 am with normal white blood cell count values have lower survival (85.4%) than patients tested at 4 pm with either abnormally low (93.0%, P<0.001) or high (91.4%, P<0.001) values. This finding is counterintuitive unless one considers an aspect of healthcare processes, which is that doctors generally only see sick patients in the middle of the night. In other words, even if a 4 am white blood cell count value is normal, it is abnormal for a patient to have a white blood cell count test ordered at that hour of the day (fig 3a).

For a similar reason, patients with a normal white blood cell count value on Sunday have the same survival rate (87.8%) as patients on Wednesday with either abnormally low (87.4%, P=0.59) or high (88.8%, P=0.08) values (fig 3d). The amount of time between consecutive white blood cell count tests is also associated with survival. For example, patients with a normal white blood cell count value less than one day after another white blood cell count test had a lower survival (78.9%) than patients with either

abnormally low (97.4%, P<0.001) or high (95.3%, P<0.001) white blood cell count values when it has been at least one year since the patient had another white blood cell count test (fig 3f). Doctors typically do not order a white blood cell count test for a patient on the weekend (fig 3c) or for a patient who just had a white blood cell count less than one day earlier (fig 3e), unless they believe the patient is sick.

Laboratory tests serve as biomarkers or proxies for complex biological processes that are difficult to measure directly. For example, after several days, blood cultures might confirm that a patient has a bacterial infection, but an elevated white blood cell count value is a much faster way to assess the patient's state of health. It is the bacteria, not the elevated white blood cell count, which is the cause of the patient's illness; and, if the physician had a way to instantly detect the bacteria, the white blood cell count test might not be necessary. However, in practice, the white blood cell count value is often the best information available. In a similar way, the healthcare process aspects of a white blood cell count test can be proxies for other processes within the healthcare system. For example, early morning tests are much more likely to be done in an inpatient setting than afternoon tests (fig 4a). Indeed, controlling for the clinical setting explains some, but not all, of the associations between hour of the day and survival (fig 4b). Countless other factors, such as the schedules of the clinics, doctors, nurses, phlebotomists, lab technicians, and patients might also be playing a role. The point is that the hour of the day of the white blood cell count test is not affecting the patient's health, but it is a readily available variable that encapsulates a great deal of information about the patient's interaction with the healthcare system.

Other laboratory tests

Table 1 shows that in the same way that abnormal values of different types of laboratory tests have different clinical significance, tests also vary to the degree and manner in which their healthcare process dimension can be used to predict outcomes.

For example, the presence of a laboratory test in a patient's record, regardless of any other information about the test result, has a significant association with the odds ratio of death in 233 of 272 (86%) tests (see supplementary material fig S1 and tables S5 and S6), based on Bonferroni adjusted P<0.05 (P<0.000184) to account for multiple hypothesis testing. Of these, the odds ratio of death is greater than one (lower survival rates) for 211 tests, with blood gasses having some of the highest odds ratios. However, 22 tests are associated with odds ratios less than one (higher survival rates), such as tests typically ordered during routine checkups at the two hospitals, including lipids (eg, low density lipoprotein, high density lipoprotein, etc) and prostate specific antigen.

Table 2 summarizes the results of the predictive models. Table S7 in the supplementary material provides details for each of the 272 tests. As an example, models for three year survival based on
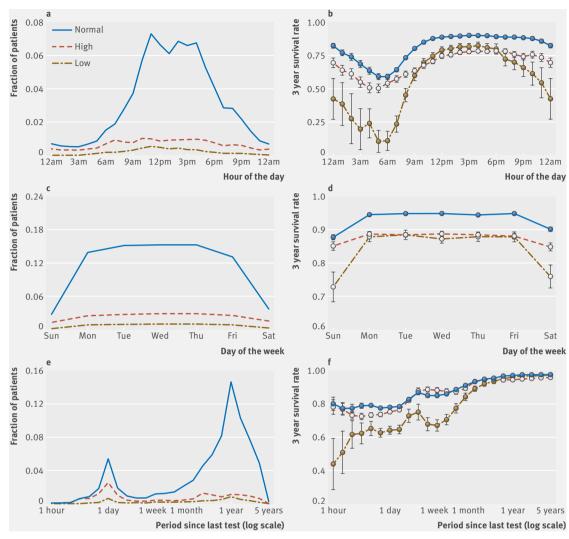
Fig 3 | Healthcare process dimensions of white blood cell count laboratory tests and survival. Note that (b) and (f) were smoothed using a three point running average

two consecutive tests were constructed for 210 tests. White blood cell is one of 127 (60%) tests where including both patient pathophysiology and healthcare process variables in the models is better than patient pathophysiology or healthcare process alone. Folate and triglycerides are examples of the 21 (10%) tests where healthcare process alone is better. Fibrinogen and testosterone are among the 26 (12%) tests where patient pathophysiology alone is better. For the remaining 36 (17%) tests, neither the patient pathophysiology nor the healthcare process variables improve a model based only on ASR. Overall, in the 174 tests where patient pathophysiology or healthcare process, or both variables improved the ASR model, healthcare process is better than patient pathophysiology in 118 (68%) tests. The time interval between consecutive tests is the single most predictive variable for 76 of 210 (36%) tests, followed by the value of the test result in 56 (27%) tests, and the hour of the day in 47 (22%) tests.

In a separate analysis described in the supplementary materials, we repeated the experiments using 30 day readmission as the outcome measure, rather than three year survival, and found similar results. For example, in the two-test models, the healthcare process variables are better than patient pathophysiology in 56 of 70 (80%) tests, with the hour of the day the best single variable in 46 of 107 (43%) tests, followed by the value of the test result in 16 (15%) tests, and the time interval between consecutive tests in 11 (10%) of tests (see supplementary materials, table S3 and S4).

Discussion

The speed by which technology is making Big Data available to biomedical researchers is outpacing the development of new analytical techniques to analyze these data and to understand the implicit processes that lead to their generation. Investigators are often unaware of the complexities of working with observational data
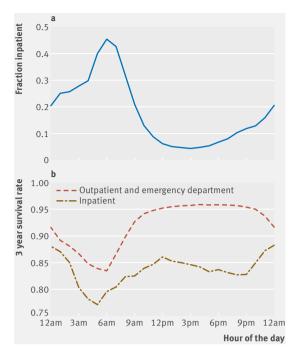
Fig 4 | White blood cell count by hour of the day. Note that (b) was smoothed using a three point running average

and do not appreciate the importance of healthcare processes. Savvy data analysts often have a toolbox of heuristic algorithms to clean up observational data. However, in these situations they are typically treating either patient pathophysiology or healthcare processes as noise and losing valuable information. Moreover, most of the noise models assume randomness whereas doctor and patient behaviors contribute to healthcare processes in purposefully biased ways.

Strengths and limitations of this study

In this study, we show the importance of healthcare processes in analysis of electronic health record (EHR) data using a large patient population and across many types of laboratory tests. To do this we are intentionally using overly simplistic (but equivalently constructed) models to isolate and compare the predictive value of individual patient pathophysiology and healthcare process variables within the context of messy, complex EHR data, and to show how easily it is to misuse and misinterpret EHR data by ignoring healthcare processes.

Obviously, a more complete model for predicting survival would include many more variables that describe patients' state of health, such as the diseases they have, drugs they take, smoking status, and family history. On the healthcare process dimension, we would analyze the data from the two hospitals independently,[39] separate the data by clinic and provider, and potentially include information about many other healthcare processes, such as the amount of data patients have,[40] hospital shift times, and the time between when diagnostic tests are ordered and when their results become available. However, the point of this study is not to develop a model that accurately predicts survival. Such a model might only be useful at the two hospitals where we conducted our study, since healthcare processes can be different at another healthcare facility, in the same way that patient characteristics vary across sites. Our dataset is also nearly a decade old. Although it is unlikely that this affects our overall conclusions, models incorporating healthcare process variables should be updated over time to capture changes in healthcare processes.

The key finding of this study is that the predictive value of healthcare process variables is often stronger than the result of the test when blindly using raw EHR data. Furthermore, the relative predictive value of the patient pathophysiology and healthcare process dimensions vary greatly between different test types, emphasizing the need to understand why a test would be ordered and what its result means within different contexts.

A limitation of this and other healthcare process research is that it can be difficult to identify the various processes that are being measured by a healthcare process variable. For example, Hripcsak and Levine show that different clinical contexts can result in similar ordering patterns, but for different reasons.[22 31] A healthcare process variable might also be related to patient pathophysiology. For example, certain laboratory tests have been shown to have true diurnal variations in controlled settings.[41 42] Thus, the information value of the time of day of a laboratory test might derive from both healthcare processes as well as biological processes. Additional research is needed to separate the two. Future healthcare process research should also involve discussions with patients to understand their effects on healthcare processes. For example, the decision to order an optional screening test can be influenced by patients' preferences, which in turn might vary based on their state of health.

Table 1 | Summary of results for three year survival models. Values are numbers (percentages)

Characteristic	Single test	Pair of tests
ASR (adjusted OR of death)		
Total	272 (100)	272 (100)
<1	22 (8)	19 (7)
>1	211 (78)	193 (71)
Not significant*	39 (14)	60 (22)
Predictive models		
Total	248 (100)	210 (100)
Best combined model:		
ASR, patient pathophysiology, and healthcare processes	168 (68)	127 (60)
ASR and healthcare processes	32 (13)	21 (10)
ASR and patient pathophysiology	30 (12)	26 (12)
ASR	18 (7)	36 (17)
Best single model:		
ASR and hour of day	104 (42)	47 (22)
ASR and day of week	14 (6)	7 (3)
ASR and time interval	NA	76 (36)
ASR and laboratory value	106 (43)	56 (27)
ASR and high or low flag	20 (8)	12 (6)
ASR	4 (2)	12 (6)

ASR=age, sex, and race; NA=not applicable
*OR significance is based on Bonferroni adjusted P<0.05 (P<0.000184).

Table 2 | Predicting three year survival using the healthcare process (HCP) and patient pathophysiology (PP) dimensions of laboratory tests

	HCP model better than PP model	PP model better than HCP model
Combined model with both HCP and PP better than HCP or PP alone	97 tests: 25VITD*; Abs Bands Manual; Abs Basos; Abs Basos Auto; Abs Eos; Abs Eos Auto; Abs Lymphs; Abs Lymphs Auto; Abs Monos; Abs Monos Auto; Abs Neuts; Abs Neuts Auto; ALKP; Alpha-Fetoprotein; ALT; Anion; AST; Atypical Lymphs; B12; Bands Manual; Basos; Basos Auto; Basos Manual; BUN; CA; CA15-3; CA19-9; CEA; CL; CO2; Cortisol; CPK; CPK-MB; CRE; CRP; CSF/F Unident; DBILI; Digoxin; Eos; Eos Auto; Eos Manual; FE; Ferritin; FIO2; Free T4; GLOB; GLU; HCG Quant; HCT; HDL; Hgb A1c; IgA; IgG; IgM; Ionized Ca; Ionized Ca Serum; K; LDH; LDL*; LIPS; Lymphs; Lymphs Manual; MCH; MCHC; MCV; MG; Monos; Monos Auto; Monos Manual; NA; Neuts; Neuts Auto; Neuts Manual; pH Blood; PHOS; PLT; Protein; PSA*; PT; PT-INR; PTH; PTT; RBC; Retics; T4; TBILI; TSH*; Urate; Urine Casts; Urine CRE; Urine Hgb; Urine pH; Urine RBC; Urine SpGr; Vancomycin Trough; VLDL*; WBC	30 tests: ALB; AMY; Base Excess Arterial; BNP; CA125; CHOL*; CSF/F Monos; CSF/F Nonhematics; ESR*; GGT; HGB; Lactate; LDL Calc*; Lymphs Auto; NT-proBNP; O2 Sat; OSM; pCO2 Arterial; pCO2 Blood; pH Serum; pO2 Arterial; pO2 Blood; RDW; T3; T3 Uptake*; Temp; TIBC; Troponin-I; Troponin-T; Urine WBC Sed
HCP or PP alone better than combined model	21 tests: Base Excess; CD16+CD56; CD19; CSF Basos; CSF Eos; CSF Neuts; CSF Reactive Lymphs; CSF/F Atyps; CSF/F Bands; CSF/F Basos; Dilantin; Folate; Glucose Blood; Metamyelos Manual; NH3; TRIG*; Urine K; Urine OSM; Urine Tot Prot; Urine Tot Vol; Urine WBC Screen	26 tests: CK-MB; CRP High Sens*; CSF Glucose; CSF Lymphs; CSF/F Lymphs; CSF/F Other Hematics; CSF/F RBC; CSF/F WBC; Fibrin D-dimer EIA; Fibrin D-dimer IA; Fibrinogen; Fluid Macros; Fluid Tot Prot; K Blood; MHCT; NA Blood; pCO2 Venous; pH Arterial; Testosterone*; Total Cells Counted; Urine ALB*; Urine ALB/CRE*; Urine CRE Timed; Urine MALB/CRE*; Urine NA; Vancomycin
Neither HCP nor PP improve ASR model	15 tests: Anticardiolipin IgM; Bili Conj*; CSF Bands; CSF RBC; CSF Unidentified; CSF WBC; CSF/F Eos; Fluid Blast; Fluid LDH; Lp(a); Myelos Manual; pH Venous; Urine GLU; Urine KET; Vancomycin Random	21 tests: ALC Toxic Screen; Bili Indir*; CSF Monos; CSF Non-Hematic; CSF NRBC; CSF Other Hematic; CSF Total Protein; CSF/F Polys; Ethanol; Fluid Glucose; Fluid NRBC; Haptoglobin; HCO3; Nuc RBC; pO2 Venous; Promyelos Manual; Urine BILI*; Urine CL; Urine NIT; Urine Urea Nit; Urobilinogen*

ASR=age, sex, and race

Tests are grouped based on whether adding the HCP or PP variables, or both, of the laboratory test improve a model based on only ASR.

*The ASR adjusted odds ratio of death is less than one in patients who were simply ordered any of the indicated tests.

Clinical and policy implications

Our findings warn about the naive use of EHR data. However, they also show that explicitly modeling the healthcare process dimension can both address some of the limitations of the data and increase the predictive value of the data. Box 1 shows a wide range of applications for this.

Comparison with other studies

The results of this study are consistent with previous research related to the healthcare process model that looked at either individual healthcare process variables, healthcare processes in small patient populations, or healthcare processes for a limited number of laboratory test types.[15-40] As in these other

---

Box 1: Applications for healthcare process modelling

Clinical care

- A clinician would not delay ordering a laboratory test to increase a patient's chance of survival. However, the clinician might use healthcare processes to see what tests thousands of other clinicians have ordered when treating similar patients; and, hospital administrators might look for outlier clinicians or outlier practices who are ordering tests in unusual patterns.
- Clinicians could also use healthcare processes as part of the move towards precision medicine by identifying subpopulations that have distinct healthcare process patterns after a new diagnosis or change in treatment strategy.[43]

Clinical research

- The effects of healthcare processes are often what clinical trials are designed to avoid. That is, variation in practice and clinical context are minimized to obtain the clearest perspective on pathophysiology or pharmacological differences. Thus, there might be a benefit to stratifying study subjects based on healthcare process variables. However, this should be done with caution since changes along the healthcare process dimension, such as increased ordering of laboratory tests, could be an early sign that certain patients are responding poorly to a treatment.
- In cases where patient pathophysiology and healthcare process are expected to be highly correlated, healthcare process variables can be used as proxies for missing patient pathophysiology data. For example, for certain laboratory tests, researchers using a claims database that does not include test result values could predict which ones are abnormal by searching for small repeat intervals.
- In many studies, researchers simply need to know the overall health status of a patient, in which case the combination of patient pathophysiology and healthcare processes create a much clearer picture than either one alone. This is important in comparative effectiveness research and pharmacovigilance studies, where looking for changes in either patient pathophysiology or healthcare processes could magnify the statistical power of the data.

Healthcare economics

- Insurance companies can incorporate healthcare processes in models of life expectancy or healthcare costs. This can potentially lead to more accurately aligned incentives for both patients and providers by rewarding behaviors, such as appropriate use of screening tests, that result in better health.

Healthcare policy

- Policy makers can study healthcare processes to identify overuse of diagnostic tests or disparities in access to healthcare among underserved populations. They can also track if regulatory changes or adoption of accountable care programs are having their expected effects on healthcare processes.

studies, we found that healthcare process aspects of EHR data can be used to infer information about patients' state of health that would not be known from patient pathophysiology alone. However, here we demonstrated the effects of healthcare processes on a large scale, enabling us to measure the relative predictive value of several patient pathophysiology and healthcare process variables across many different types of laboratory tests.

Conclusion

EHR data, without consideration to context, can easily lead to biases or nonsensical findings, making it unsuitable for many research questions. However, the same healthcare processes that make EHR data complex also leave a signal that can be useful if recognized and accounted for in models of patient health. This and other studies of healthcare processes have shown that it is a distinct dimension of observational data with a predictive value complementary to the patient pathophysiology dimension. For example, a normal laboratory test result is only one indicator of a patient's health. The fact that it was ordered at 4 am captures the physician's experience, intuition, and assessment of the patient's main complaint, baseline status, and physical exam, which are usually not explicitly coded elsewhere in an EHR or claims database. By ignoring healthcare processes or treating it as noise, investigators risk misinterpreting the actual patient pathophysiology and losing valuable information content.

1   National Research Council. The Learning Healthcare System: Workshop Summary (IOM Roundtable on Evidence-Based Medicine). Washington, DC: The National Academies Press, 2007.

2   Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Transl Med* 2010;2:57cm29. doi:10.1126/scitranslmed.3001456.

3   Centers for Medicare & Medicaid Services (CMS). "EHR Incentive Programs". http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms/30_Meaningful_Use.asp.

4   Nalichowski R, Keogh D, Chueh HC, Murphy SN. Calculating the benefits of a Research Patient Data Repository. *AMIA Annu Symp Proc* 2006;•••:1044.

5   Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE--An integrated standards-based translational research informatics platform. *AMIA Annu Symp Proc* 2009:391-5.

6   Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010;17:124-30. doi:10.1136/jamia.2009.000893

7   Copeland LA, Zeber JE. Advancing research in the era of healthcare reform: the 19th annual HMO Research Network Conference, April 16-18, 2013, San Francisco, California. Clin Med Res. 2013 Sep;11(3):120-2. doi:10.3121/cmr.2013.1175

8   Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. *JAMA* 2014;311:2479-80. doi:10.1001/jama.2014.4228.

9   Kohane IS. Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet* 2011;12:417-28. doi:10.1038/nrg2999

10   Weber GM, Murphy SN, McMurry AJ, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. *J Am Med Inform Assoc* 2009;16:624-30. doi:10.1197/jamia.M3191

11   McMurry AJ, Murphy SN, MacFadden D, et al. SHRINE: enabling nationally scalable multi-site disease studies. *PLoS One* 2013;8:e55811. doi:10.1371/journal.pone.0055811

12   Klann JG, Buck MD, Brown J, et al. Query Health: standards-based, cross-platform population health surveillance. *J Am Med Inform Assoc* 2014;21:650-6. doi:10.1136/amiajnl-2014-002707.

13   Patient-Centered Outcomes Research Institute (PCORI). PCORnet. http://www.pcori.org/funding-opportunities/pcornet-national-patient-centered-clinical-research-network.

14   Mandl KD, Kohane IS, McFadden D, et al. Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS): architecture. *J Am Med Inform Assoc* 2014;21:615-20. doi:10.1136/amiajnl-2014-002727.

15   Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc* 2013;20:117-21. doi:10.1136/amiajnl-2012-001145.

16   Homer JB, Hirsch GB. System dynamics modeling for public health: background and opportunities. *Am J Public Health* 2006;96:452-8. doi:10.2105/AJPH.2005.062059

17   Healey F, Scobie S, Oliver D, Pryce A, Thomson R, Glampson B. Falls in English and Welsh hospitals: a national observational study based on retrospective analysis of 12 months of patient safety incident reports. *Qual Saf Health Care* 2008;17:424-30. doi:10.1136/qshc.2007.024695.

18   Foy R, Leaman B, McCrorie C, et al. Prescribed opioids in primary care: cross-sectional and longitudinal analyses of influence of patient and practice characteristics. *BMJ Open* 2016;6:e010276. doi:10.1136/bmjopen-2015-010276.

19   McLintock K, Russell AM, Alderson SL, et al. The effects of financial incentives for case finding for depression in patients with diabetes and coronary heart disease: interrupted time series analysis. *BMJ Open* 2014;4:e005178. doi:10.1136/bmjopen-2014-005178.

20   Chen YF, Boyal A, Sutton E, Armoiry X, Watson S, Bion J, Tarrant CHiSLAC Collaborative. The magnitude and mechanisms of the weekend effect in hospital admissions: A protocol for a mixed methods review incorporating a systematic review and framework synthesis. *Syst Rev* 2016;5:84. doi:10.1186/s13643-016-0260-2

21   Kitagawa Y, Namiki M. Prostate-specific antigen-based population screening for prostate cancer: current status in Japan and future perspective in Asia. *Asian J Androl* 2015;17:475-80. doi:10.4103/1008-682X.143756

22   Levinson W, Kallewaard M, Bhatia RS, Wolfson D, Shortt S, Kerr EA, Choosing Wisely International Working Group. 'Choosing Wisely': a growing international campaign. *BMJ Qual Saf* 2015;24:167-74. doi:10.1136/bmjqs-2014-003821

23   Kelly C, Hulme C, Farragher T, Clarke G. Are differences in travel time or distance to healthcare for adults in global north countries associated with an impact on health outcomes? A systematic review. *BMJ Open* 2016;6:e013059. doi:10.1136/bmjopen-2016-013059.

24   Coleman JJ, Hodson J, Thomas SK, Brooks HL, Ferner RE. Temporal and other factors that influence the time doctors take to prescribe using an electronic prescribing system. *J Am Med Inform Assoc* 2015;22:206-12. doi:10.1136/amiajnl-2014-002822.

25 Jones RG, Johnson OA, Batstone G. Informatics and the clinical laboratory. *Clin Biochem Rev* 2014;35:177-92.

26 Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 2013;20:144-51. doi:10.1136/amiajnl-2011-000681

27 Albers DJ, Hripcsak G. A statistical dynamics approach to the study of human health data: resolving population scale diurnal variation in laboratory data. *Phys Lett A* 2010;374:1159-64. doi:10.1016/j.physleta.2009.12.067

28 Albers DJ, Elhadad N, Tabak E, Perotte A, Hripcsak G. Dynamical phenotyping: using temporal analysis of clinically collected physiologic data to stratify populations. *PLoS One* 2014;9:e96443. doi:10.1371/journal.pone.0096443

29 Hripcsak G, Albers DJ, Perotte A. Parameterizing time in electronic health record studies. *J Am Med Inform Assoc* 2015;22:794-804. doi:10.1093/jamia/ocu051

30 Hripcsak G, Albers DJ. Correlating electronic health record concepts with healthcare process events. *J Am Med Inform Assoc* 2013;20(e2):e311-8. doi:10.1136/amiajnl-2013-001922.

31 Levine ME, Albers DJ, Hripcsak G. Comparing lagged linear correlation, lagged regression, Granger causality, and vector autoregression for uncovering associations in EHR data. *AMIA Annu Symp Proc* 2016:779-88.

32 Lasko TA, Denny JC, Levy MA. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. *PLoS One* 2013;8:e66341. doi:10.1371/journal.pone.0066341

33 Pivovarov R, Albers DJ, Sepulveda JL, Elhadad N. Identifying and mitigating biases in EHR laboratory tests. *J Biomed Inform* 2014;51:24-34. doi:10.1016/j.jbi.2014.03.016.

34 Weber GM, Kohane IS. Extracting physician group intelligence from electronic health records to support evidence based medicine. *PLoS One* 2013;8:e64933. doi:10.1371/journal.pone.0064933.

35 Lin JH, Haug PJ. Exploiting missing clinical data in Bayesian network modeling for predicting medical problems. *J Biomed Inform* 2008;41:1-14. doi:10.1016/j.jbi.2007.06.001

36 Hagar Y, Albers D, Pivovarov R, Chase H, Dukic V, Elhadad N. Survival analysis with electronic health record data: Experiments with chronic kidney disease. Statistical Analysis and Data Mining: The ASA Data Science Journal 7, no. 5 (2014): 385-403. http://onlinelibrary.wiley.com/doi/10.1002/sam.11236/full

37 Collins SA, Cato K, Albers D, et al. Relationship between nursing documentation and patients' mortality. *Am J Crit Care* 2013;22:306-13. doi:10.4037/ajcc2013426.

38 Dahlem D, Maniloff D, Ratti C. Predictability Bounds of Electronic Health Records. *Sci Rep* 2015;5:11865. doi:10.1038/srep11865.

39 Madigan D, Ryan PB, Schuemie M, et al. Evaluating the impact of database heterogeneity on observational study results. *Am J Epidemiol* 2013;178:645-51. doi:10.1093/aje/kwt010.

40 Albers DJ, Hripcsak G. Using time-delayed mutual information to discover and interpret temporal correlation structure in complex populations. *Chaos* 2012;22:013111. doi:10.1063/1.3675621.

41 Jones AR, Twedt D, Swaim W, Gottfried E. Diurnal change of blood count analytes in normal subjects. *Am J Clin Pathol* 1996;106:723-7. doi:10.1093/ajcp/106.6.723

42 Sennels HP, Jørgensen HL, Hansen AL, Goetze JP, Fahrenkrug J. Diurnal variation of hematology parameters in healthy young males: the Bispebjerg study of diurnal variations. *Scand J Clin Lab Invest* 2011;71:532-41. doi:10.3109/00365513.2011.602422.

43 National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. National Academies Press (US), 2011. doi:10.17226/13284

Supplementary material: Additional information