Successful system, incomplete data

Caveats in reusing activity data from emergency-department whiteboards

Hertzum, Morten

Published in:
MEDINFO2017 Proceedings

DOI:
10.3233/978-1-61499-830-3-1033

Publication date:
2017

Document version
Other version

MEDINFO 2017: Precision Healthcare through Informatics
A.V. Gundlapalli et al. (Eds.)

doi:10.3233/978-1-61499-830-3-1033

Successful System, Incomplete Data: Caveats in Reusing Activity Data from Emergency Department Whiteboards

Morten Hertzum

Royal School of Library and Information Science, University of Copenhagen, Copenhagen, Denmark

Abstract

During the daily treatment of patients large quantities of data are recorded in electronic health records (EHRs). Compared to data in paper records, these EHR data are easily available for reuse in research and quality improvement. However, the opportunities for reuse depend on the quality of the data. In this study we analyze the completeness with which main treatment activities are recorded on emergency department (ED) whiteboards and whether completeness varies with the severity of the patients' condition. Data from 381,231 ED visits show that after the whiteboard had been in successful use for several years the clinicians recorded four of the five main treatment activities with a completeness of less than 50%. Completeness tended to increase with three indicators of the severity of the patient's condition: triage level, length of stay, and patient age. We conclude that the low completeness of the activity data probably prevents most types of reuse.

Keywords:

Quality Improvement; Electronic Health Records; Emergency Medical Services.

Introduction

The substitution of electronic for paper records has been a major development in healthcare organizations over the past decades. Electronic health records (EHRs) contain large quantities of data. These data are recorded during the daily treatment and care of patients but, subsequently, become available for other uses. EHRs have been associated with unprecedented opportunities for improving healthcare through the reuse of EHR data for clinical research, quality improvement, and other data-driven efforts to learn from past events [7]. However, such learning requires quality data. In this regard Weiskopf and Weng [16] contend that EHRs have "led not to improvements in the quality of the data being recorded, but rather to the recording of a greater quantity of bad data" (p. 144). This study focuses on the data recorded on electronic whiteboards in emergency departments (EDs), which are the common entry point to hospital treatment for most patients.

ED whiteboards contain data about the patients and their flow through the ED, such as the patients' time of arrival, triage level, current treatment activity, responsible clinician, and lab test results. Ready access to these data is central to the coordination of ED work and to each ED clinician's sense of overview [8; 11]. In addition to supporting work in the ED as it unfolds, the whiteboard recordings provide opportunities for learning. These learning opportunities include determining whether the recordings can be used for forecasting temporal patterns in patient arrivals, determining waiting and boarding times, identifying bottlenecks in the patient flow, assessing whether length of stay (LOS) targets are met, comparing the patient mix of EDs, and researching how the coordination of ED work is accomplished. Learning about these issues is important because the ED is a stressful environment for patients [5] as well as clinicians [4]. Improved knowledge of patient flows and resource bottlenecks can help EDs streamline work procedures, dynamically match resources to patient volumes, and prevent ED crowding [13]. We are involved in such learning efforts in the EDs in Region Zealand, one of the five healthcare regions in Denmark [e.g., 9; 11]. In the present study we analyze almost three years of log data from the whiteboards in the four EDs of the region *to assess the completeness with which main treatment activities are recorded and whether completeness varies with indicators of the severity of the patient's condition*. Information about the time spent on the different treatment activities is, for example, crucial to understanding how crowding arises and when countermeasures are required. Thus, the generation of accurate forecasts of crowding presupposes reasonably complete recordings of the treatment activities.

Weiskopf and Weng [16] identify three fundamental dimensions of data quality – completeness, correctness, and currency – and two auxiliary dimensions (concordance and plausibility) that often serve as proxies for the fundamental dimensions when they cannot be assessed directly. Completeness, the focus of this study, concerns whether a piece of data about a patient is present in the EHR. Correctness concerns whether the data that are present in EHRs are also accurate. And currency concerns whether the data are representative of the patient's state at a desired point in time, often interpreted as whether the data are recorded in the EHR within a reasonable period of time following measurement. Previous studies have demonstrated that the data quality is modest for multiple types of EHR data [2; 7; 12; 14]. For example, Chan et al. [2] reviewed the completeness of blood-pressure recordings across multiple studies and found that the number of complete recordings ranged from 0.1% to 51%. In addition, Brennan et al. [1] found that British hospital statistics for 2009-2010 showed nearly 20,000 adults attending pediatric outpatient services and over 8,000 males admitted to gynecology inpatient wards. Several reasons have been proposed to explain modest data quality, including habits, lack of time, failure to capture data that became available to clinicians who were not part of the department, and transcription errors in transferring data from paper charts to EHR [e.g., 12; 14]. At root, data quality suffers from differences in priorities between day-to-day clinical work and work such as research and quality improvement. In day-to-day clinical work data quality is secondary to patient treatment.

Method

The four EDs in Region Zealand were part of medium-sized hospitals and collectively served a population of approximately 817,000 citizens. Prior to conducting the study we obtained approval from the healthcare region.

The Whiteboard

The four EDs installed the same electronic whiteboards, which were introduced in December 2009 (ED1), January 2010 (ED2), January 2011 (ED3), and May 2011 (ED4), respectively. The whiteboard supplemented the electronic patient record by providing procedural information about the patients. Some of the whiteboard data, such as lab test results, were automatically updated when new data became available, but the majority of the whiteboard data were entered and updated manually. In this study we focused exclusively on data that were entered and updated manually. Manual data entry and updating could be done on the wall-mounted whiteboard displays (see Figure 1) as well as on any other computer in the EDs.

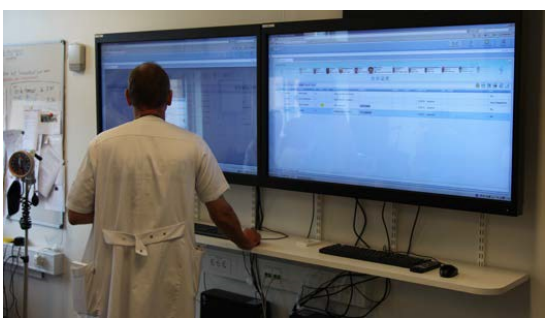

Figure 1 – The whiteboard at ED4. For each patient the whiteboard gives one row of information, including time of arrival, triage level, first name, age, responsible physician, current treatment activity, lab-test results, and next stop.

Before turning to analyze the completeness of the whiteboard data and thereby their reuse potential, it is important to note that the whiteboard was successful with respect to its primary purpose of supporting the clinical work in the EDs. A physician at ED3, for example, said in an interview that the whiteboard "*gives a great overview. I cannot imagine that we could do without it.*" A nurse at ED1 expressed agreement: "*We use it all the time*". More formally, a survey at ED1 and ED2 showed that the clinicians experienced an improved overview of their work when the electronic whiteboard replaced the former dry-erase whiteboards [8]. The survey also showed that, to a larger extent, the electronic whiteboard made information available where and when clinicians needed it.

Log Data

All changes of the whiteboard content were automatically logged. For the purpose of this study the whiteboard vendor, Imatis, produced a version of the logs from which all patient names, clinician names, and other information that might identify persons had been removed. These anonymized log data covered the three-year period 2012-2014. However, we had to discard the periods January 2013-January 2014 (ED1) and November 2013-January 2014 (ED2-ED4) from the analysis because they contained long intervals of no data. After also removing 741 outliers (defined as ED visits longer than seven days, i.e. more than 50 times the median length of stay), the dataset comprised 381,231 ED visits. Table 1 shows the division of the visits into EDs and years.

Table 1 – The 381,231 ED visits divided onto ED and year

ED	2012 *(Jan-Dec)*	2013 *(Jan-Oct)*	2014 *(Feb-Dec)*
ED1	33,040	-	30,719
ED2	40,445	32,844	37,396
ED3	32,677	28,527	34,304
ED4	38,670	32,628	39,981

The data for the 381,231 ED visits consisted of over 10 million log entries, each documenting an event that changed the whiteboard content. A log entry contained a timestamp, an event type, any values associated with the event, and a system-generated identifier of the visit to which the event pertained. For example, the event type 'WAITING_FORChanged' along with the event value 'Nurse' indicated that the patient was now waiting to be seen by a nurse. This event type marked the start of a treatment activity in the patient's progress through the ED workflow. Across the EDs different sets of treatment activities (i.e., different sets of event values) were used for indicating the stages of this workflow. However, the workflows of the EDs shared five main activities: triage, waiting to be seen by a nurse, waiting to be seen by a physician, examination (by a junior physician), and review (by a senior physician). No other treatment activity was recorded more consistently than these five main activities. The log data also contained information about changes in, among other things, responsible physician and lab test results; but in this study we focus on the five main treatment activities because the flow of the patients through the ED is important to, for example, forecasting and preventing crowding.

Results

The ability to record the patients' progress through the ED workflow was included on the whiteboard because at-a-glance access to this information was deemed important to the ED clinicians' overview of their collective work. Yet, even the five main treatment activities were recorded for only a subset of the ED visits (see Table 2). The main activity recorded most and least often differed across the EDs. In ED3 and ED4 waiting to be seen by a nurse was recorded for 74% and 76%, respectively, of the patients in 2014 and for similarly high percentages of patients in the preceding years. The only other activity recorded for the majority of the patients was waiting to be seen by a physician in ED2 (52% in 2014). Conversely, the activity of triage was recorded for less than 1% of the patients in ED1 and ED3. For all EDs at least one of the five activities was recorded for no more than 20% of the patients.

It could be hoped that the completeness of the recordings increased over time because the clinicians appreciated the improved overview or became more conscientious in their use of the whiteboard. The data provided little ground for such hopes. Rather, the trends in the data from 2012 to 2014 showed a mixed picture with nine increasing trends, four decreasing trends, and seven unchanging trends (see Table 2). It should also be noted that prior to 2012 the whiteboard had been in operational use at the EDs for between half a year and two years; thus, work procedures involving the whiteboard had had time to stabilize.

Table 2 – Completeness of the recordings of the five main treatment activities

Treatment activity	2012		2013		2014		Trend [a]
	N	%	*N*	%	*N*	%	
ED1							
Triage	12	0.04	-	-	3	0.01	→
Nurse	267	0.81	-	-	143	0.47	→
Physician	5,552	17	-	-	6,404	21	↗
Examination	11,292	34	-	-	11,252	37	↗
Review	8,264	25	-	-	8,935	29	↗
ED2							
Triage	10,028	25	9,521	29	11,871	32	↗
Nurse	4,604	11	3,565	11	3,435	9	↘
Physician	19,616	49	14,919	45	19,459	52	↗
Examination	14,217	35	11,161	34	11,217	30	↘
Review	17,772	44	14,436	44	17,228	46	↗
ED3							
Triage	37	0.11	37	0.13	22	0.06	→
Nurse	24,270	74	17,896	63	25,436	74	→
Physician	10,443	32	9,576	34	12,024	35	↗
Examination	9,224	28	7,732	27	10,269	30	→
Review	1	0.00	1	0.00	0	0.00	→
ED4							
Triage	9,867	26	5,815	18	7,398	19	↘
Nurse	27,366	71	25,018	77	30,201	76	↗
Physician	2,770	7	10,332	32	15,569	39	↗
Examination	8,531	22	7,163	22	9,577	24	→
Review	7,793	20	5,112	16	6,091	15	↘

[a] Trend in completeness from 2012 to 2014: ↗ - an increase of more than 2.00 percentage points, ↘ - a decrease of more than 2.00 percentage points, → - a change of at most 2.00 percentage points.

The first activity in the ED workflow, triage, illustrated the important point that failing to record triage as the current treatment activity for a patient did not indicate that the patient was not triaged. The whiteboard gave the triage code for 77% (ED1), 37% (ED2), 10% (ED3), and 40% (ED4) of the patients. For all four EDs the number of patients with a triage code exceeded the number of patients for which triage was recorded as the current treatment activity. Probably, triaging a patient and recording the triage code were experienced as more clinically relevant by the ED clinicians than recording that their current treatment activity was to triage the patient, especially because triage was a brief procedure.

An important consideration in assessing the possibilities for data reuse is whether completeness varies with indicators of the patient's condition. Tables 3 to 5 show how completeness varied with three indicators of the severity of the patient's condition. To save space each table gives the data for only one of the four EDs.

Table 3 shows how completeness varied with triage level in ED1. Waiting to be seen by a physician was more often recorded as the current treatment activity for patients triaged 4 and 5 (i.e., the most severe cases) and examination and review were most often recorded for patients triaged 2 and 3. It might have been more clinically relevant to record the activity in progress for the patients triaged 2 and 3 because they were in the ED longer than the other patients. For example, many of the patients triaged 4 and 5 were quickly transferred to inpatient departments for specialist treatment. In ED1 the completeness of the treatment activities triage and waiting to be seen by a nurse was largely unaffected by triage level. The patterns for physician and examination were roughly similar in the other EDs, whereas the patterns for triage, nurse, and review were different. For example, waiting to be seen by a nurse in ED4 was recorded for 80% of the patients triaged 2 and 3 but only for about half as many at the other triage levels.

When completeness varied with the triage level, it tended to be by higher completeness at medium or high triage levels.

Length of stay (LOS) directly indicates a patient's demand on ED resources and indirectly indicates the severity of the patient's condition. Table 4 shows that the activity of triage tended to be recorded more often for the patients that ended up staying longer in ED2. Waiting to be seen by a physician was most often recorded for the patients who stayed 3-8 hours; ED3 was similar in this respect. In ED2, the activity of review was recorded as the current treatment activity for fewer and fewer patients as LOS increased. The completeness of the review recordings also varied systematically with LOS in ED1 and ED4 but in the opposite direction: completeness increased with increasing LOS. Waiting to be seen by a nurse was the only treatment activity the recording of which did not vary appreciably with LOS in any of the EDs. The four other treatment activities were, in one or two of the EDs, recorded more often with increasing LOS. The only instance of a decreasing trend was for review in ED2.

Table 3 – Completeness of treatment-activity recordings for ED1, by triage level

Triage level	ED visits	Triage %	Nurse %	Phys. %	Exam. %	Review %
1	1482	0.00	0.00	7	34	38
2	11857	0.01	0.03	12	72	50
3	12131	0.02	0.19	15	70	56
4	23230	0.05	1	28	16	12
5	397	0.00	5	24	1	2

Note. The table includes only the 49,097 (77%) ED1 visits for which the triage level was available.

Because older patients tend to be weaker than younger patients, we use age as a third indicator of the severity of the patient's condition. In ED4, waiting to be seen by a nurse was recorded less often with increasing patient age, while waiting to be seen by a physician, examination, and review were recorded more often with increasing patient age (see Table 5). In the other EDs waiting to be seen by a nurse was recorded

about equally often for the different age groups but the pattern for waiting to be seen by a physician and for examination resembled that in ED4. In ED1 the pattern for the activity of review also resembled that in ED4, but in ED2 it was reversed (as it was for LOS). The pattern that completeness tended to increase with increasing patient age was stronger than the patterns for triage and LOS.

Table 4 – Completeness of treatment-activity recordings for ED2, by length of stay (LOS)

LOS *(hours)*	ED visits	Triage %	Nurse %	Phys. %	Exam. %	Review %
0-2	63,490	22	9	41	27	54
3-5	32,087	35	12	61	41	39
6-8	5,533	38	14	67	47	27
9-11	1,896	32	9	55	39	19
12-	7,679	42	11	53	37	11

Table 5 – Completeness of treatment-activity recordings for ED4, by patient age

Patient age *(years)*	ED visits	Triage %	Nurse %	Phys. %	Exam. %	Review %
0-19	12,982	16	82	15	8	7
20-39	15,753	28	77	30	24	19
40-59	19,265	29	73	37	31	23
60-79	21,026	27	71	42	39	29
80-	11,531	17	68	43	40	30

Note. The table includes only the 80557 (72%) ED4 visits for which the patient age was available.

For all three indicators the overall, but not unanimous, pattern was that completeness increased when the indicator pointed toward patients with more severe conditions.

Discussion

After the whiteboard had been in successful use in the EDs for 2-4 years the five main treatment activities were recorded in 2014 for an average of 13% (triage), 40% (nurse), 37% (physician), 30% (examination), and 23% (review) of the patients. These averages hide considerable variation across the EDs; but no ED recorded more than one treatment activity with a completeness of more than 50%, and all EDs recorded at least one treatment activity with a completeness of less than 20%. The substantial incompleteness in the recording of the patients' current treatment activity is the result of a constant tension between treating patients and documenting treatments. Another result of this tension is that the incompleteness is not randomly distributed. Rather, completeness tended to increase with increasing triage level, LOS, and – most strongly – patient age. We make two conclusions from the analysis:

- The incompleteness of the activity data is substantial, probably preventing most types of reuse.
- Reusing the data incurs a bias toward patients with more severe conditions.

The former finding confirms previous research [2; 14], thereby indicating that EHRs should mainly be assessed on the basis of how well they fulfill their primary purpose of supporting day-to-day clinical work. The latter finding acknowledges that the mere existence of large quantities of real-world EHR data provides impetus for their reuse [7]. If the data are reused it is important to be aware of their limitations. The bias toward patients whose triage level, LOS, and age indicate a more severe condition extends the finding by Rusanov et al. [15] that EHRs contain more data about the sicker patients.

Part of the reason for data incompleteness is that the decision about whether to record the data is often left to the clinicians' discretion. This practice acknowledges the primacy of their day- to-day clinical work. Somewhat surprisingly it remains unclear in many studies of the quality of EHR data [e.g., 2; 12] whether it was mandatory for the clinicians to record the data. Frequently, transitional artifacts fill a gap between the clinical work and the formal documentation of it [3]. Such transitional artifacts hold procedural information important in performing the work, but at the same time the transitional artifacts are exempted from the formal documentation requirements. The whiteboard in the present study is an example of a transitional artifact. That is, the clinicians were not formally required to keep the whiteboard current. While this may contribute to explaining the incomplete data, a formal requirement to record the treatment activities will not necessarily result in complete data. For example, Granlien and Hertzum [6] found that none of eight mandated work procedures associated with an electronic medication record were followed consistently by more than 48% of the wards at the studied hospitals.

Another reason for the incomplete data is that the treatment activities had to be recorded manually. Activity data are, for example, pertinent in moving the modelling of ED crowding beyond models based solely on when patients arrive in the ED [10]. Thus, the introduction of a tool that forecasts ED crowding on the basis of activity data might motivate the clinicians to record these data more consistently. Alternatively, it might be possible to derive activity data automatically from other data. For example, the recording of a patient's triage level indicates the end of the activity of triage and is, in most cases, also a good indicator that the patient has now transitioned to the activity of waiting to be seen by a nurse. The triage level was not recorded for all patients, but in all EDs it was recorded more often than that the patient was waiting to be seen by a nurse. Thus, the requirement for manual data entry can probably be reduced by deriving additional activity data from other whiteboard data or from EHR data. While automatic data derivation will likely improve completeness [2], data correctness may suffer because manual data entry likely captures some nonsensical data. In balancing manual data entry against automatic data derivation it should also be considered that automatic data derivation frees clinician time for other activities.

Finally, the bias of the recorded data toward patients with more severe conditions is reassuring from a clinical point of view because it suggests that the clinicians attend more to the patients who are more in need of clinical attention. It may, however, be impossible to compensate for this bias when the data are reused, thereby confounding any analyses performed by reusing the data.

Limitations

Two limitations should be remembered in interpreting the results of this study. First, the data are from EDs in one healthcare region of one country. While the four EDs show that the results of the study are not peculiar to one ED, it would be valuable to replicate the study in other countries with other healthcare systems. Second, we can merely speculate about the reasons for the incomplete recording of the

treatment activities. While the log data quantify the magnitude of this incompleteness, interviews or other qualitative data would be needed to explain why the clinicians like the whiteboard but often leave the recording of the treatment activities incomplete.

Conclusion

Changes to the content of electronic ED whiteboards are logged and thereby available for later inspection and reuse. Such log data provide opportunities for forecasting ED crowding, identifying bottlenecks in the ED workflow, and – more generally – for reusing EHR data for the purposes of research and quality improvement. In this study we have analyzed the completeness with which five main treatment activities are recorded on the ED whiteboards in a Danish healthcare region. We find that the low completeness of the activity data probably prevents most types of reuse and that, if reused, the activity data incur a bias toward the patients with the more severe conditions as indicated by triage level, length of stay, and patient age. The incomplete activity data cannot be explained by the whiteboard being disliked by the ED clinicians or unused for its primary purpose of supporting day-to-day ED work. On the contrary, the whiteboard had been in successful use in the EDs for several years. It is in spite of successful primary use that the incompleteness of the activity data probably prevents most secondary uses. We point to non-mandated use and manual data entry as reasons that contribute to the poor data quality.

Acknowledgements

This study is part of the Clinical Communication project, which is a research and development collaboration between Roskilde University, University of Copenhagen, Imatis, and Region Zealand. The two interview quotes in the Method section are from interviews conducted in collaboration with Jesper Simonsen and Arnvør á Torkilsheyggi, both from Roskilde University. Special thanks are due to Rasmus Rasmussen at Imatis for making the anonymized version of the log data.

Conflicts of interest: None to declare.

References

[1] L. Brennan, M. Watson, R. Klaber, and T. Charles, The importance of knowing context of hospital episode statistics when reconfiguring the NHS, *British Medical Journal* **344** (2012), e2432.

[2] K.S. Chan, J.B. Fowles, and J.P. Weiner, Electronic health records and the reliability and validity of quality meaures: A review of the literature, *Medical Care Research and Review* **67** (2010), 503-527.

[3] Y. Chen, Documenting transitional information in EMR, in: *Proceedings of the CHI2010 Conference on Human Factors in Computing Systems*, ACM Press, New York, 2010, pp. 1787-1796.

[4] L. Flowerdew, R. Brown, S. Russ, C. Vincent, and M. Woloshynowych, Teams under pressure in the emergency

[5] J. Gordon, L.A. Sheppard, and S. Anaf, The patient experience in the emergency department: A systematic synthesis of qualitative research, *International Emergency Nursing* **18** (2010), 80-88.

[6] M.S. Granlien and M. Hertzum, Barriers to the adoption and use of an electronic medication record, *Electronic Journal of Information Systems Evaluation* **15** (2012), 216-227.

[7] W.R. Hersh, M.G. Weiner, P.J. Embi, J.R. Logan, P.R.O. Payne, E.V. Bernstam, H.P. Lehmann, G. Hripcsak, T.H. Hartzog, J.J. Cimino, and J.H. Saltz, Caveats for the use of operational electronic health record data in comparative effectiveness research, *Medical Care* **51** (2013), S30-S37.

[8] M. Hertzum, Electronic emergency-department whiteboards: A study of clinicians' expectations and experiences, *International Journal of Medical Informatics* **80** (2011), 618-630.

[9] M. Hertzum, Patterns in emergency-department arrivals and length of stay: Input for visualizations of crowding, *Ergonomics Open Journal* **9** (2016), 1-14.

[10] M. Hertzum, Forecasting hourly patient visits in the emergency department to counteract crowding, *Ergonomics Open Journal* **10** (2017), to appear.

[11] M. Hertzum and J. Simonsen, Visual overview, oral detail: The use of an emergency-department whiteboard, *International Journal of Human-Computer Studies* **82** (2015), 21-30.

[12] W.R. Hogan and M.M. Wagner, Accuracy of data in computer-based patient records, *Journal of the American Medical Informatics Association* **4** (1997), 342-355.

[13] N.R. Hoot and D. Aronsky, Systematic review of emergency department crowding: Causes, effects, and solutions, *Annals of Emergency Medicine* **52** (2008), 126- 136.

[14] A. Parsons, C. McCullough, J. Wang, and S. Shih, Validity of electronic health record-derived quality measurement for performance monitoring, *Journal of the American Medical Informatics Association* **19** (2012), 604-609.

[15] A. Rusanov, N.G. Weiskopf, S. Wang, and C. Weng, Hidden in plain sight: Bias toward sick patients when sampling patients with sufficient electronic health data for research, *BMC Medical Informatics and Decision Making* **14** (2014), 51.

[16] N.G. Weiskopf and C. Weng, Methods and dimensions of electronic health record data quality assessmernt: Enabling reuse for clinical research, *Journal of the American Medical Informatics Association* **230** (2013), 144-151.

Address for correspondence

Morten Hertzum, University of Copenhagen, Njalsgade 76, Bldg. 4, Copenhagen, Denmark, hertzum@hum.ku.dk.