Research Paper ■

Using Empiric Semantic Correlation to Interpret Temporal Assertions in Clinical Texts

GEORGE HRIPCSAK, MD, MS, NOÉMIE ELHADAD, PHD, YUEH-HSIA CHEN, MS, LI ZHOU, BMED, PHD, FRANCES P. MORRISON, MD, MPH

Abstract **Objective:** To measure the uncertainty of temporal assertions like "3 weeks ago" in clinical texts.

Design: Temporal assertions extracted from narrative clinical reports were compared to facts extracted from a structured clinical database for the same patients.

Measurements: The authors correlated the assertions and the facts to determine the dependence of the uncertainty of the assertions on the semantic and lexical properties of the assertions.

Results: The observed deviation between the stated duration and actual duration averaged about 20% of the stated deviation. Linear regression revealed that assertions about events further in the past tend to be more uncertain, smaller numeric values tend to be more uncertain (1 mo v. 30 d), and round numbers tend to be more uncertain (10 versus 11 yrs).

Conclusions: The authors empirically derived semantics behind statements of duration using "ago," and verified intuitions about how numbers are used.

■ **J Am Med Inform Assoc.** 2009;16:220–227. DOI 10.1197/jamia.M3007.

Introduction

One of the goals of natural language processing is to provide a natural and flexible mechanism for human beings to describe the world, but still produce computable information. In health care, clinicians describe patients' conditions in narrative documents. If the information were available in a computable format—structured and coded according to a defined terminology—then it could be used for retrospective clinical research, quality assurance, and automated decision support.

As shown in Fig 1, the goal is to represent the truth about a patient in a computable form. This generally involves a human being observing and interpreting the patient's condition. The human being then authors a narrative document, generally with the expectation that the document will be read by another human being. A natural language processing system must not only abstract the information that is explicitly stated in the report, but also add implicit information that the author assumed would be added by the reader. Errors may occur at any of the steps: observation and interpretation, authoring, or processing. Deviations between what is read and what is true may therefore may be due to error or due to a misinterpretation of meaning.

Affiliation of the authors: Department of Biomedical Informatics, Columbia University, New York, NY.

This work was funded by a grant from the National Library of Medicine (NLM) "Discovering and applying knowledge in clinical databases" (R01LM006910).

Correspondence: Dr. George Hripcsak, Department of Biomedical Informatics, Columbia University, College of Physicians and Surgeons, 622 West 168th Street, VC-5, New York, NY, 10032.

Received for review: 9/19/08; Accepted for publication: 12/1/08

Sometimes information that is stated in a narrative report is collected as structured facts by independent and reliable means. If these facts can be correlated with the semantic information that is abstracted from the narrative reports, then several opportunities arise. The accuracy of the processor output can be estimated. One may be able to deduce how language is used. This can lead to improved systems and accurately setting parameters according to empiric evidence rather than instinct. We have called this process "empirical semantic correlation," signifying matching the semantic information abstracted from the reports with independent empiric data.

Background

Motivation

In this paper, we present a method for the empiric semantic correlation of statements of duration in clinical texts. In particular, we are interested in statements which use the word "ago," like "the patient was discharged 3 months ago." Several examples follow:

- he was admitted six weeks ago
- recently admitted about 2 weeks ago
- the patient was recently discharged from the emergency room about 10 days ago
- her private medical doctor ordered a chest CT about 1 month ago
- history of liver transplant 8 yrs ago
- seen in emergency room 1 month ago
- patient was recently admitted several weeks ago
- mammogram 2 years ago
- liver transplantation, 2 years ago
- severe aplastic anemia diagnosed 5 years ago

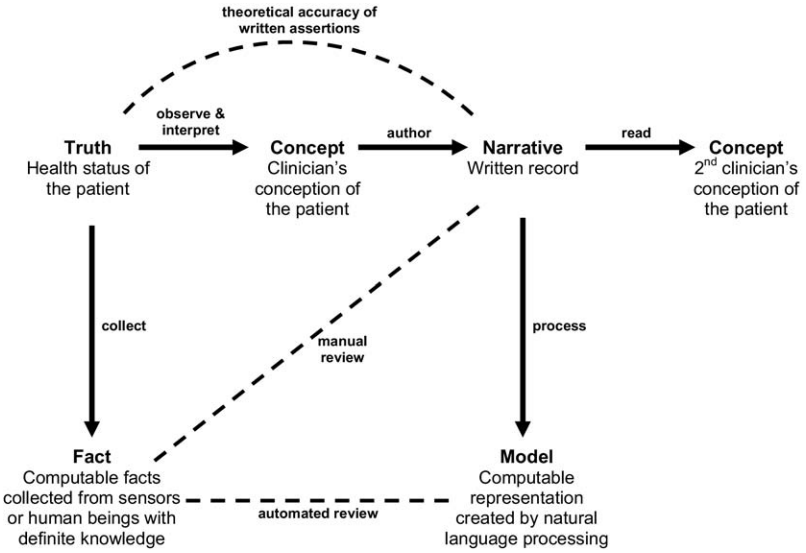

Figure 1. Framework for empiric semantic correlation.

- admission one month ago
- last electrocardiogram was one year ago

In this paper, the named event that occurred in the past is called the "referred-to event" (e.g., admitted, admitted, and discharged from the emergency room in the first three examples). The reference point from which the duration is counted back is the "anchoring event." Usually for statements that use "ago," the anchoring event is the authoring of the text, although due to cut-and-paste, the anchoring event may be admission despite the note being authored later in the hospitalization.

There are several factors affecting these durations. The events are often recounted by the patient to the clinician, and memory of when the event occurred is imperfect. The clinician develops a concept of when the event occurred and writes a shorthand version of that concept, assuming that the reader will interpret it correctly. For example, when a clinician writes "5 weeks" he or she does not generally mean 35 days, but some wider range around 5 weeks. Furthermore, whereas "11" weeks, months, or years may mean 11 U (which might be 10–12 in reality), "10" may mean one group of ten with a correspondingly wider variance (say, 5–15 in reality).

Recognizing, representing and interpreting temporal references in clinical texts automatically can benefit many applications in clinical information systems and data mining applications. Beyond health applications, our method contributes to the recent advances in processing of temporal expressions in the field of computational linguistics.

Related Work

Identifying, representing, and reasoning over temporal information is one of the classic problems of artificial intelligence.[1] Even in a specific domain like in the biomedical domain, while much research has been devoted to tagging, syntactic, and semantic parsing of medical texts, temporal reasoning with medical data are still a challenging task.[2]

In the general domain, advances have been made at a steady pace.[3,4] TIMEX[5] and TimeML,[6] a robust specification language for events and temporal expressions in natural language, provide a good framework for research in this area.[7] In the biomedical domain, emphasis has recently been placed on studying how to annotate and mark temporal information with an under-specified context.[8–10]

The following three steps can been characterized in a system that reasons over temporal information:[11] (1) recognition, that is to detect the extent of the temporal expression in text; (2) interpretation, that is to use the information from the context to turn the recognized expression into a fully specified date and time; and (3) normalization across expressions, so that further processing can be carried out smoothly.

At the semantic interpretation stage, one question, which is similar to the one we investigate in this study, is how to interpret temporal statements with an underspecified context, as in "last Friday" or "2 weeks ago." In many systems, however, the assumption is that the expression "2 weeks ago" refers to an event that happened exactly 14 days from the time the statement was made. In this study, we hypothesize that in clinical texts, such statements are not always that precise. Instead, we propose to acquire the semantics of these expressions from a corpus of statements of durations.

We focus our discussion of previous work on two papers that investigate how to automatically acquire the duration of events in an empiric fashion[12,13] (for a more detailed discussion of previous work in the field of temporal expressions, we refer the reader to a review of the field[2]). The goal of the two papers is slightly different from the one described in this study: they are concerned with the typical duration of events, such as "to finish", or "to meet", while we examine statements of durations like "2 weeks ago." Nevertheless, their goals are relevant to this work. Mani and Wellner[12] show a proof of concept for the lexical acquisition of metrical constraints from corpora. For instance, they collect occurrences of an event in a corpus of news sentences which have been annotated with temporal information automatically. The verb "to lose," for instance, appears 25 times in their corpus with different durations ranging from one day to one decade, the most frequent being "3 months" (the event refers to financial loss, which typically occurs over a trimester). They propose to turn this sample into a distribution of duration for that particular term. One big challenge with this approach is data sparseness.

Pan, Mulkar, and Hobbs[13] describe a method for learning implicit typical durations of events at a coarse level (short or long duration). They rely on a manually collected corpus of statements and typical durations. For instance, the verb "to meet" in a particular sentence was annotated by a set of annotators with lower and upper bounds on the duration of the event. The dataset is then used to train a binary classifier to categorize events into either short or long durations. Features included lexical context around the target word to classify, syntactic features as well as semantic features obtained through WordNet look-up. In our approach, we do not rely on manual annotation. Instead, we are interested in discovering from empiric data the semantic interpretation behind a temporal reference. We create an automatically annotated dataset by relying on the time stamps of the different reports that potentially anchor the events stated in the patient record.

Sullivan, Irvine, and Haas[14] carry out very similar work in the sense of correlating narrative expressions with known temporal facts. They correlated phrases like "yesterday," "this morning," and "last night" with the actual time of occurrence to determine at what point authors consider the day to have changed. For example, at 1 am, "tonight" may refer to the previous hours whereas at 10 am, it may refer to the evening to come. They plotted the frequency of use of phrases versus time of day and suggest that authors consider the night's ending to be the time of awakening or the start of daylight. Reiter and colleagues[15] carried out similar work in the domain of weather forecasting. They studied the meaning of phrases like "by late evening" and "by midday." They mapped the phrases stated by forecasters to the actual time of events and showed that forecasters differ in their use of such phrases; they then defined a standard phrase for each meaning.

Methods

We studied the use of temporal assertions in discharge summaries and compared them to events in the electronic health record.

Manual Review

We screened 5 years of discharge summaries looking for occurrences of the word "ago." We manually reviewed those occurrences looking for temporal assertions in which the anchoring event was the current admission to or discharge from the hospital; the referred-to event was a previous hospital admission, a hospital discharge, a medical procedure, or a medical test; and the referred-to event was stated as having occurred some duration in the past ("patient was discharged 3 wks ago"). The temporal assertions generally fit the following form:

$$<event>\ <duration>\ AGO$$

Where <duration> was defined as follows:

[<modifier>] <number> (DAY[S] | WEEK[S] | MONTH[S] | YEAR[S])

And where <event> specified a medical occurrence (the referred-to event) and (<modifier>) was intended to include optional modifiers like "about" and "approximately."

We then looked up the referred-to event in the electronic health record to find the actual time of occurrence of the event. We recorded the time of the anchoring event (generally the time of the authoring of the note), the stated duration time unit (day, week, months, or year), the stated duration numeric value, and the actual time of the referred-to event.

If no correlated event could be found in the electronic health record, then the occurrence was marked as such. Generally, this meant that the event occurred at another hospital for which we had no data. When several potential matching events were found, we selected the closest one to the stated time. Our goal was to analyze 50 successfully matched occurrences.

Automated Review

The manual process was tedious mainly due to the time to find a correlated event in the electronic health record, which limited the number of cases available for statistical analysis. We therefore applied an automated system to generate a larger sample and used the results of the manual process to corroborate the automated distribution. We narrowed the scope to physician statements about previous admissions. This problem was sufficiently constrained that the anchoring and referred-to events could be abstracted and matched to previous hospital visits automatically. The TimeText system[9] found temporal assertions with "ago," parsed and extracted the duration unit and value, and extracted the events. The stated time of the referred-to event was calculated and matched to the visit history of the electronic health record, selecting the closest visit in time. The visit history was obtained from the institution's patient registration system and includes the admission and discharge dates and times. The actual time of that event was recorded. Where multiple visits might match and where it was unclear whether the anchoring event was the current admission or discharge, the closest possible time was selected.

Analysis

We quantified the accuracy of temporal assertions as follows. Given the known time of the anchoring event (current admission) and the actual time of the referred-to event, we calculated the actual duration between the referred-to event and the anchoring event. Then, given the stated duration back in time when the referred-to event was supposed to have occurred and the actual duration, we calculated the deviation and proportional deviation, which were defined as follows:

$$stated\_duration = stated\_numeric\_value \times stated\_unit$$

$$deviation = stated\_duration - actual\_duration$$

$$proportional\_deviation = \frac{deviation}{stated\_duration}$$

A proportional deviation of 0 implies that the stated and actual times match perfectly. A proportional deviation of 1 implies that the referred-to event actually occurred simultaneously with the anchoring event (e.g., that a referred-to Laboratory test just occurred rather than occurring in the past), and a proportional deviation of -1 implies that the referred-to event actually occurred twice as long ago as stated in the corpus.

To aggregate the proportional deviations from several occurrences, we used the root mean square (rms) proportional deviation because we know that a proportional deviation of 0 is most desirable and rms deviation uses 0 as the under-

lying mean. We also applied the analyses using standard deviations and interquartile ranges, but these were inappropriately small when a group of occurrences had similarly large deviations in the same direction.

We used the bootstrap to calculate confidence intervals and p values.[16] Institutional review board approval was obtained for the project.

Research Hypotheses

We hypothesized the following:

1. Deviation is grossly proportional to duration. That is, proportional deviation is an appropriate first approximation.
2. The rms proportional deviation varies with how long ago the event occurred (or was supposed to have occurred), possibly because of imperfect memory.
3. After correcting for the effect of duration, assertions using larger numeric values will tend to be more accurate than those using smaller values. For example, a statement about 1-months will be less accurate than one about 30 days.
4. Among larger numeric values, those that represent round numbers, such as multiples of 5 or multiples of 6 months, will be less accurate than others. For example, when writing about a date in the past that is known only approximately, the author will tend to use round numbers.

The four hypotheses are clearly correlated because stated duration depends on both the stated numeric value and the stated unit. We therefore did exploratory analysis, correlating rms proportional deviation and the factors. We used linear regression with the absolute value of the proportional deviation as the dependent variable and transformations of stated duration (unit × value), stated unit, and stated numeric value as the independent variables. The stated unit was represented categorically using dummy variables or as a continuous variable set to the approximate number of days in the unit (e.g., 1, 7, 30, and 365).

Results

Manual Review

The domain expert analyzed 50 occurrences of ago that mapped to events stored in the electronic medical record. The absolute values of the deviations rose proportionally with duration. The Pearson correlation coefficient of the logarithm-transformed absolute value of deviation versus the logarithm-transformed duration was 0.71. That is, whether the stated duration is several days or several years, the degree to which the stated duration differs from the actual duration is roughly proportional to the duration.

Based on this result, we plot proportional deviation. Figure 2 shows the distribution of proportional deviation, stratified by the stated unit (day, week, mo, and year). "Year" was used most often (28 of 50), and "day" was used least often.[2] The rms proportional deviation was 0.24 (95% CI 0.17–0.32), the standard deviation (of the proportional deviation) was 0.25 (95% CI 0.17–0.33), and the interquartile range was 0.17 (95% CI 0.05–0.29). The average of the absolute values of the

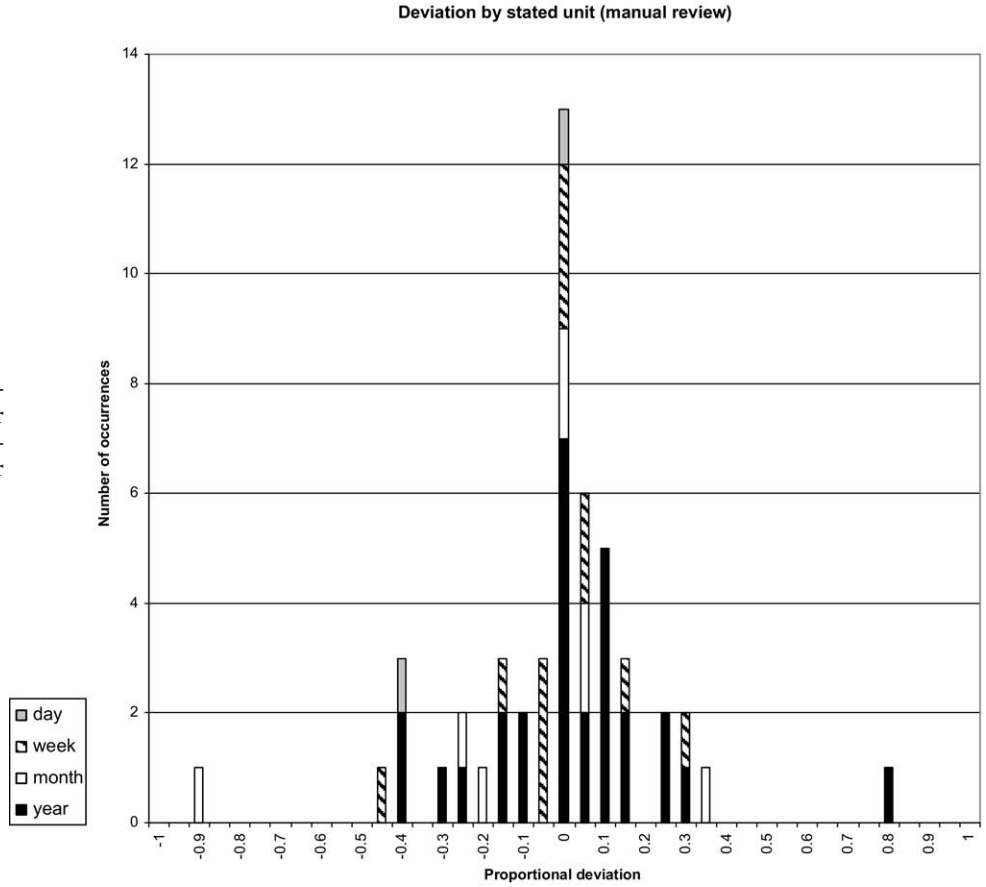

Figure 2. Distribution of deviation by unit (manual review). For the manual review data, the proportional deviation is plotted for each stated unit.

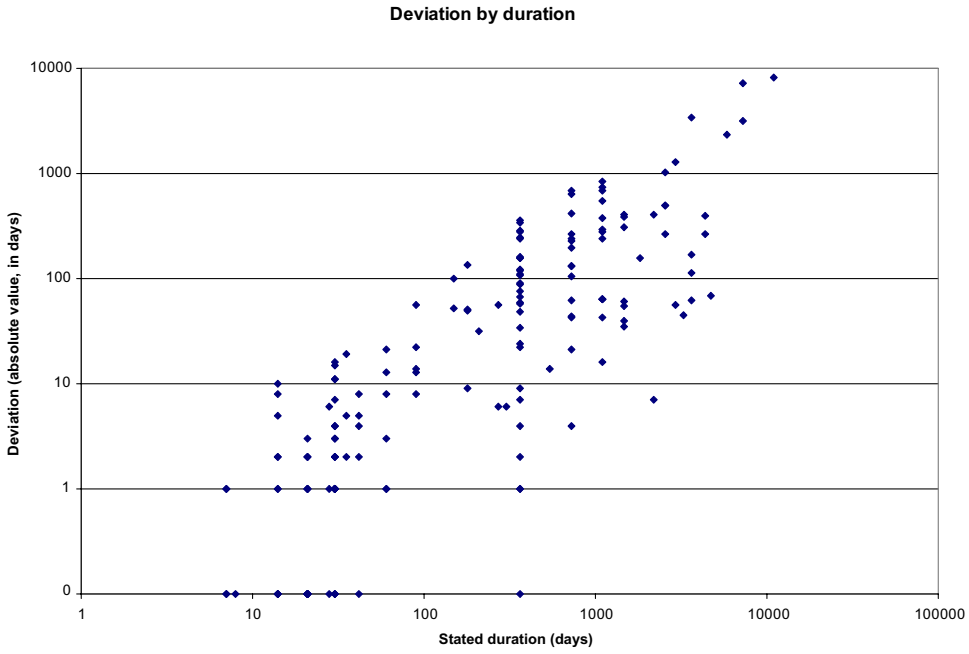

Figure 3. Distribution of deviations (automated review). For the automatically gathered data, the absolute value of the raw deviation (not proportional) measured in days is plotted against the stated duration in days, which was calculated as the stated numeric value times the size of the stated unit. Both axes are plotted on a logarithmic scale, except that a deviation of 0 is plotted where 0.1 would normally be on the y-axis. Deviation appears to be roughly proportional to stated duration.

proportional deviations was 0.16 (95% CI 0.11–0.21), and the median was 0.084 (95% CI 0.03–0.14), which corresponds to half the interquartile range of the raw sample. The sample was too small to make assertions about the dependence of proportional deviation on time and stated parameters.

Automated Review

There were 126,314 discharge summaries in the 5-year period. We found 178 occurrences of temporal assertions that were in the appropriate "ago" format (defined above), whose event referred to a previous admission, and that matched visits in the electronic health record. Figure 3 shows the absolute value of the raw deviation with respect to the stated duration on logarithmic scales. The Pearson correlation coefficient of the logarithm-transformed absolute value of deviation versus logarithm-transformed duration was 0.83. This confirms the approximate proportional relationship.

Figure 4 shows the distribution of proportional deviations, stratified by stated unit. The overall rms proportional deviation was 0.31 (95% CI 0.27–0.36), the standard deviation (of the proportional deviation) was 0.31 (95% CI 0.27–0.36), and the interquartile range was 0.19 (95% CI 0.10–0.28). The average of the absolute values of the proportional deviations was 0.20 (95% CI 0.16–0.24), and the median was 0.095 (95%

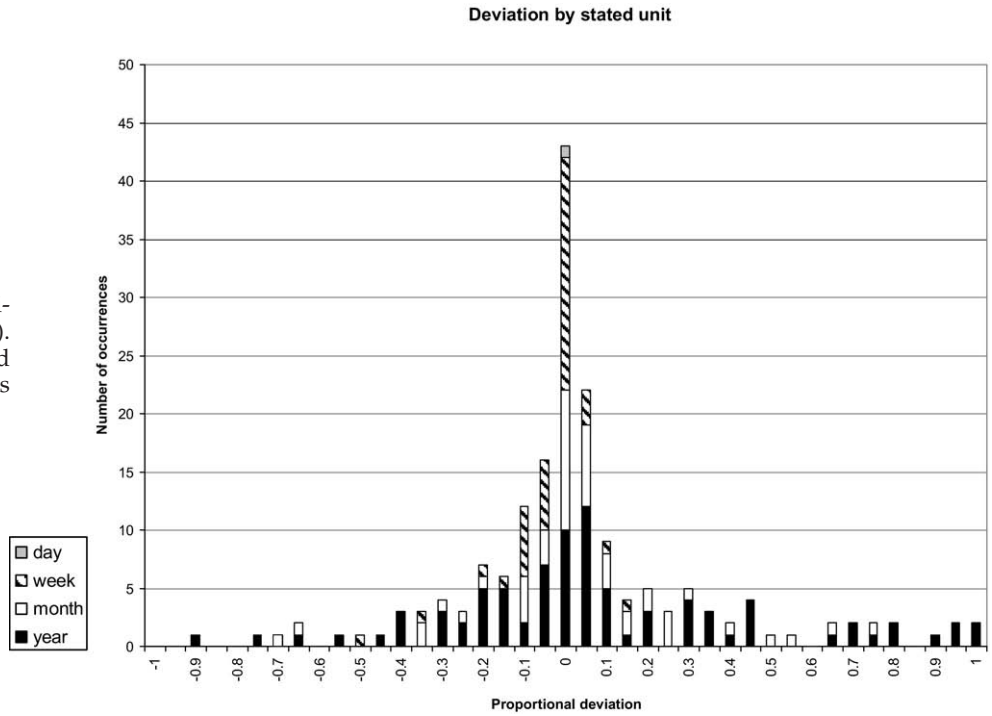

Figure 4. Distribution of deviation by unit (automated review). For the automatically gathered data, the proportional deviation is plotted for each stated unit.

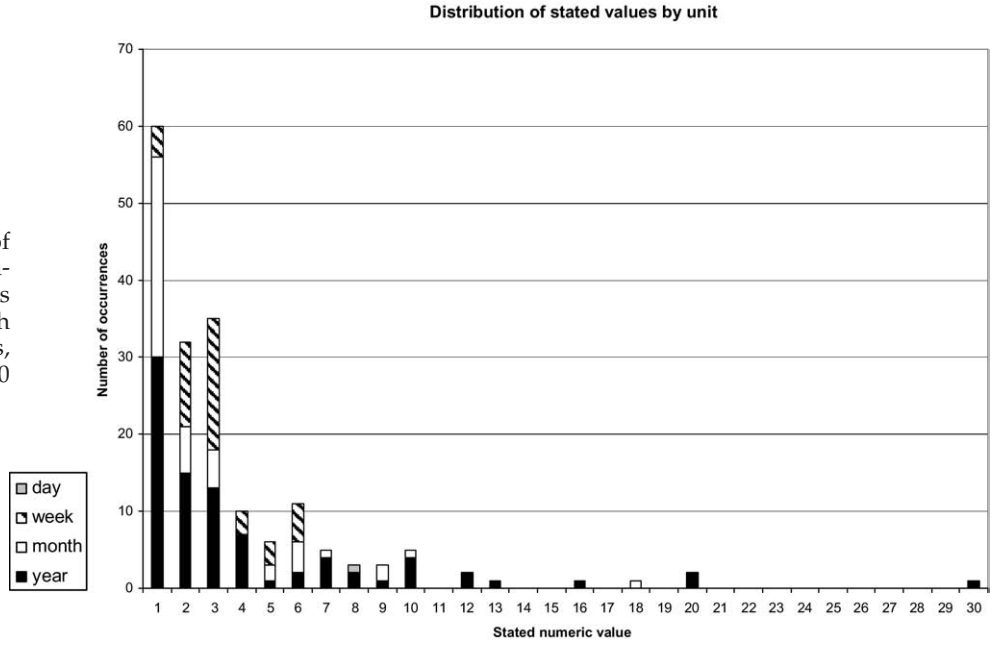

Figure 5. Distribution of stated numeric values by unit (automated review). Larger numbers tended to be associated with years, and among larger numbers, round numbers like 10, 20, and 30 were most common.

CI 0.05–0.14). Only one occurrence used a "day" unit (8 d, which was an exact match to the actual duration). This sample was limited to previous admissions, and they usually occur on a time scale longer than days.

Figure 5 shows the stated numeric values stratified by unit. Larger numbers were generally used with "year," possibly because there was no natural larger unit to switch to. Among larger numbers, round numbers like 10, 20, or 30 were most common.

Figure 6 illustrates the accuracy of the proportional relationship between deviation and duration. The rms proportional deviation is plotted against duration as calculated from the stated unit and value. If deviation were truly proportional, deviation would be flat across the plot. Instead, the figure shows that proportional deviation rises somewhat with duration. Short durations (e.g., less than 1 mo) appear to be more accurate than those of several years. The bins were chosen purposely. The third bin, "3 months to < 1 year", includes relatively longer durations but excludes years, yet it appears to be larger than the shorter duration bins; thus the effect is not simply due to the use of "years." The largest bin, "18 years or more", has enormous error—as expected—because the database only goes back 18 years and any supposed matches are likely to be mistakes.

We compared round numbers to other numbers. For occurrences in which the stated numeric value was greater than or equal to 5, the proportional deviations for values that are round numbers (multiples of five for any unit or multiples of 6 mo) were compared to other values. Occurrences outside the range of the database (over 18 yrs) were excluded. The rms proportional deviation was 0.39 (95% CI 0.24–0.55) for round numbers and 0.19 (95% CI 0.12–0.25) for other num-

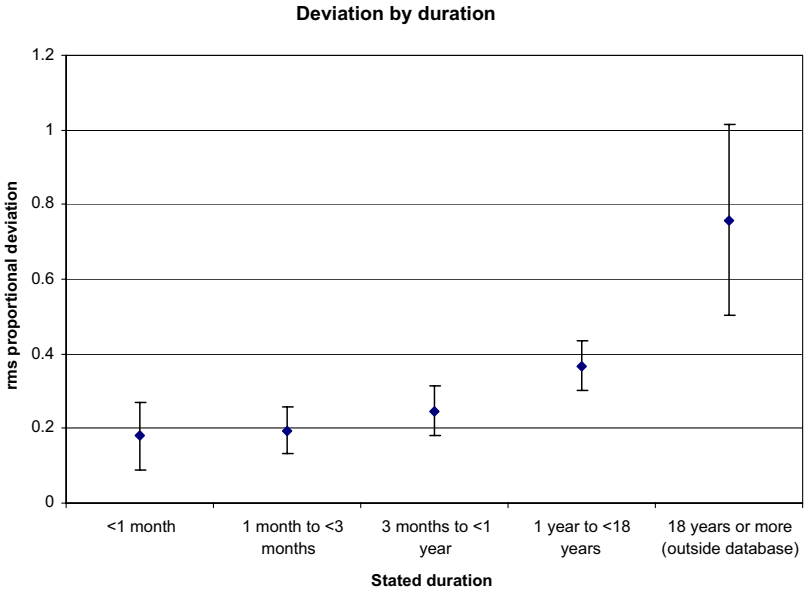

Figure 6. Proportional deviation as a function of stated duration (automated review). The rms proportional deviation (and 95% confidence intervals) is plotted against stated duration, which is calculated as the stated numeric value times the size of the stated unit. If it were truly proportional, rms proportional deviation would be flat across this plot. Instead, deviation appears to rise slightly faster than duration (i.e., the proportion rises with duration), with a large jump for assertions that lie outside of the database (18 yrs or more).

Table 1 ■ Linear Regression Model Variables

Variable	Definition	Coefficient	Significance
Value	stated numeric value in the temporal assertion (1–30 in this sample)	0.0414	<0.001
Round Number	true if value is a multiple of 5 (any unit) or 6 (with months)	−0.0218	0.002
ln(duration)	logarithm of stated duration in days, which equals the product of unit and value	0.1500	0.023
Gt 18 yrs	true if duration ≥ 18 yrs, so the event should not be in the database	0.8160	<0.001
Intercept		0.4060	0.416

bers. Thus, values that represented round numbers were less accurate than the others ($p = 0.03$) and more variable than the others. That is, it appears to be true that values like 11 are more accurate than those like 10. One would also expect greater variability in the round numbers, because 10 may mean "about one group of ten" or it may literally mean 10 U (and not 9 or 11).

The various factors—duration, stated numeric value, groupings, etc—are clearly related to one another. We therefore used linear regression to measure effects after correcting for other effects. We tested a series of linear regression models on the duration, the unit, the numeric value, the presence of a round number, and extension outside the 18-years limit of our database, including logarithm-transformed and binned versions of the continuous variables. The best fitting model was highly significant, with $p < 0.001$ and with R-squared 0.182. Table 1 contains the definitions of the variables, the coefficients, and the significance of the variables in that model. Proportional deviation increases with logarithm of the stated duration, corroborating Fig 6. After correcting for duration, larger numeric values were associated with less error. This may simply be because larger values allow a proportionally finer granularity in specifying the duration (e.g., compare 1 mo to 30 d). Among the larger values, round numbers were associated with greater deviation. Durations greater than 18 years were associated with greater deviation; these value were outside of our database (see rightmost bin in Fig 6).

We had intended to study the effect of modifiers like "about." There were too few occurrences, however, to draw conclusions. In general, temporal assertions that used "about" did not seem less accurate than unmodified statements.

Discussion

Our data supported our four hypotheses. The observed difference between the stated duration back to the event and the actual duration back to the event was roughly proportional to the duration. For example, the deviation was on average about 20% of the duration.

Several other effects were found beyond that first order approximation. Longer durations were associated with slightly larger deviations. For example, "11 days" was proportionally more accurate than "11 years." This may be a memory effect, such that events from longer ago are remembered less well. It may be that more recent events need to be stated more accurately. For example, stating that a radiograph was done 11 days ago might prompt a clinician to look for the examination under the indicated date, whereas an examination done 11 years ago may be less relevant and not require pinpointing. The automated review focused only on one event type (admissions), so it cannot be due to a bias related to the different types of events occurring at different times.

Opposing this effect is the observation that larger stated numeric values were proportionally more accurate than smaller stated numeric values. For example, "31 days" was more accurate than "1 month." This may be due to the granularity available to the temporal assertion author: Fractions were not observed in the sample, so statements about 1 month, for example, cannot be made as accurate as statements about 31 days. The author may have picked a finer unit (with a correspondingly larger numeric value) if a more accurate duration was both known and needed.

Furthermore, when the stated numeric value was a round number like 5, 10, 15, etc. or a multiple of 6 for months, it had greater deviation. For example, "11 years" was more accurate than "10 years" even after correcting for any duration or size effect. This may be because when the temporal assertion author knew the duration only approximately, the author tended to select round numbers. It was impossible to tell when "10" units meant approximately 10 units and when it meant precisely 10 units. If "11" units were specified, it is likely that the author meant 11.

One might postulate that the decreased proportional deviation observed with greater numeric values was merely because the round-number variable removed the most inaccurate occurrences. In fact, even when the round-number variable was eliminated, the former effect persisted.

These data therefore empirically confirm several intuitions about how authors use numbers and temporal durations, and they quantify the effects. The manual results (Fig 2) verify that the automated analysis is probably valid, and automated results have sufficient sample size to quantify the effects. The equivalent standard error of the deviations was generally around 0.2 of the stated duration, so a window that is plus or minus 0.4 of the stated duration will include most deviations. A more accurate window can be generated by using the results of the regression in Table 1. Some simple heuristic rules may also be used to adjust the window: increase the window for round numbers.

These results have a concrete effect on our TimeText system.[9] The system parses temporal assertions in narrative medical reports and uses a temporal constraint satisfaction problem formalism to infer the temporal relationships among events in the reports. As part of its knowledge processing component, it adds implicit temporal knowledge to the explicitly stated facts. An important example of that is broadening the temporal window around assertions. For example, the statement that an event occurred 8 weeks ago does not imply it actually occurred 56 days ago, but it is likely to have occurred within a reasonable window of that time. We had chosen to widen the duration to plus or minus

1 unit, with a minimum of half a unit when the stated number was one. When a modifier like "about" is included, TimeText uses plus or minus 50% of the stated duration.

Based on the current results, we used too small a window for unmodified assertions. They probably need to be broadened and probably need to be proportional to the stated duration. Using plus or minus 50% of the stated duration will, on average, cover about 90% of occurrences (assuming a standard deviation of 0.31). Thus, the algorithm that we use for "about" could be applied to all assertions. Alternately, we can use the factors uncovered in this study to tailor the broadening to the particular combination of numeric value and unit.

The proposed change to TimeText illustrates the importance of empiric data. Knowledge engineering based only on intuition and expert opinion may lead the researcher astray. Modern databases have the potential to improve the accuracy of natural language processing. In medicine, the emergence of large clinical databases and the opportunity to correlate structured data and narrative text may become essential. The processing may need to be automated, however, to achieve sample sizes sufficient to draw conclusions.

One of the limitations of the study was the potential error in matching events in the temporal assertion to actual occurrences in the electronic health record. For example, if a referred-to admission occurred at another hospital so that it was not in our database, then another admission may be selected erroneously, inappropriately increasing the deviation. During the manual review, if a potentially matching actual event was found to be too far from the referred-to event, the reviewer marked it as unmatched. For example, an admission last week was an unlikely match for "the patient was admitted one year ago." The automated system allowed more distant matches, and this may explain the slightly increased deviation in the automated sample. Nevertheless, the similarity of the manual and automated results serves to validate the automated processing.

We picked one particular problem for the automated analysis: the statement by a physician while caring for a patient in the hospital that the patient was previously admitted a particular duration "ago." We selected this event because it is relatively common and a relatively wide range of durations are used. Picking a single problem like this may limit the generalizability of the results, but it avoids a critical problem of confounding. If several event types are included and if different event types tend to have occurred different durations ago, then measured inaccuracies in temporal assertions may be due to the event type rather than the duration, the stated units, or the stated number of units.

Conclusions

In summary, empiric correlation of temporal assertions with facts in an electronic health record revealed useful information about how to interpret the assertions, and previous intuitions were confirmed and quantified. Stated durations back to past events deviated from the true duration in proportion with the duration. Various additional effects were found, including how long ago the event occurred and the stated numeric value. These results have direct consequences for the TimeText temporal system, and empiric correlation is likely to be useful in other contexts.

References ■

1. Allen J. Time and time again: The many ways to represent time. Int J Intell Syst 1991;6(4):341–55.
2. Zhou L, Hripcsak G. Temporal reasoning with medical data—A review with emphasis on medical natural language processing. J Biomed Inform 2007;40(2):183–202.
3. Katz G, Pustejovsky J, Schilder F, Annotating e, Reasoning. About time and events. Dagstuhl Seminar Proceedings, 2005.
4. Boguraev B, Munoz R, Pustejovsky J (eds.). Proceedings of the Workshop on Annotating and Reasoning About Time and Events, Association for Computational Linguistics (COLING-ACL), 2006.
5. Mani I, Wilson G. Robust temporal processing of news. In proceedings of the conference of the Association for Computational. Linguistics (ACL) 2000;69–76.
6. Pustejovsky J, Castano J, Ingria R, et al. Robust specification of event and temporal expressions in text. In Proceedings of the International Workshop on Computational Semantics. 2003.
7. Mani I, Pustejovsky J, Gaizauskas R (eds.). The Language of Time, Oxford University Press, 2005.
8. Sinclair G, Webber B. Marking time in developmental biology. In proceedings of the BioNLP. Workshop 2007;197–198.
9. Zhou L, Parson S, Hripcsak S. The evaluation of a temporal reasoning system in processing clinical discharge summaries. J Am Med Inform 2008;15:99–106.
10. Mowery D, Harkema H, Chapman W. Temporal annotation of clinical text. In proceedings of the BioNLP ACL. Workshop 2008;106–107.
11. Dale R, Mazur P. Local semantics in the interpretation of temporal expressions. In: Proceedings of the Workshop on Annotating and Reasoning About Time and Events, Association for Computational Linguistics (COLING-ACL), 2006, pp 9–16.
12. Mani I, Wellner B. A pilot study on acquiring metric temporal constraints for events. In: . Proceedings of the Workshop on Annotating and Reasoning About Time and Events, Association for Computational Linguistics (COLING-ACL), 2006, pp 23–9.
13. Pan F, Mulkar R, Hobbs J, Extending t. ML with typical durations of events. In: . Proceedings of the Workshop on Annotating and Reasoning About Time and Events, Association for Computational Linguistics (COLING-ACL), 2006, pp 38–45.
14. Sullivan T, Irvine A, Haas S. It's all relative: Usage of relative temporal expressions in triage notes. ASIS & T. Annu Meet 2008(AM08: Columbus, Ohio, October 24–29, 2008.
15. Reiter E, Sripada S, Hunter J, Yu J, Davy I. Choosing words in computer-generated weather forecasts. Artif Intell 2005;167: 137–69.
16. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer Verlag, 2001.