Temporal trends of hemoglobin A1c testing

Rimma Pivovarov,1 David J Albers,1 George Hripcsak,1 Jorge L Sepulveda,2 Noémie Elhadad1

ABSTRACT

Objective The study of utilization patterns can quantify potential overuse of laboratory tests and find new ways to reduce healthcare costs. We demonstrate the use of distributional analytics for comparing electronic health record (EHR) laboratory test orders across time to diagnose and quantify overutilization.

Materials and methods We looked at hemoglobin A1c (HbA1c) testing across 119,000 patients and 15 years of hospital records. We examined the patterns of HbA1c ordering before and after the publication of the 2002 American Diabetes Association guidelines for HbA1c testing. We conducted analyses to answer three questions. What are the patterns of HbA1c ordering? Do HbA1c orders follow the guidelines with respect to frequency of measurement? If not, how and why do they depart from the guidelines?

Results The raw number of HbA1c orderings has steadily increased over time, with a specific increase in low-measurement orderings (<6.5%). There is a change in ordering pattern following the 2002 guideline (p<0.001). However, by comparing ordering distributions, we found that the changes do not reflect the guidelines and rather exhibit a new practice of rapid-repeat testing. The rapid-retesting phenomenon does not follow the 2009 guidelines for diabetes diagnosis either, illustrated by a stratified HbA1c value analysis.

Discussion Results suggest HbA1c test overutilization, and contributing factors include lack of care coordination, unexpected values prompting retesting, and point-of-care tests followed by confirmatory laboratory tests.

Conclusions We present a method of comparing ordering distributions in an EHR across time as a useful diagnostic approach for identifying and assessing the trend of inappropriate use over time.

BACKGROUND AND SIGNIFICANCE

A recent report from the Institute of Medicine estimates that as much as 30% of healthcare costs in the USA are a result of unnecessary care. Finding ways to reduce unnecessary care can ease some of the healthcare cost burden without affecting the quality of patient care.1 One major contributor to excessive healthcare costs is the overordering of laboratory tests.

Laboratory test orders recorded in an institution’s electronic health record (EHR) can be analyzed to identify patterns of ordering across a large patient population, to study adherence to existing ordering guidelines, and to quantify potentially unnecessary care. This approach is especially attractive for high-volume tests, for which robust pattern analysis can be conducted and for which guidelines have been specifically constructed through detailed analysis of the latest research and expert panel discussions to maximize the test’s utility.

One frequently ordered laboratory test with specific ordering guidelines is glycated hemoglobin A1c (HbA1c). HbA1c is the measure of average blood sugar control over 6–12 weeks. The healthy range of HbA1c is between 4% and 6%, and diabetic patients have higher HbA1c values. Although diabetic classification as controlled and uncontrolled is usually determined with blood glucose measurements, it is commonly reported that the desired HbA1c level for a patient with controlled diabetes is <7%. For patients with uncontrolled diabetes, HbA1c levels often rise much higher.

Historically, HbA1c has been a standard test for the monitoring of diabetes: in 2002 the American Diabetes Association (ADA) established that patients with uncontrolled diabetes should have their HbA1c measured every 3 months, and those with controlled diabetes should have it measured every 6 months.2 New evidence suggests that HbA1c can be used for the diagnosis of diabetes as well.34 The 2009 ADA guidelines incorporated this finding and began recommending the use of HbA1c for the diagnosis of diabetes.56 These guidelines state that, if a patient has an HbA1c value of 6.5% or more for the first time, they should be retested (on a different day) to confirm the diabetes diagnosis; unless the patient exhibits clinical symptoms or has a blood glucose ≥200 mg/dL6, then no retesting is necessary. Both the presence of guidelines (both in 2002 and 2009) and the sharp distinction of how HbA1c should be ordered for monitoring and for diagnosis provide a point of comparison when analyzing patterns of HbA1c ordering.

Despite these widely publicized guidelines for diabetes care, there are numerous reports of overordering of HbA1c laboratory tests. In a study focusing on patients with newly diagnosed diabetes, HbA1c orders were analyzed over a period of 2 years.8 It was found that 8.4% of patients (N=11,003) received at least one repeat HbA1c within 30 days of their initial test, and 30.8% (N=40,162) within 90 days. A more recent 10-year retrospective analysis at a UK university hospital found that 21% of 519,664 HbA1c orders were ordered too soon (as defined by sooner than 6 months for patients with 7% HbA1c and less than 2 months for patients with 7% or over).9

Striking differences have been shown in the frequency of HbA1c orders across different healthcare settings. In a study at a Turkish university hospital, 10.3% of all 10,496 HbA1c orders over a 2-year study period were performed within less than a month of one another, and when only inpatient orders were looked at, 33.8% were found to be
ordered within less than a month. Other studies have also found inappropriate repeat testing to be more frequent in hospitalized patients.

In this paper, we focus on the overall temporal trends of HbA1c ordering across a 15-year span. Thus, we explore the ordering patterns across both inpatient and outpatient data points, as there may be an impact on ordering patterns for patients transitioning between outpatient and inpatient settings.

OBJECTIVE
In this work, we analyze HbA1c laboratory test order data over a 15-year longitudinal time scale that covers the release of two separate ADA guidelines, thereby giving us an opportunity to retrospectively study the influence of both guidelines over time. Guidelines for HbA1c provide instruction on when to measure HbA1c for different types of patients. Specifically, we study how well both the diagnosis guidelines and the monitoring guidelines are being followed at our institution.

Because of the large patient population with diabetes, HbA1c is a high-volume test and is one of the most frequently ordered tests in our institution. We comprehensively examine all of the HbA1c measurements in our clinical data warehouse, regardless of who is coded for having diabetes to generate a more complete view of the HbA1c measurement trends. We are also in position to link clinical notes to ordering patterns in order to qualitatively assess the reasons behind the ordering patterns we observe.

We ask the following research questions. (i) What are the patterns of ordering HbA1c in a large patient population? (ii) Do HbA1c orders follow guidelines with respect to frequency of measurement? (iii) If patterns of ordering do not follow guidelines, in which ways do they depart from the guidelines and what are potential explanations for the departure?

MATERIALS AND METHODS
Dataset
After obtaining appropriate institutional review board approval, we collected all HbA1c measurements in the NewYork-Presbyterian Hospital’s clinical data warehouse between January 1996 and December 2010. All measurements were included in the dataset—that is, there were no selection criteria for the patients. Each data point consisted of a tuple (patient identifier, timestamp of individual measurement, and corresponding value).

Patterns of HbA1c ordering through time
To capture total yearly ordering patterns, we looked at all HbA1c orders performed in our institution. Across all 15 years, we report counts of HbA1c orders and stratify the counts by the numerical HbA1c values.

HbA1c ordering patterns pre versus post 2002 guidelines
The 2002 ADA guidelines established that patients with controlled diabetes should be monitored every 6 months, and patients with uncontrolled diabetes should be monitored every 3 months. To verify the extent to which these guidelines had an impact on practice, we aggregated our institution’s HbA1c measurement data into two subsets: pre-guideline measurements (1996–2001) and post-guideline measurements (2003–2010).

For each year, we include only tests that are conducted within the same calendar year; therefore, tests that are conducted in December and repeated in January of the next year are not included in our results.

To visualize repeat-ordering patterns, we aggregated data in the following way. For each patient’s HbA1c time series, we calculated the days between two consecutive measurements within a year and aggregated across all patients. A histogram was created for each year, mapping the gap between consecutive measurements and number of such measurement pairs across the dataset that year. For instance, if a patient had two consecutive measurements 132 days apart and both measurements occurred in 2007, this would contribute twice to the 132-day gap for 2007.

To test whether there was a change between the pre- and post-guideline ordering patterns, we performed a two-sample Kolmogorov–Smirnov (K-S) test. The K-S test measures whether the pre- and post-guideline samples come from the same distribution. To quantify the specific differences in pre- and post-guideline periods, we performed an L1 distance calculation separately on the 0–90 day and the 91–365 day sections of the distribution. For the L1 distance calculation, we transformed the discrete measurement counts for the pre- and post-guideline periods to probability density functions (PDFs) representing the probability of every measurement gap across the time period. We quantified the difference between the pre- and post-guideline PDFs using the absolute difference between the distributions (ie, the L1 distance) defined by:

\[ d(a,b) = \int_a^b |f_1(x) - f_2(x)|dx \]

Here, \( f_1 \) and \( f_2 \) are the PDFs, and \( a, b \) represent two time points.

In addition, to visualize measurement frequency differences between patients with controlled and patients with uncontrolled diabetes, we created a density plot of the joint probability of HbA1c value and time to next measurement. This comparison is important, as ADA guidelines are defined on the basis of controlled and uncontrolled diabetes, which correspond to glucose and HbA1c levels. We hypothesize that a high HbA1c value alerts the physician to uncontrolled diabetes, thereby influencing the patient’s time to next measurement (probably closer to 3 months). Alternatively, a patient with well-controlled diabetes would probably not be tested for another 6 months.

Finally, to enable comparison with previous work that focuses on overtesting of patients with well-controlled diabetes, we used HbA1c <7.0% as a proxy measure for well-controlled diabetes and calculated how many of the total HbA1c orders each year are unnecessary (repeated within 180 days for a patient with HbA1c <7.0%).

Diagnostic use of HbA1c ordering
Considering the 2009 ADA’s recommendation for diagnostic use of HbA1c (ie, when a patient has an HbA1c of ≥6.5% for the first time, it is recommended that their HbA1c be retested on another day to confirm a diabetes diagnosis), we tracked the values of the measurements across gaps. Our hypothesis was that rapid re-measurement of HbA1c (within 10 days) is due to the occurrence of a first high value. To test this hypothesis, we looked at the proportion of rapidly retested HbA1c values that follow the diagnostic guidelines. Finally, we qualitatively assessed reasons that guideline deviations may occur by reading a sample of clinical notes of patients who had HbA1c tests repeated within 10 days.

RESULTS
Overall, our dataset consisted of 397 926 HbA1c orders, measured for 119 691 unique patients across 15 years. The maximum number of orders per patient was 150, and the
average was 3.32, with a large SD of 5.83 orders. The high variance we observed is due to the various characteristics of our EHR data. As we are not filtering our population for only diabetic patients, we have large differences between regularly monitored patients and those who were tested once for screening purposes. In addition, we have a sparse dataset of patients, some of whom may not be regularly followed-up at the outpatient clinics and others who receive their care in other institutions.

**Patterns of HbA1c ordering**
Over the 15-year period from 1996 to 2010, there was an increase in the raw number of HbA1c tests at the NewYork-Presbyterian Hospital (figure 1). This increase follows a general increase in measured patients (6232 patients who had their HbA1c measured in 1996 to 31765 patients in 2010). Over the 15-year period, the rates of HbA1c testing have remained fairly steady, between 2.09 and 2.7 tests per patient per year, with a very slight upward trend. The increase in tests with <6.5% values is consistent with the use of HbA1c for screening purposes, as most screened patients will be normal.

To adjust for diabetes screening, we report statistics for patients who have had at least two HbA1c measurements within their record. There is still a steady increase in measured patients (4434 in 1996 to 19302 in 2010) and a steady increase from 2.3 tests per patient per year in 1996 to 3.09 in 2010.

**Pre- and post-2002 guideline ordering patterns**
The two-sample K-S test showed a significant difference between the measurement gap counts in the pre- and post-guideline time periods (p<0.001). By estimating the PDFs of the measurement gap counts, we found that the measurement gap distribution changed from a fairly unimodal to a bimodal distribution (figure 2). The modality shift shows that a mostly homogeneous dataset of measurement gaps transformed into a heterogeneous dataset. We observed that the PDFs of the pre- and post-guideline measurement gaps were highly similar after 3 months, both sharply peaking at the 3-month time frame and the post-guideline curve having a slight peak at 180 days.

Noting that the two distributions look very different in the 0–90-day time frame, we performed a separate L1 distance calculation for two parts of the distributions (0–90-day gap, 91–365-day gap) to quantify the differences in the distributions. We were able to detect an order of magnitude difference in the L1 distances: L1(0–90 days)=0.0012 and L1(91–365 days)=0.00029.

We observed a sharp transition in 2002 in how patients are measured on short time scales of <90 days. This implies that how patients’ HbA1c values are measured on time scales of >3 months has not changed in 1.5 decades, whereas there was a dramatic change in how patients’ HbA1c values are measured on time scales of <3 months starting in ~2002.

In addition to measurement gaps, we evaluated how values correlated with measurement gaps in both pre- and post-2002 guideline periods. Figure 3 is a density plot representation of the joint probability of each HbA1c level and the time to next measurement for both time periods.

The density plots indicate a change in ordering after the 2002 guidelines as well. Before the guidelines were released, there is almost no correlation between HbA1c value and time to next measurement; most of the population is measured between 60 and 500 days irrespective of the patient’s value. After the 2002 guidelines, we see a much more prominent correlation between HbA1c value and time to next measurement as shown by the L-shaped curve: having a higher HbA1c value prompts quicker retesting (around 100 days), while lower HbA1c values have higher probabilities of being measured at longer time scales. If a patient’s diabetes is more controlled (towards the left of the x axis), the time to the next measurement is longer.

However, when looking specifically at unnecessary repeat tests for patients with well-controlled diabetes (repeats within 180 days for patients with HbA1c <7%), we still find that rates of inappropriate use increase over time. Between 1996 and 2000, 3.8–6% of the HbA1c tests were repeated inappropriately. After 2000, inappropriate testing rose to 11.8–19.5%, growing to over 20% in 2004 and remaining stable between 19% and 20% until 2010.

**Diagnostic use of HbA1c**
After 2002, we found growing numbers of HbA1c tests repeated within 10 days. Orders repeated within 10 days accounted for 1–2% of all HbA1c orders in 1996–2001, and grew to 2.8–3.8% after 2001. Therefore, we examined the dataset in multiple ways to...
unaware of their outpatient counterparts ordering the HbA1c.

HbA1c are admitted to hospital a few days later, and have rapid repeats. From 1996 to 2001, we appropriately retested HbA1c tests. From 1996 to 2001, we investigated whether the repeats were for justified diagnostic purposes or whether overutilization may be occurring. The 2009 guidelines for diabetes diagnosis allow HbA1c rapid retesting if the patient exhibits an HbA1c value of ≤6.5% for the first time.

Looking across all 15 years, we tested whether the patients with rapid retests meet the criteria for a diagnostic measure. Surprisingly, in 1996, 37% of retests were justified, whereas only 13% in 2010 were justified—according to the 2009 guidelines (figure 4).

We find that the rapidly repeated tests cannot be explained by adherence to guidelines. In fact, over time, a larger portion of rapid repeats is conducted on HbA1c tests with lower values even though there is no justification for any ≤10-day repeats when a value is <6.5%. In particular, beginning around the year 2000, between 40% and 50% of the rapidly repeated tests were ordered for the lowest HbA1c values (figure 5).

The overall rate of rapid retesting has grown similarly across all HbA1c values (figure 6), but, as the number of initial <6.5% orders has increased over time, so has the number of inappropriately retested HbA1c tests. From 1996 to 2001, we find that 1–2% of HbA1c tests with an initial <6.5% value are rapidly retested, and, after 2001, 2–3.6% are rapidly retested.

As a means to better understand the rapidly retested HbA1c phenomenon, we conducted a manual chart review for a subset of 100 randomly chosen patients who had HbA1c rapid repeats. A common pattern we identified was outpatients receive an HbA1c test, are admitted to hospital a few days later, and have a repeat HbA1c upon admission to the floor. This could be indicative of care-coordination issues where inpatient physicians are unaware of their outpatient counterparts ordering the HbA1c test only a few days prior (even though the laboratory results are visible in the EHR). Multiple times, consecutive outpatient visits within 10 days resulted in HbA1c being tested during several of the visits. This could also point to care-coordination issues among clinicians with different specialties (even though, once again, the laboratory results are visible in the EHR). Such care-coordination issues could lead to providers ordering a second ‘initial’ HbA1c test, without realizing that an initial test result already exists.

Another pattern we identified was a retest of HbA1c because the value was much lower than expected—specifically, an HbA1c result seemed too low given a patient’s previous history and therefore the same physician reordered the test; sometimes the clinical notes even referred to the surprisingly low first value. Finally, we found that between 5% and 10% of the rapid-repeat tests were the result of physicians conducting point-of-care HbA1c testing and ordering a confirmatory laboratory-run HbA1c test.

DISCUSSION

Distributional analysis over time of consecutive HbA1c measurement frequencies has uncovered a number of phenomena. Overall, the raw number of HbA1c tests has increased over time. This can be explained in part by the natural increasing number of patients coming to our institution for care and in part by the increasing number of patients with diabetes. In addition to the general increase in testing, we also see evidence for an increase in inappropriate use from the growing number of very short gap measurements (between 0 and 10 days) over the 15-year period.

We note that some of the measurement patterns established by the 2002 and 2009 guidelines predate their releases, although there were publications that hinted at diagnostic and monitoring recommendations before the guidelines as well. It seems that physicians have been using HbA1c as a diagnostic measure for at least 6 years before the official ADA guidelines that recommended HbA1c use for diagnostic purposes were released; we can see this by the prominent peak of short-gap measurements that begins to appear around 2003. In addition, we are able to see that, during the mid 1990s, before the 2002 guidelines specifying 3- and 6-month measurement intervals were released, there were already slight peaks at those two measurement gaps.

Overuse of HbA1c measurements: retesting within 10 days

While clinicians roughly follow the 3-month and 6-month guidelines, we find that there is a strong signal for seemingly
unnecessary repeated measurements within 10 days. The findings from this study are well aligned with recent literature about HbA1c measurement dynamics and the overutilization of HbA1c over short periods of time. For instance, Lyon et al\textsuperscript{17} also uncovered a highly prevalent short-gap peak, but their peak was at \(\sim 30\) days, not as short as the time gap in our dataset.

By comparing ordering distributions in our dataset, we find that the trend towards repeated testing over very short time periods (\(\leq 10\) days) is increasing in volume and is most common in 2010. In conjunction with a general increase in tests with HbA1c values \(<6.5\%\), we uncover a troubling trend: a growing proportion of rapid retests are conducted on tests with values \(<6.5\%\), despite ADA guidelines recommending rapid retests only on a subset of \(\geq 6.5\%\) tests. Moreover, the signal stays salient, even after controlling for the usage of HbA1c as a diagnostic tool through two tests a few days part.

On the basis of the results illustrated in figure 4, we postulate that 65–85\% of the HbA1c repeat measurements occurring within 0–10 days are unnecessarily ordered—a total of 9491 tests from 1996–2010. The 9491 tests represent an HbA1c repeat test overutilization rate of 2.3\% across the entire 15-year period. We use the Medicare reimbursement of US$13.24 per test to estimate an unnecessary expenditure of \(\sim\)US$125 600 at just one hospital in New York City. This figure does not account for personnel, laboratory time, ordering or interpretation time, any patient care costs, or over-ordered laboratory tests as a result of transferred patients and a lack of health information exchange, which has also been shown to contribute to laboratory test overutilization.\textsuperscript{18}

There are many potential explanations for the increasing rate of overuse of HbA1c tests. The addition of HbA1c to preset laboratory order panels may lead to the retesting, as often it is more efficient for the clinician to order an entire panel rather than remembering to exclude the HbA1c test. Another reason for this trend could be a consequence of guideline-induced over-vigilance; providers are intent on following the monitoring HbA1c guidelines but do not always remember to check whether an HbA1c result has been recently recorded.

**Impact on informatics research**

As the number of EHRs across the country increases, there is a growing potential for pertinent and effective information technology interventions to help optimize healthcare resources, as well as to ensure that that clinicians adhere more closely to national guidelines for testing.\textsuperscript{19} For instance, identifying rapid retests of HbA1c as a strong pattern in an institution, the EHR could now implement a module that denies rapid retest of HbA1c without appropriate reasoning from the ordering physician, and perhaps require a phone call to the laboratory to verify the need for a second test. Alternatively, there could be a systematic review and removal of HbA1c from laboratory test order sets.

Our work on trends in measurement gaps is relevant to large-scale data analysis work in informatics as well. To properly study large patient populations (in our case, across the entire EHR), it is helpful to stratify or decompose the population into homogeneous data points.\textsuperscript{20} In the pre-2002 guideline period, the distribution of HbA1c data points by gap is mostly unimodal, but the post-guideline period shows that the distribution changes towards a bimodal distribution. Thus, the population of laboratory measurements is decomposable according to different reasons, some health-related (measurements 3 months apart for patients with uncontrolled diabetes), but some that might not be health-related (measurements <10 days apart). The ability to
quantify and recognize decomposable and non-decomposable distributions of data points when performing large-scale research on the EHR is critical to ensure precise and robust inference.

Limitations
A limitation of this work is that the data are derived from a single institution. Nevertheless, the findings we report are consistent with previous work,8 9 and the distributional analytics methodology we employ is generalizable to other EHR systems. We also recognize that the truncation of the yearly data on December 31st leads to missing tests, but as we perform the truncation for all years, we are able to create datasets that can be compared over time and used to discover temporal trends.

Future work
The chart review suggested potential reasons for unnecessary reordering, but we emphasize the need for a more in-depth understanding of this phenomenon. We hope to achieve this with a mix of further quantitative analyses, interviews with practicing endocrinologists and primary providers, and more detailed chart reviews that include factors known to contribute to inappropriate ordering of laboratory tests such as length of stay in an inpatient setting and patient age.21

In future work, we hope to augment and extend this research to HbA1c tests performed all over New York using the Health Information Exchange Network, once the hurdles to enable research studies with these data are cleared and more longitudinal data have been collected.

In addition, once we complete more in-depth analyses on why rapid retesting may be occurring, we are interested in using the results of this research to inform interventions. In conjunction with the pathology department, we will explore options for deploying informatics interventions through the EHR in an effort to restrict HbA1c rapid reorder and reduce the unnecessary overutilization of HbA1c laboratory testing.

CONCLUSION
In this paper, we provided a methodology for studying utilization patterns that serves as a useful diagnostic approach for identifying trends of inappropriate laboratory test use over time. We analyzed the laboratory measurements of HbA1c for all the patients in our institution, over 15 years. Our study replicates prior work on HbA1c overutilization and offers new insight into the trends of laboratory ordering over time, in particular pre- and post-guidelines for diabetes monitoring, correlations between HbA1c value and testing frequency, and the use of HbA1c as a diabetes diagnostic tool. With the number of diabetic patients expected to continue to grow,22 it is essential to identify the ways in which HbA1c ordering is misused. Our study contributes to this effort.

More generally, with the interest in detecting redundant and inappropriate laboratory test utilization,23 24 the methodology presented in this paper can easily be replicated to other high-volume tests. As more data-derived patterns of use are detected across laboratories, we can start to understand better key aspects of clinicians’ workflow and what informatics solutions can be put in place to support clinicians for quality care, all while ensuring cost effectiveness.

Acknowledgements The authors would like to thank Dr Adler Perotte and Dr Hojjat Salmadian for their helpful discussions, as well as two anonymous reviewers for their valuable comments on earlier versions of this paper.

Contributors RP, DJA, GH, and NE came up with the experiment. RP, DJA, and JLS performed much of the data analysis. GH and JLS helped with the results interpretation. RP and NE wrote the paper.

Funding This work is supported by National Science Foundation IGERT #1144854, National Library of Medicine award R01 LM06910 (to GH), and National Library of Medicine award R01 LM010027 (to NE).

Competing interests None.

Ethics approval Columbia Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES


