

Five-Year Two-Center Retrospective Comparison of Central Laboratory Glucose to GEM 4000 and ABL 800 Blood Glucose: Demonstrating the (In)adequacy of Blood Gas Glucose

Journal of Diabetes Science and Technology
2020, Vol. 14(3) 535–545
© 2019 Diabetes Technology Society



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1932296819883260
journals.sagepub.com/home/dst



George Cembrowski, MD, PhD^{1,2} , Joanna Jung, PhD¹,
Junyi Mei, BSc³, Eric Xu, MD³, Tihomir Curic, BS⁴,
RT Noel Gibney, MB⁵, Michael Jacka, MD, MBA⁵,
and Hossein Sadrzadeh, PhD^{4,6}

Abstract

Purpose: To evaluate the glucose assays of two blood gas analyzers (BGAs) in intensive care unit (ICU) patients by comparing ICU BGA glucoses to central laboratory (CL) glucoses of almost simultaneously drawn specimens.

Methods: Data repositories provided five years of ICU BGA glucoses and contemporaneously drawn CL glucoses from a Calgary, Alberta ICU equipped with IL GEM 4000 and CL Roche Cobas 8000-C702, and an Edmonton, Alberta ICU equipped with Radiometer ABL 800 and CL Beckman-Coulter DxC. Blood glucose analyzer and CL glucose differences were evaluated if they were both drawn either within ± 15 or ± 5 minutes. Glucose differences were assessed graphically and quantitatively with simple run charts and the surveillance error grid (SEG) and quantitatively with the 2016 Food and Drug Administration guidance document, with ISO 15197 and SEG statistical summaries. As the GEM glucose exhibits diurnal variation, CL-arterial blood gas (ABG) differences were evaluated according to time of day.

Results: Compared to the GEM glucoses measured between 0200 and 0800, the run charts of (GEM-CL) glucose demonstrate significant outliers between 0800 and 0200 which are identified as moderate to severe clinical outliers by SEG analysis ($P < .002$ and $P < .0005$ for 5- and 15-minute intervals). Over the entire 24-hour period, the rates of moderate to severe glucose clinical outliers are 3.5/1000 (GEM) and 0.6/1000 glucoses (ABL), respectively, using the 15-minute interval ($P < .0001$).

Discussion: The GEM ABG glucose is associated with a higher frequency of moderate to severe glucose clinical outliers, especially between 0800 and 0200, increased CL testing and higher average patient glucoses.

Keywords

blood gas analyzer, glucose accuracy, GEM 4000, ICU, Radiometer ABL 800, surveillance error grid

Introduction

In 2001, Van den Berghe et al demonstrated that surgical adult intensive care unit (ICU) patients who had their blood glucose maintained between 80 and 110 mg/dL experienced significantly less morbidity and mortality than patients with glucoses between 144 and 180 mg/dL.¹ While most hospitals used blood glucose meters (BGMs) to monitor and control glucose, Van den Berghe's patients received more accurate glucose testing from an ICU-based blood gas analyzer (BGA). The University of Alberta Hospital has been using ICU-based, respiratory therapist-operated blood gas instruments for over 25 years and we have used the same systems

¹Laboratory Medicine and Pathology, University of Alberta, Edmonton, AB, Canada

²CC Quality Control Consulting, Laboratory Concision, Edmonton, AB, Canada

³College of Medicine, University of Manitoba, Winnipeg, MB, Canada

⁴Calgary Laboratory Services, AB, Canada

⁵Critical Care, School of Medicine, University of Alberta, Edmonton, AB, Canada

⁶Cumming School Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:

George Cembrowski, MD, PhD, Laboratory Medicine and Pathology,
University of Alberta, Edmonton, AB, Canada T5N 3M7.
Email: cemr001@gmail.com

to assess newer glucose measurement technologies.² We have long maintained that BGA glucose measurements are sufficiently accurate for critical care and were not surprised that Normoglycemia in Intensive Care Evaluation-Survival Using Glucose Algorithm Regulation (NICE-SUGAR)³ did not replicate Van den Berghe's findings. It seemed that glucose measurement technology was largely ignored by the NICE-SUGAR investigators as their input forms only required glucose level, but not methodology. Many hospitals that participated in NICE-SUGAR (including ours) used hematocrit-sensitive BGMs⁴ which provided factitiously elevated glucoses in anemic patients and predisposed to iatrogenic hypoglycemia.⁵ Gunst and Van den Berghe⁶ maintain tight glycemic control was too rapidly and broadly implemented without the implementation of five critical factors including highly accurate BGMs or ICU-based BGA for the analysis of arterial (not capillary) blood.

A 2013 systematic review of blood glucose accuracy showed that the accuracy of arterial blood-glucose measurements was superior to capillary measurements and that BGA arterial measurements were more accurate than BGM using capillary blood and tended to be more accurate than BGM measurements of arterial blood.⁷ The authors stated that blood-glucose monitoring was less accurate within or near the hypoglycemic range, especially in patients with unstable hemodynamics or receiving insulin infusions. They concluded: "we should be aware that current blood glucose-monitoring technology has not reached a high enough degree of accuracy and reliability to lead to appropriate glucose control in critically ill patients."

In 2008, we attempted to indirectly compare the accuracy of BGA and BGM glucose analyses in the General Systems ICU at the University of Alberta Hospital.⁸ Our approach measured the short-term variation of series of inpatient glucose measurements. The resulting statistic summarized the magnitude of the methodology's analytical and subject's biological variation as well as any intervening pre-analytic variation. We have termed this statistic PreAnalytic (including biologic variation) and ANalytic variation (PAAN). While PAAN was lower in patients who had their blood glucose measured by BGA than BGM, we discovered that patients with dysglycemia tended to have their glucose monitored by BGM. In the same ICU, patients not expected to have large glucose excursions had their glucoses measured by the BGA. As a result of this selection bias in glucose measurement, the biologic/analytic variation differences were not readily generalizable.

For the last decade, we have been assessing PAAN in cohorts of patients with frequently repeated tests and have discovered that all BGAs do not deliver equivalent analytic accuracy. In 2017, we discovered an imprecision issue, associated with elevated PAAN, in virtually all of the analytes measured by tandem GEM 4000 BGAs. We theorize that the GEM's superior morning agreement and higher afternoon imprecisions are related to the analysis of Process Control

Solution C that is analyzed at 0200. It is the only GEM process control solution that is run every 24 hours and suspicions increase given the fact that it is generally run at 0200, concurrent with the beginning of the more accurate testing. Once this solution is run, the imprecision of the analyzers drops to highly acceptable levels and both GEMs produce concordant results for the next few hours; subsequently, the imprecision increases,⁹ with the imprecision's effects amplified when multiple analyzers analyze serial inpatient samples.^{1,2}

The North Star for glucose measurement improvement has been the ever-tightening consensus guidelines for glucose measurement accuracy. In 2013, ISO 15197 revised its acceptability limits for BGMs,¹⁰ and in late 2016, the US Food and Drug Administration (FDA) issued new guidance criteria for evaluating BGM analytical accuracy in hospital environments.¹¹ The ISO and FDA accuracy requirements supplement successive generations of clinician-designed error grid diagrams which has morphed from Clarke's 1987,¹² to Parke's 2000¹³ to Klonoff's 2015 three-dimensional surveillance error grid¹⁴ (SEG) that typically would not be used in initial laboratory and regulatory evaluations but in postmarketing quality improvement exercises. The reference glucose values are plotted on the *x*-axis, the test system on the *y* axis, and an overlying almost continuous spectrum of color-coded discrete points which can be translated into eight different levels of acute clinical risk. The SEG is based on surveys of 206 diabetes clinicians and provides nearly concordant results for pediatrics and adult type 1 and type 2 patients and diverse clinician groups including US and non-US physicians.¹⁵ In the development of the SEG, minimum analytic specifications were derived for the reference glucose analyzer, limits that are readily achievable by today's larger central laboratory (CL) analyzers.

Until the recent ISO and FDA revisions, BGA evaluations of glucose accuracy were highly formulaic and much less purposeful than evaluations of CL chemistry analyzers. A more interesting and more relevant approach was employed by Liang and Wanderer¹⁶ who retrospectively evaluated the glucose accuracy of the GEM 4000 and GEM 3000 BGA at Vanderbilt University Medical Center. Paired CL and GEM BGA glucose results were extracted from a database if the CL glucose collection time was within five minutes of the BGA glucose collection in the same patient, and both CL and BGA tests were completed within one hour. Central laboratory and BGA samples were arterial or venous. Glucose levels were excluded from analysis if the BGA demonstrated significant hyponatremia, hypokalemia, or anemia, consistent with specimen dilution. The authors demonstrate acceptability with 2013 ISO and near acceptability with the 2016 FDA guidance. In this study, we follow their approach but have expanded the data inclusion criteria. We examine the glucose accuracy of two point of care BGA: tandem GEM 4000s and tandem ABL 800s that are operated by respiratory technologists and located in or just outside ICUs in two university hospitals in Alberta, Canada.

Table 1. Number of Central Laboratory and Blood Glucose Analyzer Glucoses Compared With 60-minute, 15-minute, and 5-minute Intervals as well as the Total Number of Venous Blood Gas and Arterial Blood Gas Glucoses That Were Presumably Compared to Either Arterial or Venous Central Laboratory Glucose.

Datasets	GEM				ABL			
	ABG	VBG	Other	Total	ABG	VBG	Other	Total
Original dataset (drawn within 60 min)	24 553	3244	55	27 852	8105	1887	0	9992
15-min interval	-	-	-	12 721	-	-	0	7919
5-min interval	-	-	-	7223	-	-	0	4217

Abbreviations: ABG, arterial blood gas; VBG, venous blood gas.

Materials and Methods

An Alberta Health Services data repository provided five years of ICU BGA glucose levels (May 1, 2012 to April 30, 2017) and matching CL glucose levels if the CL samples were drawn within 60 minutes of the BGA glucose. The ICU data originated from two adult ICU units, a 30-bed unit at the Calgary Foothills Medical Centre (766 bed quaternary care center) and a 32-bed unit at Edmonton's University of Alberta Hospital (885 bed quaternary care center). Two Instrumentation Laboratory GEM 4000s provided blood gas, electrolyte, and metabolite testing at the Foothills ICU, and two ABL 800s provided similar testing at the University of Alberta Hospital (UAH) ICU.

The Roche AccuChek Inform II BGM was available for glucose monitoring in both ICUs. Over most of the study duration, typical glucose ranges in Edmonton and Calgary were between 110 and 180 mg/dL although different ranges were sometimes used for individual patients. The Edmonton and Calgary CLs analyzed plasma; as such, blood specimens were processed and analyzed expeditiously (99 percentile CL turn-around times of 59 and 45 minutes compared to BGA turn-around times of 26 and 57 minutes, respectively). Initially, in Calgary, the CL instrument was the Roche Cobas 6000 c 501 which was replaced by the 8000-C702 (hexokinase enzymatic assay, traceable to an isotopic dilution mass spectrometry standard with a coefficient of variation [CV] of about 1.7%); in Edmonton, CL glucoses were generated by tandem Beckman-Coulter DxC (glucose oxidase National Institute of Standards and Technology aligned [917c and 965b] method with a CV of 3%). Table 1 identifies the mix of patient specimens whose BGA glucose accuracies were evaluated. As in Liang's paper, the CL or BGA samples were either arterial or venous.

We posited that intervals between sampling for CL and BGA glucoses exceeding 20 or 30 minutes would cause measurable differences between the CL and BGA glucoses, through both in vivo and ex vivo factors, eg, glucose-modifying therapies and glycolysis, respectively. To determine the maximum sampling interval that these differences would not be detected, we determined for both the ABL 800 and GEM 4000 the mean absolute deviations (MADs) between the POC and CL glucoses for increasing sampling intervals:

$$\text{MAD} = \Sigma |\text{BGA glucose} - \text{CL glucose}| / N$$

Because of the GEM 4000's diurnal imprecision variation, we opted to use the ABL's interval of MAD stability to define the maximum interval between CL and BGA sampling. Using this time interval, we assembled appropriate pairs of BGA glucoses and companion CL glucoses. We also used Liang's five-minute maximum interval to assemble smaller GEM and ABL datasets. We associated each BGA glucose measurement with its within-day time of blood draw, expressed as hours, ranging from 0 to 24 hours. For each of the 24-hour GEM (15- or 5-minute) datasets, we created two data subsets, the early morning 6-hour presumably stable period of the GEM (0200-0759) and the 18-hour less stable period (0800-0159). We generated analogous files for the ABL; analysis of the ABL data provides an essential perspective for interpreting the GEM data.

The datasets were then evaluated with (1) specialized Klonoff SEG software that provided bias, mean absolute relative difference (between the CL and BGA glucoses) (MARD) and CV calculations,¹⁷ (2) FDA 2016 guidance criteria for POC glucose analyzers, and (3) 2013 ISO 15197 guidance document. To visually demonstrate the time-dependent relation of glucose variation, time-ordered (by each fractional hour of the test) graphs of the BGA-CL differences were graphed along with running CVs (running CV calculated from the relative standard deviation of the last 1000 relative differences). To provide an intuitive and relevant clinical interpretation of the GEM and BGA variation, the Klonoff SEG was determined for each of the 12 datasets. The chi-square (χ^2) statistic was used to determine whether the prevalence of the nonslight risk SEG outliers (moderate plus severe plus extreme clinical outliers) differed between the ABL and the GEM for early morning, late morning to evening, and the entire day. Additionally, χ^2 was used to determine *within-analyzer* prevalence of the nonslight risk SEG outliers' distributions between early morning and late morning to evening.

Results

Figure 1 contrasts the ABL and GEM MAD graphs. We chose 15 minutes as the maximum interval to compare CL and BGA glucoses as the Radiometer MAD graph is relatively stable up to 15 minutes (mean= 6.1 [$s = 0.52$] mg/dL,

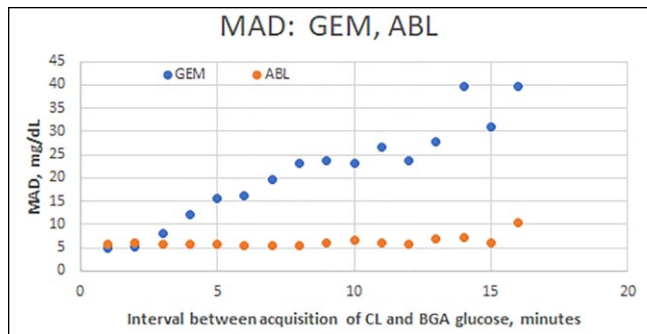


Figure 1. Mean absolute deviation graphs for blood glucose analyzer-central laboratory glucose differences vs time for University of Alberta Hospital (UAH) radiometer ABL intensive care unit patients and University of Calgary (UC) Foothills GEM intensive care unit patients.

7920 CL-BGA pairs). The GEM MAD graph is very different and starts increasing after two minutes with the five-minute MADD being three times the two-minute value. Table 2 presents the composition and statistical summary of 12 sets of the GEM and ABL blood gas data. The GEM glucose demonstrates a negative bias of approximately 2% with a hypothetical population mean of 150 mg/dL (Liang documented a bias of -3 mg/dL but did not offer a patient mean, repeat Liang reference). The ABL's positive bias of 2.5% has been attributed to glucose measurement in a whole blood matrix.¹⁸ The MARD and CV (standard deviation of the relative difference) are relevant as they are indicators of random error (imprecision). Apart from the five-minute interval early morning (0200-0800) set, the GEM CV is higher than that of the ABL, especially for the 15-minute interval 24-hour dataset. With respect to the ISO accuracy assessments, the accuracy for gluces exceeding 100 mg/dL is acceptable for the ABL and mostly acceptable for the GEM (the 15-minute nonearly morning subset is deficient). For gluces under 100 mg/dL, the GEM demonstrates more than 5% inaccuracy in the 15-minute interval datasets and especially in the nonearly morning subset. With regard to the FDA guidance for gluces exceeding 75 mg/dL, the GEM demonstrates more than 5% inaccuracy for the 15-minute interval 24-hour set and the 5-minute interval nonearly morning subset. With respect to the FDA hypoglycemic (<75 mg/dL) challenges, both GEM and ABL did poorly.

Figure 2 shows the changes in the dispersion of the BGA-CL glucose differences and running CV over 24 hours of BGA-CL differences. For the GEM, both in the 5- and 15-minute graphs, the 0200 outliers diminish in frequency and the running CV is seen dropping just after the first few visible points and reaches a nadir between 0400 and 0600 as indicated by the two-hour gray triangles. By 0800, the CV begins to increase and is accompanied by a persistence of outliers until 0200. There is dampening in the variation cycle viewed with five-minute interval lens. Review of the

ABL-CL graphs shows a small increase in outliers between 0600 and 0800 with a diminution at 1800. We determined that 0530 and 1730 are typical times for recalibrating the ABL.

Figures 3-5 present the SEG for the 24-hour sets of 5-minute interval and 15-minute interval BGA-CL comparisons, the 6-hour early morning subsets and the 18-hour late morning to evening subsets, respectively. Only the 15-minute GEM 0800 to 0200 dataset did not meet "BGM Surveillance Study Accuracy Standard," demonstrating 92.8% compliance. The SEG summary of the clinical outliers is provided in Table 3. The last column of Table 3 contains combined frequencies of the moderate, severe, and extreme clinical outliers expressed in outliers per 1000 reported gluces, to mitigate the effects of data rounding and to reinforce the importance of outliers in terms of total quality management.¹⁹ A single GEM "extreme risk" error was identified. The χ^2 analysis of the distribution of the nonslight risk outliers, summarized in the captions of Figures 3-5, demonstrates statistically significant differences between the GEM and ABL outlier prevalence. The prevalence of the nonslight risk outliers were significantly different between the GEM early morning gluces compared to the GEM late morning to late evening gluces ($P < .002$ and $P < .0005$ for 5- and 15-minute intervals, respectively). For the ABL, there was no difference in early morning and late morning and evening outlier distributions.

In Figure 6, Pareto charts provide in decreasing order the volumes of CL test usage per two-hour intervals. The top set of graphs was constructed for BGA glucose specimens and CL gluces being drawn within 60 minutes; the other two correspond to BGA and CL drawn within 15 and 5 minutes of each other. Inspection of these graphs shows that using either the 5- or 15-minute interval, 65% of the GEM-CL and 75% of the ABL-CL differences will be derived from the 0200 to 0800 and 0000 to 0600 intervals for the GEM and ABL, respectively, thus reducing the proportional contribution of observations from outside these hours. Compared to the ABL ICU, many more gluces are sent to the CL from the GEM ICU. Using χ^2 analysis and the 0200 to 0800 and 0000 to 0600 testing volumes as baselines for follow-up testing in the remaining 18 hours for the GEM and ABL, respectively, the proportions of follow-up testing are higher in the GEM (0.51-0.66) compared to the ABL (0.32-0.41, $P < 10^{-6}$).

Discussion

Drs Liang and Wanderer inspired us to broaden their BGA fitness study and make it more inclusive in terms of time interval between the acquisition of the BGA and CL samples. Selection of a five-minute interval for glucose pairs is safe in that its universe of glucose tests will include presumably pre-scheduled paired specimens which are likely to be run in the

Table 2. Descriptive Statistics (Bias, Mean Absolute Relative Difference, Coefficient of Variation, and Glucose Mean Obtained From Klonoff Surveillance Error Grid Calculator Repeat Kovatchev¹⁷) and Compliance to ISO Standards and Food and Drug Administration Guidance.

Blood gas analyzer	Measurement period (h)	Glucose comparison interval (min)	Comparisons	Bias (%)	MARD (%)	CV (%)	Glucose mean (mg/dL)	ISO (<5%) >+15 mg/dL (<100 mg/dL (%))	ISO (<5%) >+15% >100 mg/dL (%)	ISO (<5%) >+20% >100 mg/dL (%)	FDA (<5%) >12 mg/dL <75 mg/dL (%)	FDA (<5%) >15 mg/dL >75 mg/dL (%)	FDA (<2%) >15 mg/dL <75 mg/dL (%)	FDA (<5%) >12% >75 mg/dL (%)	FDA (<2%) >15% >75mg/dL (%)
GEM	24	15	12 721	-1.1	5.0	16.1	153.0	6.4	3.5	2.3	6.7	14.1	6.5	4.1	4.4
GEM	24	5	7223	-1.1	4.4	10.6	152.3	4.2	2.4	1.7	9.8	9.8	4.1	4.1	1.5
ABL	24	15	7919	2.4	4.2	6.4	144.0	7.4	1.4	0.6	13.3	9.3	3.8	3.8	2.0
ABL	24	5	4217	2.6	4.1	5.6	143.5	4.7	1.2	0.5	14.3	10.4	4.0	4.0	2.0
GEM	18 (0800-0159)	15	4550	-1.4	6.3%	10.6%	159.1	8.8	6.8	4.8	13.4	13.4	9.4	9.4	7.0
GEM	18 (0800-0159)	5	2442	-1.5	5.3	10.6	158.7	5.0	4.2	3.0	11.6	8.7	5.9	5.9	4.2
ABL	18 (0800-0159)	15	2292	2.5	4.5	6.7	151.1	6.6	2.1	1.1	18.6	12.9	4.8	4.8	2.7
ABL	18 (0800-0159)	5	1164	2.5	4.1	6.2	150.1	7.4	2.2	0.8-	32.3	16.1	4.0	4.0	2.1
GEM	6 (0200-0759)	15	8164	-0.9	4.2	10.7	149.6	4.4	2.3	1.4	14.1	14.1	4.1	4.1	2.5
GEM	6 (0200-0759)	5	4778	-1.0	3.9	7.7	149.1	3.5	1.6	1.0	11.6	11.6	3.2	3.2	1.7
ABL	6 (0200-0759)	15	5627	2.4	4.1	6.3	141.1	3.2	0.4	0.5	8.8	6.2	3.4	3.4	1.7
ABL	6 (0200-0759)	5	3054	2.6	4.1	5.3	141.1	3.8	0.9	0.3	8.5	6.4	3.9	3.9	1.9

Abbreviations: CV, coefficient of variation; FDA, food and drug administration; MARD, mean absolute relative difference.

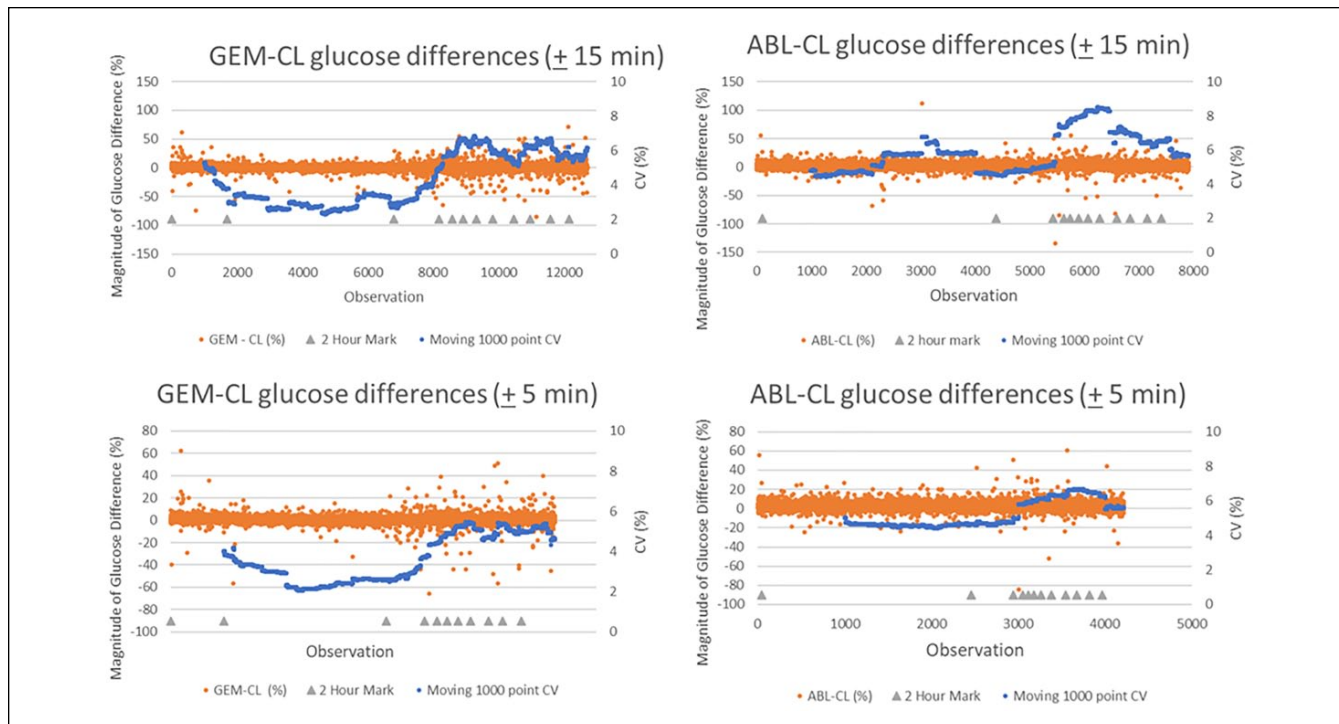


Figure 2. All arterial blood gas-central laboratory differences (red) have been ordered by their occurrence over 24 hours for the complete 5- and 15-minute interval GEM and ABL datasets. The graphs start at 0200 (indicated by first gray triangle) and end 24 hours later. Each succeeding triangle indicates a two-hour increment. The blue line represents the moving CV and is derived from the relative standard deviation of the last 1000 arterial blood gas-central laboratory differences.

early morning during the GEM's analytically stable period. We chose a 15-minute interval as it was not associated with an increase of the MAD as measured by the ABL but would probably incorporate some of CL glucoses that were ordered to clarify any outlying BGA glucose. The Pareto charts of Figure 6 indicate that many CL glucoses are ordered beyond the five-minute interval. Our study findings differ from those of Liang and Wanderer and are primarily due to the "error grid red" and "error grid yellow" GEM outliers that become much more obvious after 0800 and prevail until the 0200 running of the GEM process control material.

One of the most difficult roles of the clinical pathologist or clinical chemist is to define or communicate the clinical unacceptability of a clinical analytical system, especially if it is being used to provide clinical results. It is even more difficult to define an analytical system as unacceptable if it is used in patients who are labeled as "critically ill" and are placed in units where extraordinary testing and treatment modalities are *de rigueur*. Too often, the *milieu intérieur* of these patients is regarded as ever-changing and ever-fragile, especially when tested by an unstable analyzer. A tremendous amount of work has been accomplished in diabetes translational work that allows the apparent quality of a glucose result to be interpreted to the user of that glucose result, whether she be the patient or clinician or the manager of the laboratory operating this instrument.

The ISO glucose standards and FDA guidance documents are very helpful and binding and can be used by industry to create more compliant analytical systems. But they do not offer the visceral translation of apparent clinical outliers into a specific risk category. The Klonoff error grid provides its user the perspective of the astute diabetologist. If an analytical system produces a series of higher risk, clinically unacceptable glucoses, then the glucose measurement system itself may be unacceptable. Using the Klonoff error grid approach, the GEM-CL comparisons demonstrate severe and moderate glucose errors in far greater frequencies than the ABL. Using the 15-minute interval to determine glucose errors, the rate of moderate to severe glucose errors for the GEM and ABL is 3.5 and 0.6 per 1000 glucoses ordered. To make these figures more user-friendly, the daily number of BGA (116 GEM and 76 ABL) can be used to generate the average time between successive moderate and severe errors. For the GEM, these higher risk errors would occur every 2.4 days and every 22 days for the ABL. Using the five-minute interval, the time between moderate and severe errors extends to seven days for the GEM compared to 66 days for the ABL.

The astute clinical chemist/pathologist realizes that in the case of an imprecise analyzer, redoing the test may not provide useful information to either the analyst or the test orderer. Imprecise results cause clinician confusion, more test ordering, delays, and occasionally, patient harm. Compared to the

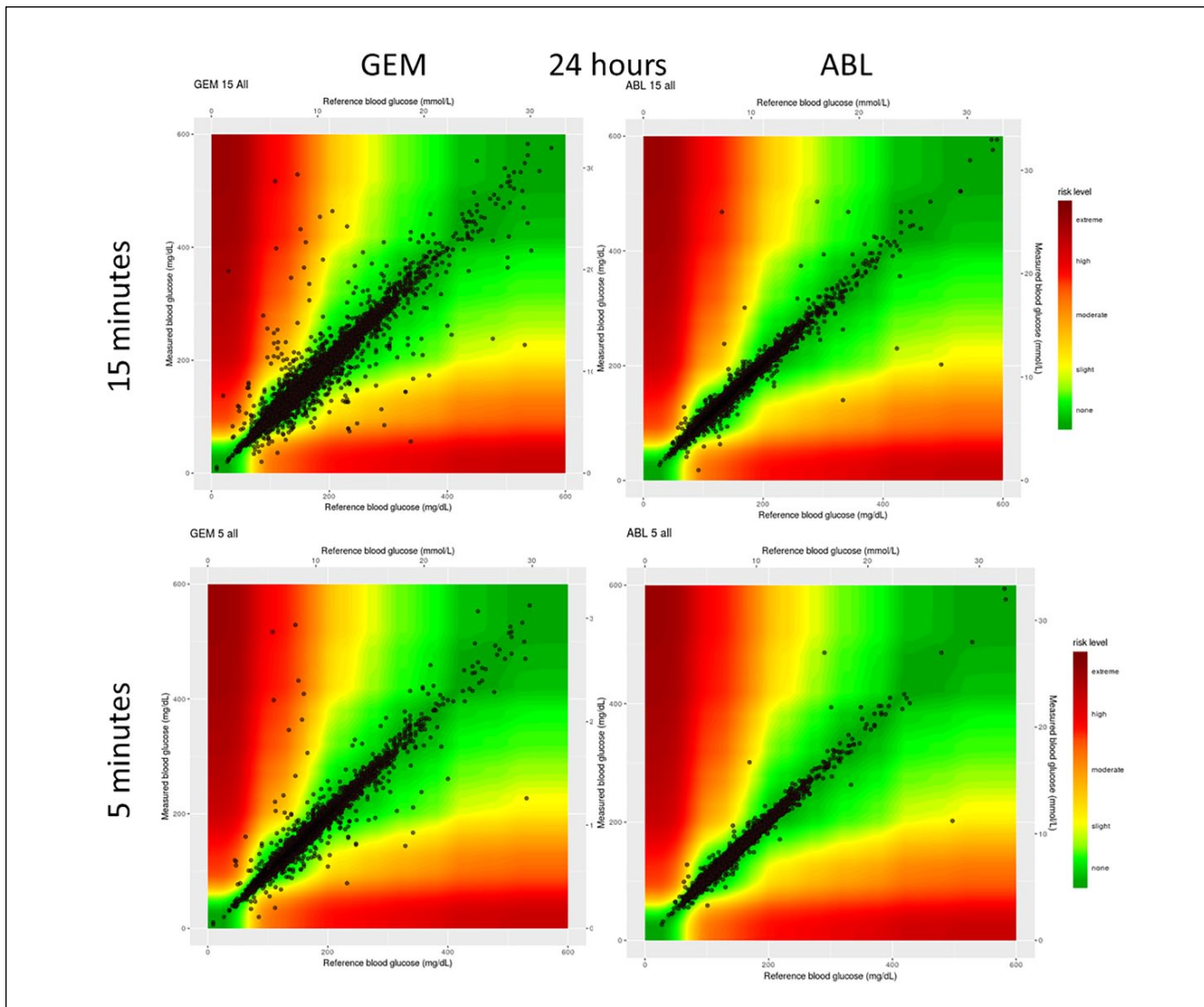


Figure 3. Surveillance error grids for 24-hour arterial blood gas-central laboratory comparisons. $\chi^2 = 17.1$ ($P < .0001$) for 15-minute ABL vs GEM prevalence of clinical outliers; $\chi^2 = 8.7$ ($P < .005$) for 5-minute ABL vs GEM prevalence of clinical outliers.

CL glucose ordering in the ABL ICU, it seems that the GEM ICU medical staff do not rest. Figure 6 (60-minute interval) shows that the seemingly extraneous CL glucose testing occurs continuously. This CL glucose testing extends to other tests and we recently reported that the 2016 cost of “GEM tests” sent to the CL (sodium, chloride, bicarbonate, potassium, and glucose) ordered by the Calgary GEM ICU was approximately \$116 000, far more than \$48 000 expended by the ABL ICU²⁰ for similar CL tests.

There is one other canary in the BGA glucose coal mine. It is interesting and perhaps relevant that the average glucoses of the ABL ICU patients are roughly 10 mg/dL less than the GEM ICU patients ($P < .01$). The bed count of the two ICUs is similar and the patient mix is probably not dissimilar as it represents patients admitted to Alberta’s two larger city academic quaternary hospitals separated by 180 miles and

supported by the provincial Alberta Health Services. We wonder if this difference might be related to the GEM’s diurnal imprecision variation. It is likely that the glucose results that would garner the most attention are those ordered in response to unexpected tests’ results, those that are obtained outside of the early morning hours. Figure 6 indicates that the 5-minute and even the 15-minute interval GEM comparisons consist primarily (>65%) of the 0200 to 0800 testing. The quality of those GEM tests is not in dispute. Only if the quality lens is focused on 0800 to 0200, the error rate is much higher. If the clinician attempts to respond to these possibly erroneous afternoon results, the clinician might eventually determine that less intervention is warranted as there is an insufficient relationship between the glucose measurement and therapy. The end result of this learned diminished intervention would be higher average glucose.

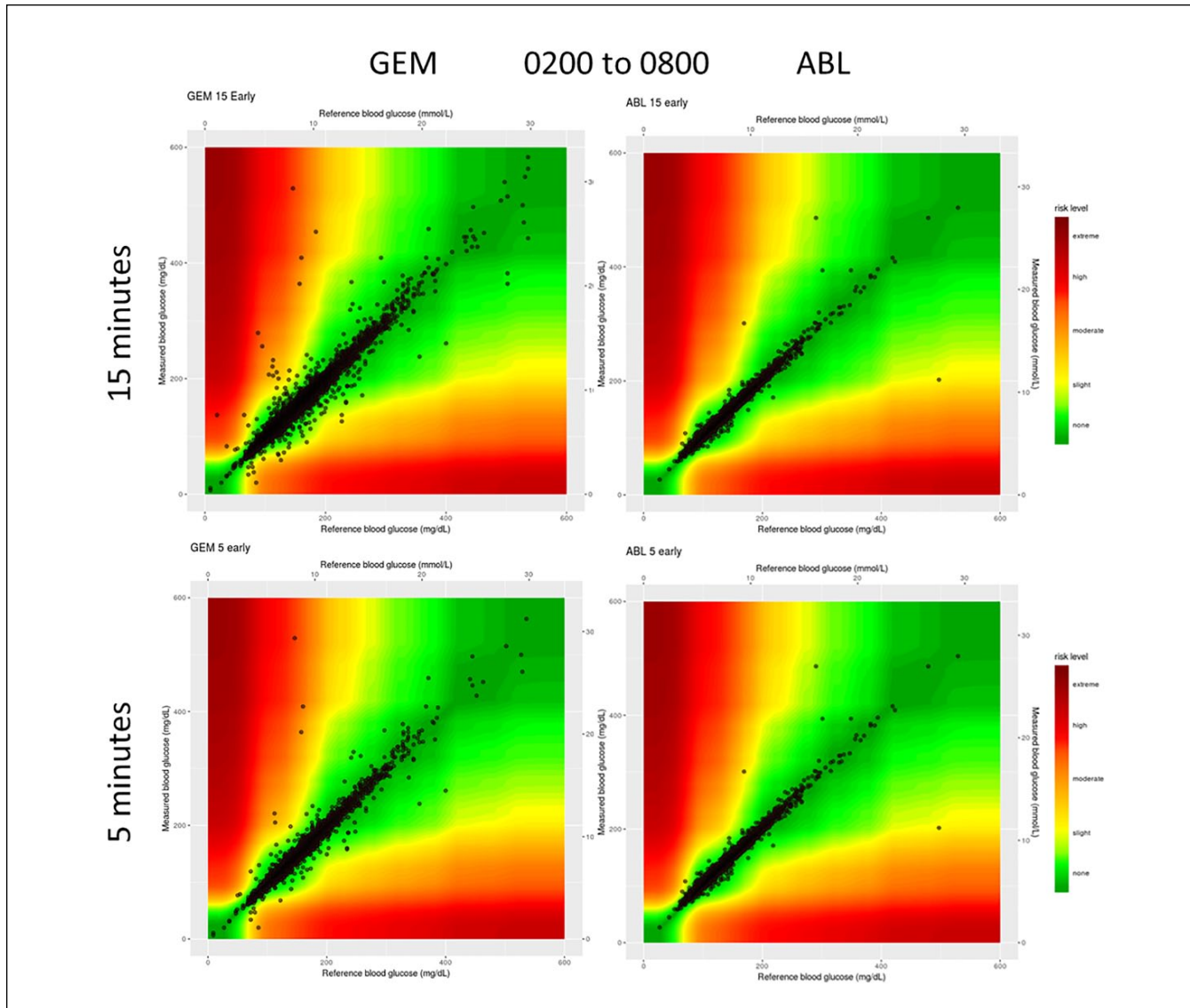


Figure 4. Surveillance error grids for early morning (0200-0800) arterial blood gas-central laboratory comparisons. $\chi^2 = 7.2$ ($P < .01$) for 15-minute ABL vs GEM prevalence of clinical outliers; $\chi^2 = 3.8$ ($P = .05$) for 5-minute ABL vs GEM prevalence of clinical outliers.

We have been wrestling with the question of why this GEM afternoon increase in glucose imprecision has not been reported by others. The answer is complex. Diane Vaughan, the originator of the expression “normalization of deviance,” attributes the ignoring of signals of potential danger by technical experts to be associated with the experts’ interpretations of the signals being shaped by a system that includes history, competition, scarcity, bureaucratic procedures, power, rules and norms, hierarchy, and culture and patterns of information.²¹

With regard to the question of the adequacy of the ABL to support tight glycemic control, our short answer is “yes” as Dr Van den Berghe uses one in her ICU. The longer answer is “probably.” A retrospective comparison of CL and BGA glucoses requires a much more accurate CL

glucose analyzer. The Beckman DxC glucose CV is about 3%; the ABL 800 is around 1.5% to 2.0%. Due to the propagation of errors, numeric comparisons between the Beckman and the ABL will demonstrate statistical variation closer to that of the Beckman, rather than the ABL. There is also the issue of the ABL’s positive 2% to 3% bias. Simulations are needed to definitively answer this question; off hand, a positive bias might result in more “fictitious” outliers that would be counted as FDA outliers. To the clinician, this bias is minute and is obviated by sequential use of a highly precise BGA. “Big data” retrospective studies such as ours are very important as they provide unique views of accuracy and imprecision that are virtually impossible to replicate with all of the current method evaluation approaches. These data, however, have

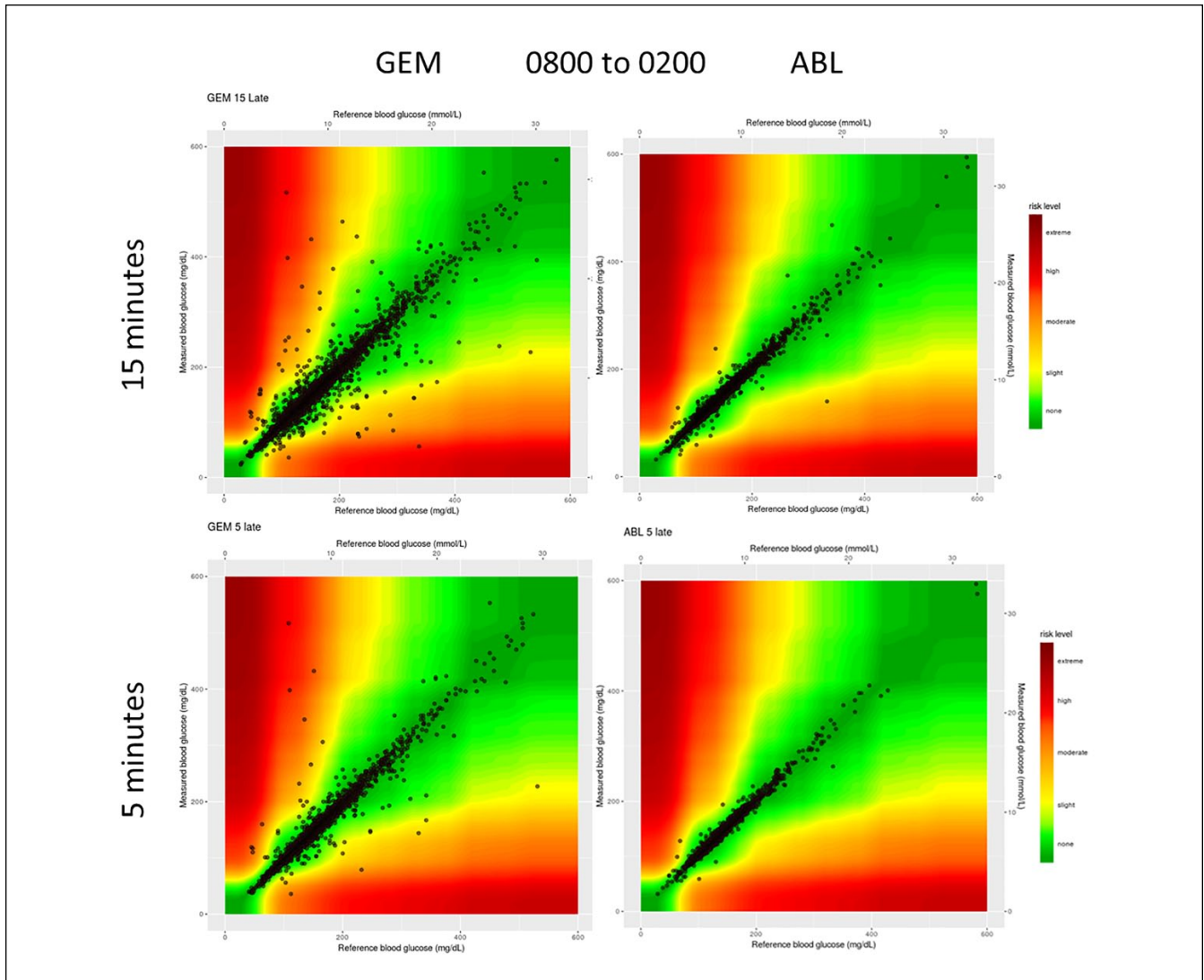


Figure 5. Surveillance error grids for later morning to evening (0800-0200) comparisons. $\chi^2 = 8.4$ ($P < .005$) for 15-minute ABL vs GEM prevalence of clinical outliers; $\chi^2 = 4.1$ ($P \leq 0.05$) for 5-minute ABL vs GEM prevalence of clinical outliers.

Table 3. Summary of Slight, Moderate, Severe, and Extreme Glucose Measurement Outliers, Derived From Klonoff Surveillance Error Grid Calculator, Repeat Kovatchev.¹⁷

Blood gas analyzer	Measurement period (h)	Glucose comparison Interval (min)	Nonerroneous glucose, N	Slight risk, lower glucose outliers, N	Slight risk higher glucose outliers, N	Moderate risk lower glucose outliers, N	Moderate risk higher glucose outliers, N	Severe risk, lower glucose outliers, N	Severe risk, higher glucose outliers, N	Extreme risk, higher and lower glucose outliers, N	Moderate to extreme risks outliers/1000 glucoses
GEM	24 (entire day)	15	12 325	280	58	25	16	3	0	1	3.5
GEM/	24 (entire day)	5	7064	117	19	8	10	1	0	0	2.6
ABL	24 (entire day)	15	7786	105	13	4	1	0	0	0	0.6
ABL	24 (entire day)	5	4150	57	6	1	0	0	0	0	0.2
GEM	18 (0800-0159)	15	4309	165	40	14	11	2	2	0	6.4
GEM	18 (0800-0159)	5	2349	65	11	4	8	1	0	0	5.3
ABL	18 (0800-0159)	15	2230	47	8	3	0	0	0	0	1.3
ABL	18 (0800-0159)	5	1139	21	2	1	0	0	0	0	0.9
GEM	6 (0200-0759)	15	8011	18	115	5	11	0	1	0	2.1
GEM	6 (0200-0759)	5	4712	52	8	4	2	0	0	0	1.3
ABL	6 (0200-0759)	15	5557	58	5	1	0	1	0	0	0.4
ABL	6 (0200-0759)	5	3011	36	4	0	0	0	0	0	0.0

The last column is sum of moderate, severe, and extreme glucose measurement outliers per 1000 glucoses ordered.

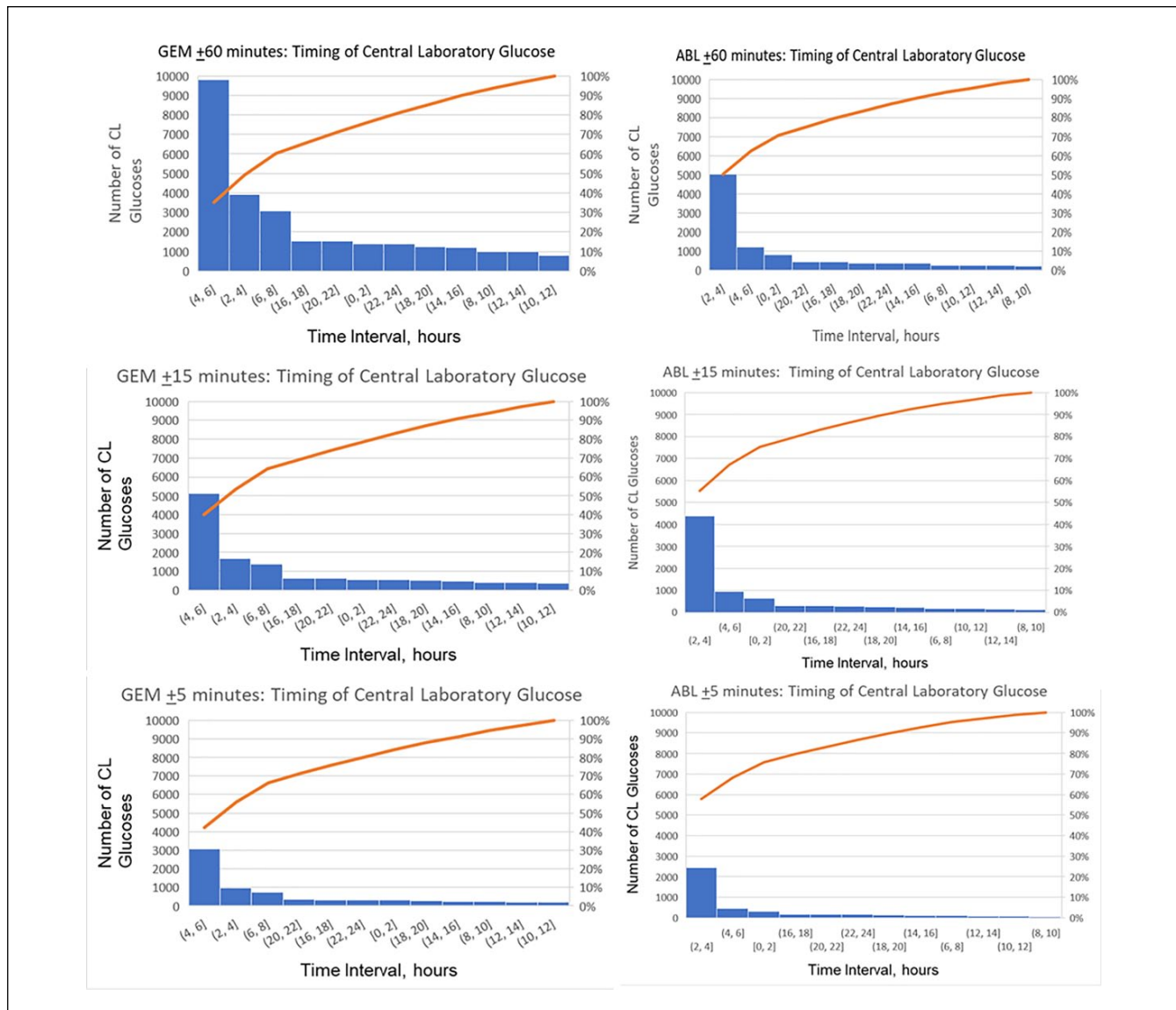


Figure 6. Pareto charts illustrating the bihourly volumes of central laboratory glucose ordering for central laboratory and blood glucose analyzer glucoses tested within 60, 15, and 5 minutes.

validated our hypothesis about two periods of testing in the GEM 4000, an early morning optimal period and the other 18 hours when factitious trends may be encountered due to consecutive inpatient samples being assayed on two GEMs rather than repeated on the same analyzer (repeat Reference⁹).


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

George Cembrowski  <https://orcid.org/0000-0001-5957-2941>

References

1. Van den Berghe G, Wouters P, Weekers F, et al. Intensive insulin therapy in critically ill patients. *N Engl J Med.* 2001;345(19):1359-1367.
2. Slater-Maclean L, Cembrowski G, Chin D, et al. Accuracy of glycemic measurements in the critically ill. *Diabetes Technol Ther.* 2008;10(3):169-177.
3. Nice-Sugar Study Investigators. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med.* 2009;360(13):1283-1297.
4. Cembrowski GS, Tran DV, Slater-MacLean L, Chin D, Gibney RN, Jacka M. Could susceptibility to low hematocrit

- interference have compromised the results of the NICE-SUGAR trial? *Clin Chem*. 2010;56(7):1193-1195.
5. NICE-Sugar Study Investigators. Hypoglycemia and risk of death in critically ill patients. *N Engl J Med*. 2012;367(12):1108-1118.
 6. Gunst J, Van den Berghe G. Blood glucose control in the ICU: don't throw out the baby with the bathwater! *Intensive Care Med*. 2016;42(9):1478-1481.
 7. Inoue S, Egi M, Kotani J, Morita K. Accuracy of blood-glucose measurements using glucose meters and arterial blood gas analyzers in critically ill adult patients: systematic review. *Crit Care*. 2013;17(2):R48.
 8. Slater-Maclean L, Tran DV, Chin D, Cembrowski G. Biologic and analytic variations of glucose determined from serial intensive care unit patient testing indicate inferiority of blood glucose meter glucose compared to blood gas glucose measurements. Paper presented at: Diabetes Technology Conference; 2008; Bethesda, MD; Abstract A154. *J Diabetes Technol*.
 9. Cembrowski GS, Xu Q, Cembrowski AR, Mei J, Sadrzadeh H. Impaired clinical utility of sequential patient GEM blood gas measurements associated with calibration schedule. *Clin Biochem*. 2017;50(16-17):936-941.
 10. International Organization for Standardization. *In vitro diagnostic test systems—requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus*. ISO 15197:2013.
 11. Food and Drug Administration. Blood glucose monitoring test systems for prescription point-of-care use—guidance for industry and food and drug administration staff. <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM380325.pdf>. Accessed October 10, 2016.
 12. Clarke WL. The original Clarke error grid analysis (EGA). *Diabetes Technol Ther*. 2005;7(5):776-779.
 13. Parkes JL, Slatin SL, Pardo S, Ginsberg BH. A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. *Diabetes Care*. 2000;23(8):1143-1148.
 14. Klonoff DC, Lias C, Beck S, et al. Development of the Diabetes Technology Society blood glucose monitor system surveillance protocol. *J Diabetes Sci Technol*. 2016;10(3):697-707.
 15. Klonoff DC, Lias C, Vigersky R, et al. The surveillance error grid. *J Diabetes Sci Technol*. 2014;8(4):658-672.
 16. Liang Y, Wanderer J, Nichols JH, Klonoff D, Rice MJ. Blood gas analyzer accuracy of glucose measurements. *Mayo Clin Proc*. 2017;92(7):1030-1041.
 17. Kovatchev BP, Wakeman CA, Breton MD, et al. Computing the surveillance error grid analysis: procedure and examples. *J Diabetes Sci Technol*. 2014;8(4):673-684.
 18. ABL800 FLEX reference manual. 200610E. Code number:989-963, Version 5.26.
 19. Nevalainen D, Berte L, Kraft C, Leigh E, Picaso L, Morgan T. Evaluating laboratory performance on quality indicators with the six sigma scale. *Arch Pathol Lab Med*. 2000;124(4):516-519.
 20. Mei J, Xu E, Kattar M, et al. Comparison of rates of nearly simultaneous, identical central laboratory testing associated with blood gas/electrolyte/metabolite point of care testing in two adult intensive care units. Paper presented at: National meeting of the American Association of Clinical Chemistry; August 2018; Chicago IL; *Clin Chem* 2018, Abstract B190, page S193, 70th AACC Annual Scientific Meeting Abstracts, 2018.
 21. Vaughn D. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. 2nd ed. Chicago: University of Chicago Press; 2016:416.