UC Davis

UC Davis Previously Published Works

Title

Computing the Surveillance Error Grid Analysis

Permalink

https://escholarship.org/uc/item/84f9j8t2

Journal

Journal of Diabetes Science and Technology, 8(4)

ISSN

0103-4065

Authors

Kovatchev, Boris P

Wakeman, Christian A

Breton, Marc D

et al.

Publication Date

2014-07-01

DOI

10.1177/1932296814539590

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Journal of Diabetes Science and Technology
2014, Vol. 8(4) 673–684
© 2014 Diabetes Technology Society
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1932296814539590
dst.sagepub.com
SAGE

Computing the Surveillance Error Grid Analysis: Procedure and Examples

Boris P. Kovatchev, PhD[1], Christian A. Wakeman, BS[1], Marc D. Breton, PhD[1], Gerald J. Kost, MD, PhD[2], Richard F. Louie, PhD[2], Nam K. Tran, PhD[2], and David C. Klonoff, MD[3]

Abstract

Introduction: The surveillance error grid (SEG) analysis is a tool for analysis and visualization of blood glucose monitoring (BGM) errors, based on the opinions of 206 diabetes clinicians who rated 4 distinct treatment scenarios. Resulting from this large-scale inquiry is a matrix of 337 561 risk ratings, 1 for each pair of (reference, BGM) readings ranging from 20 to 580 mg/dl. The computation of the SEG is therefore complex and in need of automation.

Methods: The SEG software introduced in this article automates the task of assigning a degree of risk to each data point for a set of measured and reference blood glucose values so that the data can be distributed into 8 risk zones. The software's 2 main purposes are to (1) distribute a set of BG Monitor data into 8 risk zones ranging from none to extreme and (2) present the data in a color coded display to promote visualization. Besides aggregating the data into 8 zones corresponding to levels of risk, the SEG computes the number and percentage of data pairs in each zone and the number/percentage of data pairs above/below the diagonal line in each zone, which are associated with BGM errors creating risks for hypo- or hyperglycemia, respectively.

Results: To illustrate the action of the SEG software we first present computer-simulated data stratified along error levels defined by ISO 15197:2013. This allows the SEG to be linked to this established standard. Further illustration of the SEG procedure is done with a series of previously published data, which reflect the performance of BGM devices and test strips under various environmental conditions.

Conclusions: We conclude that the SEG software is a useful addition to the SEG analysis presented in this journal, developed to assess the magnitude of clinical risk from analytically inaccurate data in a variety of high-impact situations such as intensive care and disaster settings.

Keywords

blood glucose monitoring, meter errors, hypoglycemia, hyperglycemia, error grid analysis

The surveillance error grid (SEG) analysis introduced in this issue of the *Journal of Diabetes Science and Technology*[1] is based on a carefully crafted set of questions answered by 206 clinicians with expertise in the treatment of diabetes, and on advanced mathematical interpretation of the obtained results transforming the SEG into a modern tool for the analysis of errors associated with blood glucose monitoring (BGM). While the predecessors of the SEG—the now classic Clarke[2] and Parkes[3] error grid analyses—were amenable to straightforward visual interpretation and pen-and-paper classification of meter errors into several zones, the SEG introduces a new dimension to the evaluation of the accuracy of BGM devices—the *perceived level of risk associated with meter errors*. The logic behind such a risk evaluation is self-evident: as shown in a recent report, even the best contemporary BGM systems are prone to more or less frequent errors,[4] potentially resulting in deterioration of glycemic control parameters as shown by in silico experiments. These experiments suggested a threshold effect between 10% and 15% permitted BGM error for most parameters, except for HbA1c, which appeared to be increasing relatively linearly with increasing BGM error above 10%.[5]

The missing link between the magnitude of BGM errors and the ensuing deterioration of glycemic control is therefore the level of risk created by BGM inaccuracy, which occasionally translates into inappropriate treatment decisions and, eventually, into inferior treatment outcomes. The keyword

[1]University of Virginia, Center for Diabetes Technology, Charlottesville, VA, USA
[2]University of California, Davis, Point-of-Care Testing Center for Teaching and Research, Davis, CA, USA
[3]Mills-Peninsula Health Services, San Mateo, CA, USA

Corresponding Author:
Boris P. Kovatchev, PhD, University of Virginia, Center for Diabetes Technology, PO Box 400888, Charlottesville, VA 22908, USA.
Email: boris@virginia.edu

here is "occasionally," which implies that a meter reading, no matter how inaccurate, would not always trigger inappropriate treatment. This is because a treatment decision results from a complex behavioral process that factors in user experience, practice, and the dynamics of the surrounding situation.[6] With all these objective and subjective factors in play, the best approach to risk evaluation is an interpretation of expert opinion that synthesizes the collective wisdom of a number of practicing clinicians into a formal risk assessment method.

In brief, the SEG[1] takes on the risk interpretation of expert opinion in several steps: First, 4 archetypal patient scenarios were formulated, 2 involving type 1 and 2 involving type 2 diabetic patients. Each expert was asked to rate 2 scenarios, which generated between 101 and 105 responses per scenario or 412 responses total. Each expert completed a table with the minimum and maximum of the range of blood glucose levels that would correspond to 1 of 5 types of actions: (1) emergency treatment for low BG, (2) use of oral glucose, (3) no action, (4) use of insulin or exercise and less food intake, and (5) emergency treatment for high BG. For each data pair of measured and reference glucose, the experts identified the degree of risk associated with an action taken because of a measured BG reading, compared to the action that would have been taken if the reference BG had been known. The degree of risk for hypo- or hyperglycemia identified by the experts was coded from 0 (none) to 4 (extreme). An example of how this first step of SEG development resulted in the visual (and numerical) interpretation of the perceived risk associated with BGM errors is presented in Figure 1A.

The numerical and clinical differences between boundaries of the 4 scenarios were determined to be generally insignificant and the developers of the SEG concluded that creation of a separate error grid for each scenario was not justified. All 412 survey responses to the 4 scenarios were therefore combined. The product was a single color-coded error grid plot, which averaged all the respondents' boundaries and became the SEG (Figure 1B). It is important to note that whereas each expert's boundaries between risk zones were determined individually, the SEG error grid boundaries represent the composite ratings of all the experts; thus, the lines between the risk zones are blurred and the plot is nearly continuous. For each data pair a risk value was calculated corresponding to the average perceived risk across all experts over all scenarios. Average risk ratings were created for every possible reference BG and corresponding measured BG, both ranging from 20 to 600 mg/dL, which resulted in a color-coded matrix of 337 561 risk values, which became the basis for the SEG.[1] Across the plot, the average risk rating ranges from 0 (none) to 4 (extreme).

Computation of the SEG is not practical by hand because of the fine gradation of risk levels for nearby data points. Thus, appropriate software is needed to produce the SEG graphs and, most important, to tabulate the data points into a distribution presenting each level of risk, corresponding to the errors associated with the performance of any particular BGM device. In this article we introduce the computation of the SEG implemented in Excel-based macro, and illustrate its use with computer-simulated meter errors corresponding to different levels of International Organization for Standardization (ISO) accuracy, and with real data sets collected during experiments testing BGM devices under various, previously described normal-use and stress conditions.[7-12] The software for computing the SEG in this article is copyrighted by the University of Virginia (Charlottesville, VA) and will be available through the Diabetes Technology Society website: at www.diabetestechnology.org/SEGsoftware.

Methods

Software

An Excel Macro (for Excel 2013 with Data Analysis Pack) was developed to compute the results of the SEG analysis. The user interaction with the Excel Macro is straightforward and includes the following 3 steps:

1. Enter or paste reference blood glucose data in column 1 of the Excel spreadsheet labeled reference BG.
2. Enter or paste measured BG meter data in column 2. The SEG plot is automatically generated from the data; all readings outside the SEG range (20 to 600 mg/dL) are counted as "out of range."
3. Hit "Run" to compute the counts and percentages in the SEG zones.

To reset the data, hit "Reset."

The layout of the Excel Macro is illustrated in Figure 2. The output of the procedure includes

- Device ID introduced by the user, date, and the number of valid data pairs split into:
  - Hypo. Pairs—BGM overestimates BG and could trigger risk for hypoglycemia (data pairs above the diagonal line)
  - Hyper. Pairs—BGM underestimates BG and could trigger risk for hyperglycemia (data pairs below the diagonal line)
- The number and percentage of data pairs in each of 8 SEG risk zones: none, slight (lower and higher), moderate (lower and higher), great (lower and higher), and extreme, total and split into hypo- and hyperglycemia risk pairs.

SEG and ISO 15197:2013: Suggested Guideline for SEG Cutoffs

To assess the agreement of the SEG with the current standard for BGM accuracy ISO 15197:2013,[13] which specifies that 95% of the data pairs should be within ± 15 mg/dL from

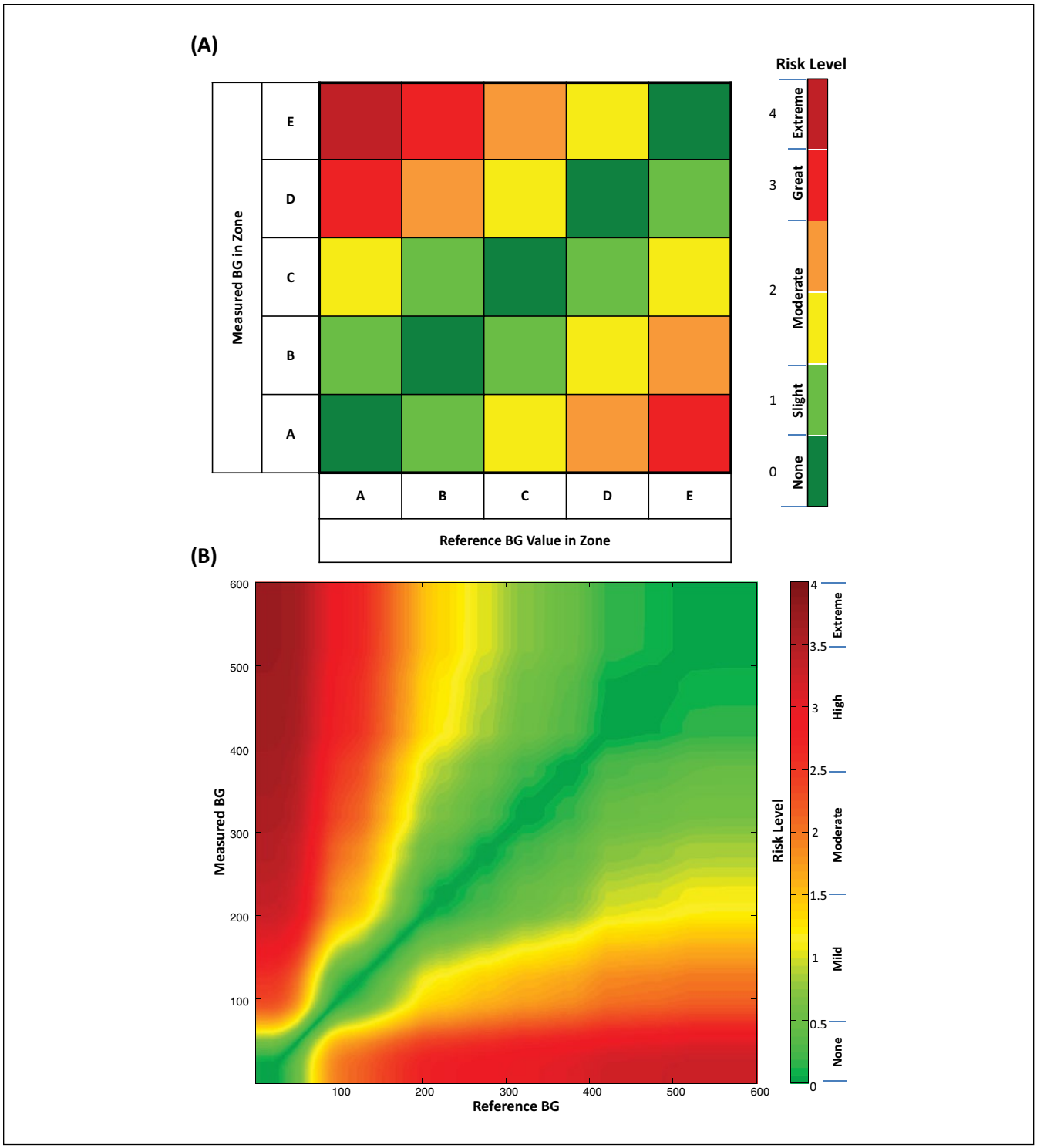

Figure 1. Constructing the zones of the Surveillance Error Grid: discrete risk assessments of treatment scenarios by each individual expert (A) and continuous aggregated risk assessment across all experts (B).

reference for reference glucose levels < 100 mg/dL, or within ± 15% from reference for reference glucose levels ≥100 mg/dL, we generated several sets of reference-meter pairs corresponding to meter errors of 0% of data points outside of the ISO 15197:2013 limits (a perfect device according to ISO), 5% outside the ISO limits (ie, 95% within limits, which is exactly the current ISO standard requirement), and then 10%, 15%, 20%, 25%, and 30% of data points outside of ISO

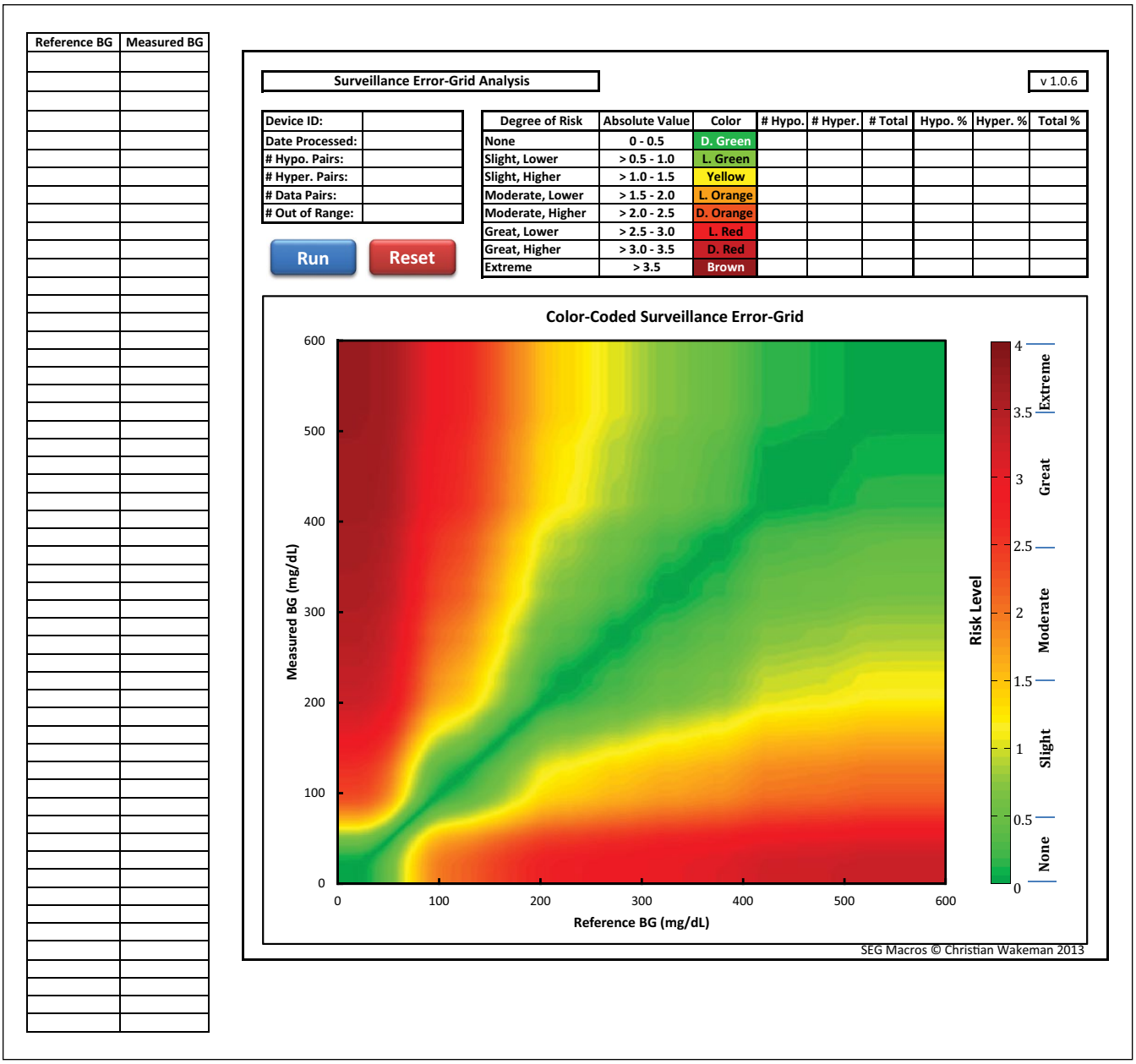

Figure 2. Data entry and output of the Surveillance Error Grid Excel Macro.

limits. The distribution of the errors was computer-simulated to approximate the shape of the error distributions observed in real devices.[4] Specifically, Freckmann et al[4] presented detailed percentiles ranging from ± 5% to ± 20%, below and above 75 mg/dl and 100 mg/dl, for the error distributions of 43 BGM devices. These percentiles allowed the reconstruction of realistic computer-simulated BGM error distribution. Then, the ISO thresholds were modeled as follows: for each error level of X%, data pairs are generated such that the percentage of data pairs outside of the ISO standard boundaries is equal to X%, with X ranging from 0 to 30. For each ISO error level we generate 3334 data pairs, which ensure that low-probability outliers are adequately covered. It should be underscored that such a manipulation of meter errors can only be performed by computer simulation since the error of a real experiment cannot be predetermined.[5,14-16] This large-scale simulation allows us to assess the percentage of *at-risk data points*—those that are outside of the green "no risk" zone of the SEG for each ISO 15197:2013 error level and to suggest a general guideline for the acceptable percentage of at-risk values that could be used for future classification of BGM devices. In addition, to illustrate the SEG-ISO relationship, we plotted the simulated data pairs for 0%, 5%, 15%, and 25% ISO error level. To clarify, 5% error level, or

95% data pairs meeting the ISO requirements is the exact cutoff level of acceptance for ISO 15197:2013.

Clinical Examples

We used several previously published data sets to illustrate the SEG procedure with data collected under various BGM testing conditions:

Example 1. In this study, 3 hospital glucose meter systems were tested against the YSI 2300 glucose analyzer (YSI Life-Sciences, Yellow Springs, OH). A total of 613 arterial whole blood samples from critically ill patients were obtained from the blood gas laboratory. Samples were tested once on each of the BGM systems in random order to decrease systematic bias. Remaining samples were spun down and plasma glucose was determined on YSI analyzer. Two meters, device 1 and device 2, were from the same manufacturer. For detailed description of these devices we refer to the original article providing the data;[7] here we focus on illustrating the performance differences between devices detected by SEG.

Example 2. Assessing the performance of BGM testing for critical care. Data from a multicenter study conducted at 59 US sites comparing a modified glucose oxidase, 4-well, interference-compensating test strip and meter system to paired reference measurements (N = 2767) obtained from 21 chemistry laboratory analyzers. Multicenter sites anonymously provided paired observations obtained from critical care, nursery and clinic areas. Here we present a SEG plot depicting the combined performance of this BGM device (device 4) against clinical laboratory measurements. Detailed description of this device and the experimental setup of testing can be found in the original articles describing these data.[8,9]

Example 3. The objective of this study was to assess the effectiveness of traditional vial versus foil packaging for preserving glucose test strip performance in humid conditions. Glucose test strips were exposed to mean relative humidity of 97.0 ± 1.1% in an environmental chamber for up to 168 hours. At defined time points, stressed strips were removed and tested in pairs with unstressed strips using whole blood samples spiked to glucose concentrations of 60, 100, and 250 mg/dL.[10] Two SEG plot analyses were conducted to depict the effectiveness of vial versus foil packaging in preserving test strips.

Example 4. For this example, the study evaluated the effects of short-term (≤1 hour) static high temperature and humidity stresses on the performance of point-of-care (POC) glucose test strips and meters. Glucose test strips and meters were exposed to a mean relative humidity of 83.0 ± 8.0% and temperature of 42 ± 3.2°C in an environmental chamber. Stressed and unstressed glucose reagent strips and meters were tested with spiked blood samples (n = 40 measurements per time point for each of 4 trials) after 15, 30, 45, and 60 minutes of exposure.[11] SEG analysis depicted test strip performance for all stress exposure times combined.

Example 5. The objective of this experiment was to characterize the performance of 2 glucose meters and test strips using simulated dynamic temperature and humidity disaster conditions. Glucose oxidase- and glucose dehydrogenase-based test strips were dynamically stressed for up to 680 hours using an environmental chamber to simulate conditions during Hurricane Katrina of 2005. Paired measurements versus control were obtained using 3 aqueous reagent levels for 2 BGM devices.[12] Here we present 4 SEG plots depicting the performance of each device with strips exposed to dynamic stress of ≤72 hours versus test strips exposed to dynamic stress of >72 hours.

Results

SEG and ISO 15197-2013

Table 1 presents the percentage of at-risk data pairs (which we defined as those outside of the green "no risk" zone of the SEG where 100% data pairs exhibited a risk score of ≤0.5) for ISO 15197:2013 error levels of 0%, 5%, . . . 30%.

Figures 3A, 3B, 3C, and 3D elaborate 4 specific cases from Table 1, visualizing ISO errors of 0%, 5%, 15%, and 25%, respectively. The "perfect BGM, which has 0% readings outside of the limits of the ISO standard still has a small percentage of at-risk data pairs. Increasing the ISO error to the exact ISO requirement (5% outside the 15 mg/dL and –15% limits, or 95% within the acceptance limits) results in a narrowly spread data cloud with >3.2% at-risk data pairs. Increasing further the magnitude of the errors to 15% and then to 25% outside the ISO limits results in increasing spread of the data cloud and outliers in the "great" and "extreme" risk zones.

Suggested Guideline for SEG Cutoffs

From Table 1 and Figure 2, it follows that the current ISO requirement—95% of all data pairs to be within the ISO limits (shaded column in Table 1)—corresponds to 3.2% at-risk data pairs outside of the SEG "green" zone. While these percentages are distribution-dependent, we can suggest that BGM devices exhibiting ≥3.2% at-risk data pairs should be classified by the SEG as insufficiently accurate. Any data points in the "great" or "extreme" risk ranges should be considered as failed data points and should be carefully evaluated by regulatory agencies during surveillance of BG monitors.

Clinical Examples

Figures 4 to 7 depict the clinical examples described in the Methods:

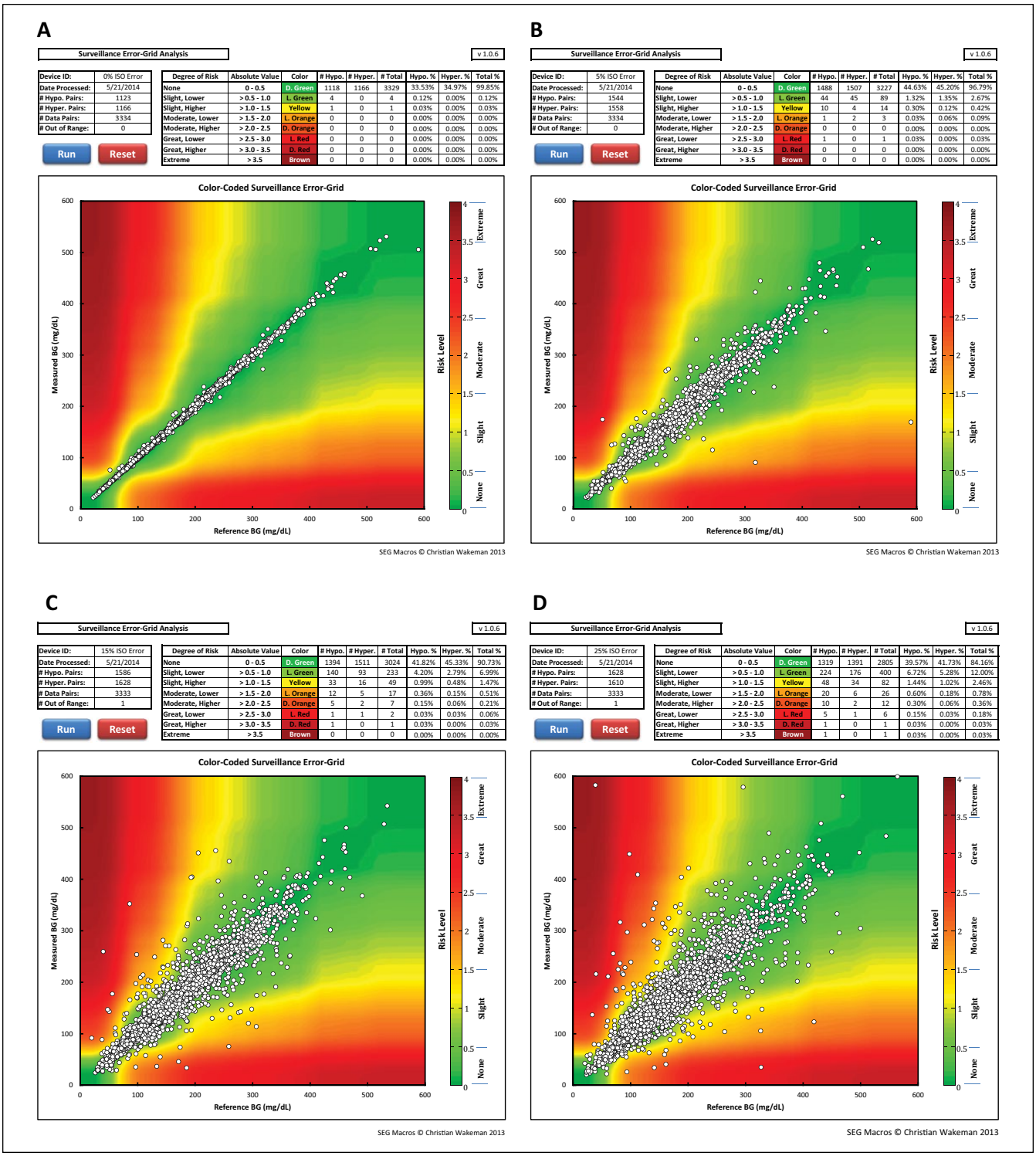

Figure 3. Computer-simulated meter errors corresponding to ISO 15197:2013 error levels of 0% (A), 5% (B), 15% (C), and 25% (D); 5% ISO error corresponds exactly to the standard requirement for accuracy.

Example 1. It is evident that the SEG differentiates the devices well, with device 3 being the worst performer visually and numerically with 20.5% at-risk data pairs (Figure 4C). Comparing the percentage of hypo- to hyper at-risk data pairs we can conclude that device 1 and device 2 tend to underestimate BG and create more risk for hyperglycemia (% hyper deviations > % hypo deviations in Figures 4A and 4B), while device 3 tends to create higher risk for

Table 1. Correspondence Between SEG and ISO 15197-2013 Error Levels.

ISO error level (%)	0	**5**	10	15	20	25	30
At-risk pairs (%)	0.15	**3.2**	6.5	9.3	12.4	15.8	19.6

hypoglycemia (% hypo deviations > % hyper deviations in Figure 4C).

Example 2. Figure 5 depicts the performance of hospital BGM devices in several clinical centers. This is an example of good BGM accuracy, shown visually and numerically by the SEG, with about 3% at-risk data pairs, none of which leave the lower slight-risk zone. Predictably all data pairs are clustered well along the primary diagonal of the SEG.

Example 3. Figure 6A presents effectiveness of vial packaging in preserving POC glucose test strip performance in humid conditions; Figure 6B presents the effectiveness of foil packaging under the same conditions. According to the SEG, in the conditions of this study, current vial packaging worked better than foil to preserve strip performance (0.56% vs 4.05% at-risk readings)—an observation that was not made in the original manuscript presenting these data (which only concluded that both packaging designs appeared to protect glucose test strips from high humidity stress), although the original manuscript recorded 7 foil-sealed strip failures and only 3 vial-packaged failures.[10]

Example 4. Figure 7 presents the deviation of short term high-temperature/high-humidity stressed tests trips from nonstressed tests strips in this experiment. While the BG testing range is narrow, it is evident that stressed tests strips produced elevated glucose results—a conclusion which confirms the original experiment, that detected stressed meter and test strip bias as high as 33 mg/dL.[11] Remarkably, the bias is so strong that the SEG indicates consistent overestimation of BG levels—159 out of 160 data pairs are above the SEG diagonal, 1 data pair is on the diagonal, and there are none below. Nevertheless, all data pairs were classified in the no-risk zone. Thus, while these stress conditions produce numerical bias, the clinical implications of this bias are insignificant.

Example 5. Figures 8A and 8B present the SEG plot for meter 1, and Figures 8C and 8D present the SEG plots for meter 2, both exposed to dynamic high-temperature/high-humidity stress for ≤72 and >72 hours. Consistent with the original report,[12] BGM performance deteriorates (albeit slightly) even with short exposure to dynamically changing environment, and continues to deteriorate with exposure >72 hours as indicated by increase in at-risk data pairs. However, this increase is not clinically significant according to the expert opinion behind the SEG because in both sets of conditions fewer than 1% of data points were at risk.

Discussion

The panel of experts which developed the SEG decided that the greatest need for the SEG would be for surveillance and postmarket assessment of cleared devices and this metric would provide a tool for use by FDA, BG monitor manufacturers, and other regulatory bodies to assess the degree of clinical risk from clinically inaccurate BG monitors for postmarket decision making. The SEG was therefore developed to assess the magnitude of clinical risk from particular analytically inaccurate data points for postmarket surveillance purposes and not for assessing the overall performance of a BG Monitor based on its mean performance.

The SEG represents a 3-dimensional interpretation of the risk associated with BGM errors, as perceived by 206 experts in diabetes care rating 4 different treatment scenarios.[1] The X- and Y-axis of the SEG plot are identical to the axes of the classic Clarke[2] and Parkes[3] error gird analyses, corresponding to reference and meter BG values, respectively. However, instead of the discrete error zones outlining various degrees of meter deviation in the Clarke and Parkes analyses, the SEG presents a third dimension—a continuous risk rating derived from the opinions of the expert survey respondents. In this third dimension each (reference, BGM) data pair gets an "elevation" proportional to the average risk rating given by the experts to that data point. The elevation is color coded from green (risk rating = 0) to brown (risk rating = 4). The shades in-between vary almost continuously because different experts endorsed different risk levels at different treatment scenarios. For convenience, the continuous SEG plot can be subdivided into 8 risk zones corresponding to risk increments of 0.5, and the zones are labeled from "no risk" to "extreme risk" accordingly. However, as opposed to Clarke's and Parkes' analyses, the risk values within a zone are not uniform: while for example all data pairs in the Clarke's or Parkes' A zone would have the same "weight," the risk within each SEG zone varies continuously from the lower to the higher ends of the zone, leaving no discrete "jumps" in risk values when a zone threshold is crossed. This is an important distinction, which solves a long-standing controversy in the risk assessment of meter errors.

Comparison of the SEG to ISO 15197:2013 using computer-simulated data pairs with realistic error distribution suggests that a device with ≤3 % errors outside of the SEG no-risk "green" zone corresponds would meet the ISO requirements of ≤5% data pairs outside the 15 mg/dL 15% standard limits, while higher percentages outside the SEG no-risk zone would indicate noncompliance with the standard (see Figure 3). It must be understood, however, that there is no one-to-one correspondence between the SEG and ISO 15197:2013. The SEG is continuous, based on the aggregated opinions of multiple experts, while ISO is binary, defining a single threshold (15 mg/dL at 15%

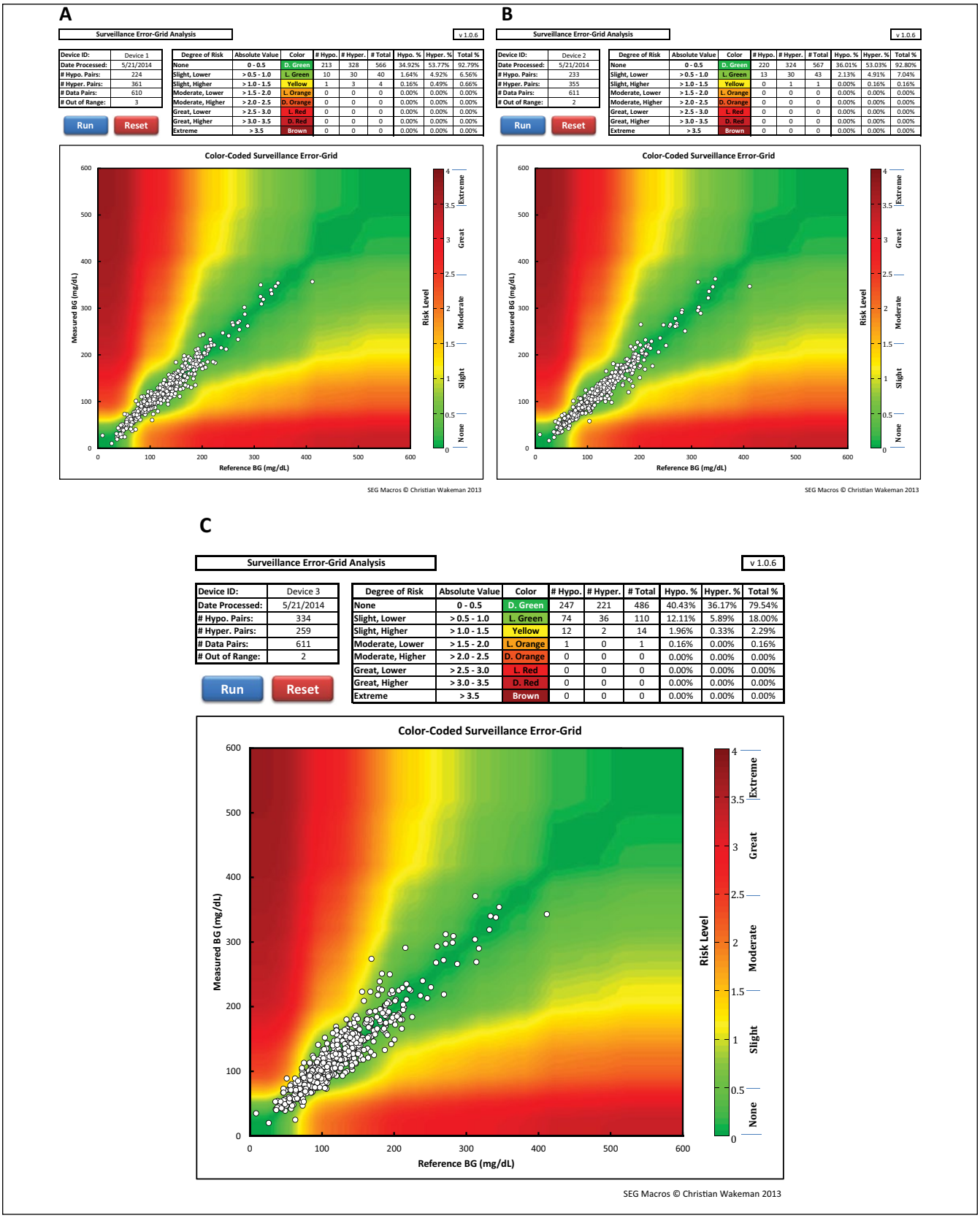

Figure 4. Performance of 3 hospital glucose meter systems against YSI 2300 plasma glucose reference method, showing (1) similar accuracy results between device 1 (A) and device 2 (B), which were from the same manufacturer; (2) inferior performance of device 3 (C), and (3) differences in device bias toward hyperglycemia observed in device 1 (A) and device 2 (B), and toward hypoglycemia observed in device 3 (C).

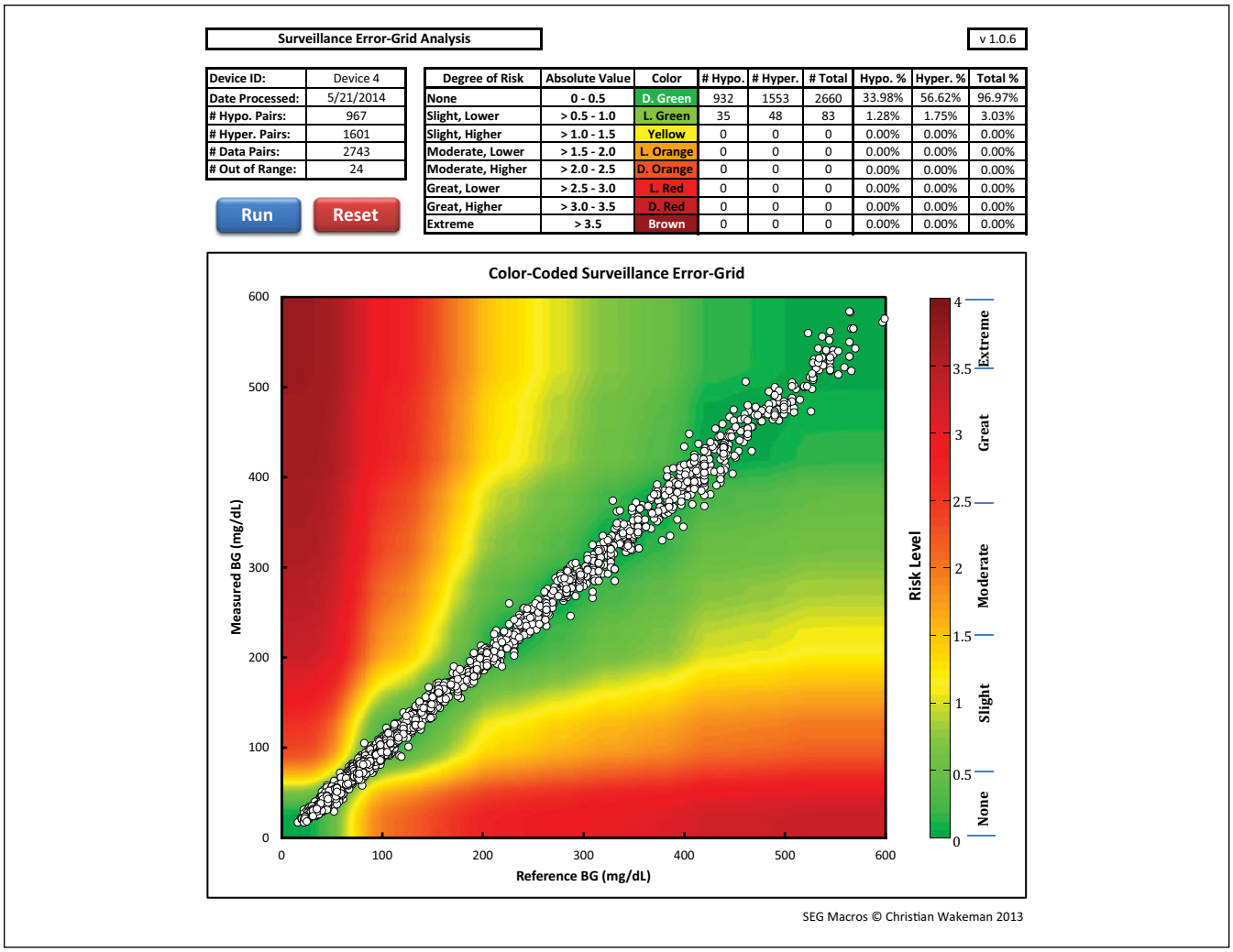

Figure 5. Performance of hospital BGM device (labeled device 4) in several clinical centers.

standard limits) that classifies the data pairs into accurate or inaccurate, without any nuance in-between. In particular, the ISO criteria does not provide indication which data pairs should be considered outliers, or dangerous meter errors - ISO 15197:2013 addresses the magnitude of *analytical errors* and all data pairs outside the 15 mg/dL−15% standard boundaries are considered equal in the ISO computation. Thus, while not meant as a replacement of ISO 15197:2013, the SEG provides additional insight because it assesses the magnitude of *clinical error* and provides fine data resolution for the surveillance of meter errors and for their relevance to the clinical practice as perceived by diabetes experts.

The continuous nature of the SEG amounts to a large matrix of 337 561 elements (which is equal to 581^2) that is needed to compute the analysis. Each element of this matrix is the aggregated expert opinion for the treatment risk associated with a (reference, BGM) data pair, both elements of which can range from 20 to 600 mg/dl. To automate the computation of the SEG, this matrix is embedded into software, such as the Excel Macro used for the computation of all examples in this article. The work with this Excel Macro is conceptually straightforward, as illustrated in Figure 2, and by simulated and real-data examples in this article.

To create these illustrations we used (1) computer simulation of meter errors that correspond to errors observed in a large survey of BGM devices[4] and (2) data from previously published experiments testing BGMs and strips under various environmental conditions.[7-12] With both simulated and real data, the SEG showed results that were intuitively clear and corresponding well to previous findings.

Thus, we can conclude that the SEG software would be a useful addition to the SEG analysis presented in this journal,[1] developed to assess the magnitude of clinical risk from analytically inaccurate data points in a variety of high-impact situations such as intensive care and disaster settings.

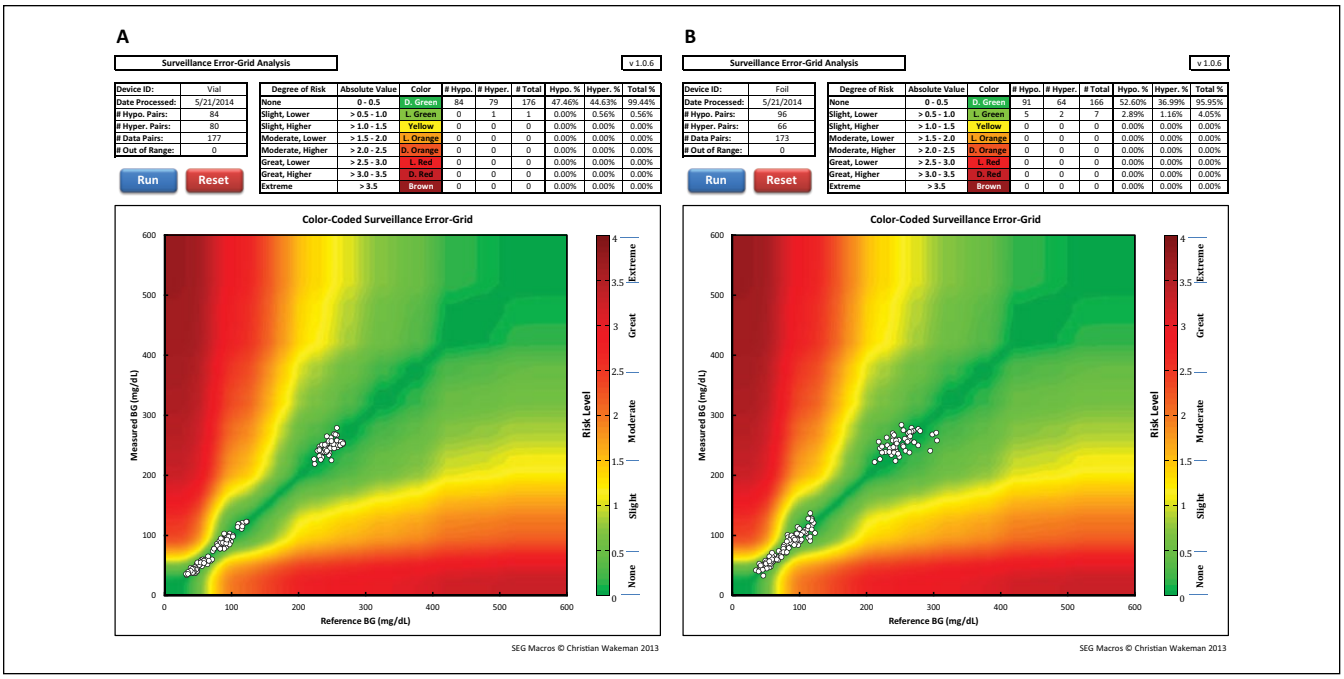

Figure 6. Effectiveness of vial test strip packaging (A) vs foil test strip packaging (B) in preserving POC glucose test strip performance in humid conditions. In this figure the axis "reference BG" refers to tests strips, which were not exposed to environmental stress.

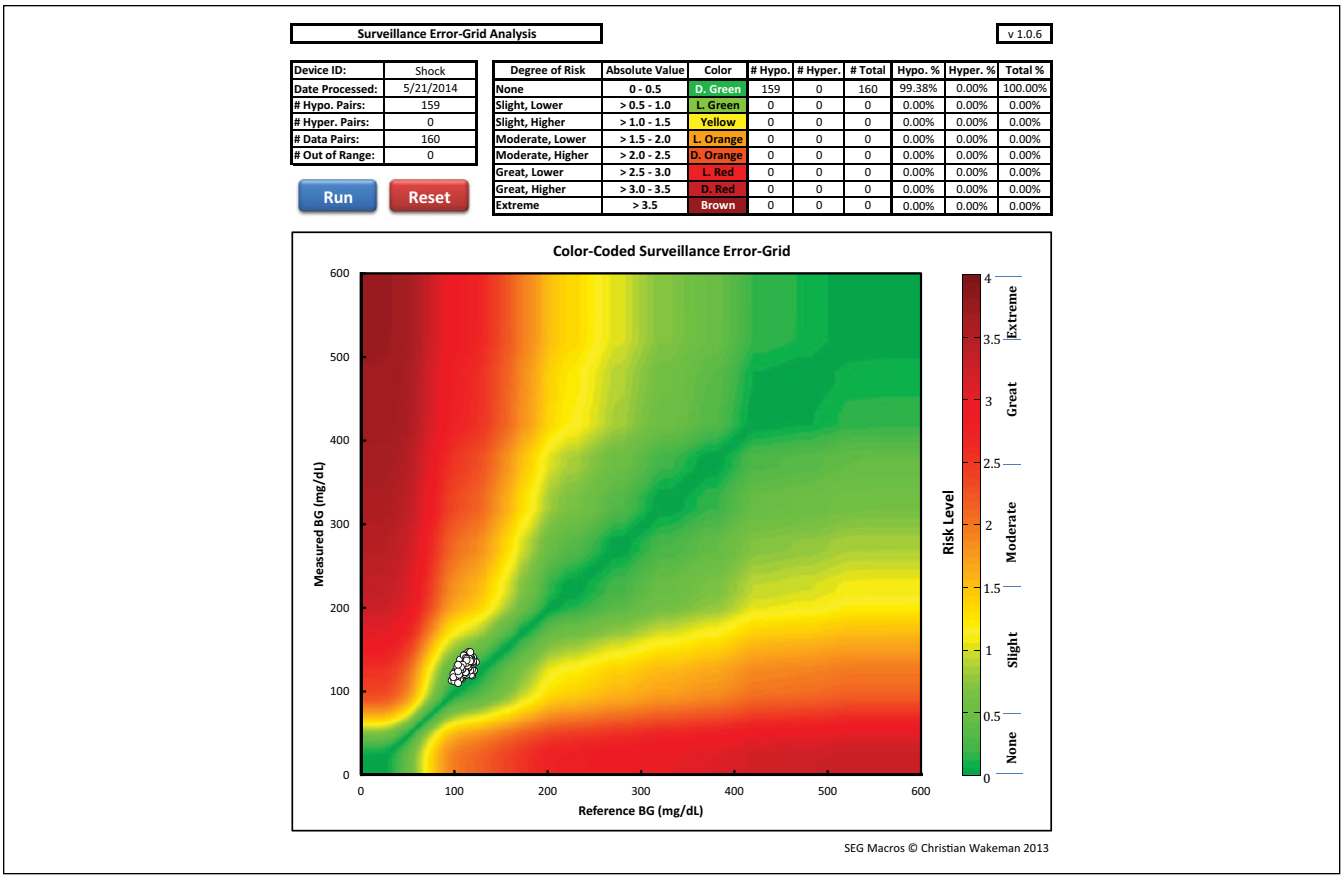

Figure 7. Effects of short-term (≤1 hour) static high temperature and humidity stresses on the performance of glucose test strips. The SEG plot presents deviation of stressed from nonstressed tests strips plotted on the reference BG axis.

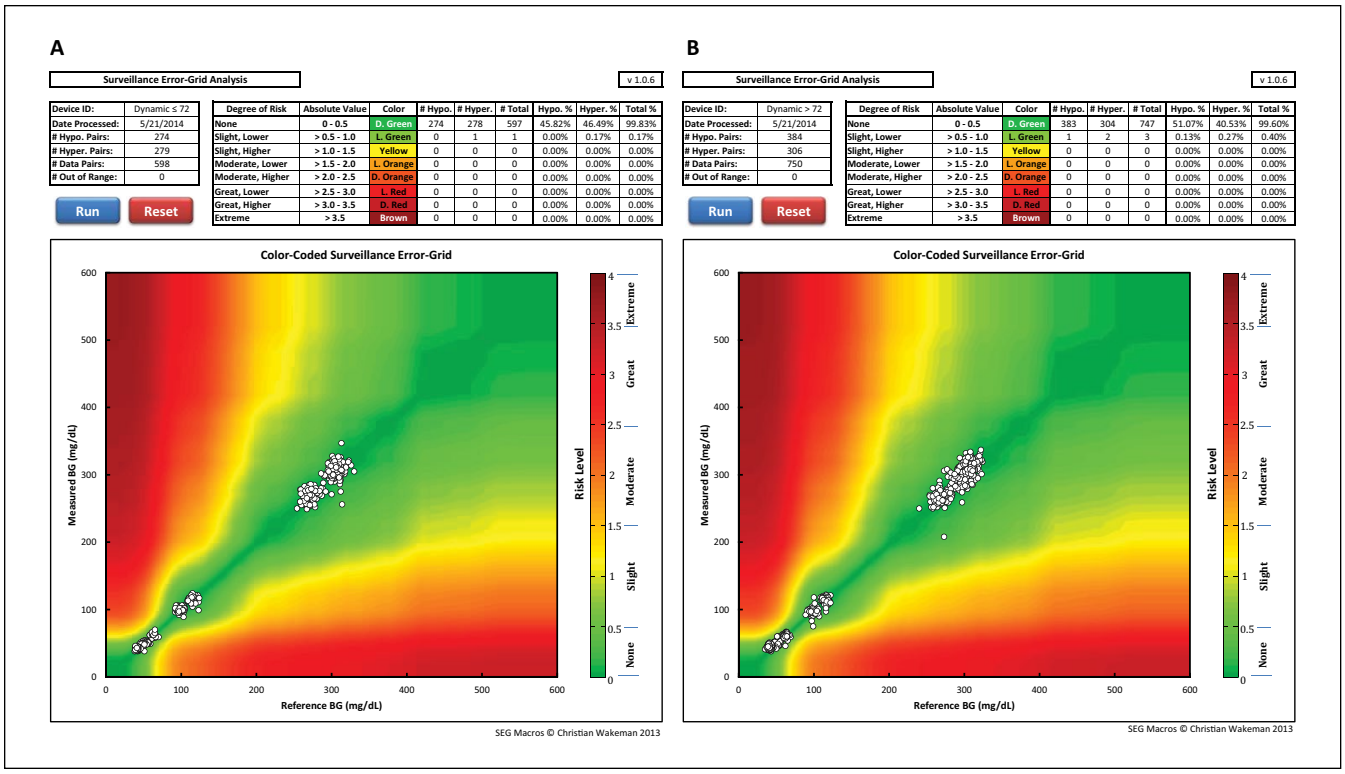

Figure 8. Performance of 2 glucose meters and test strips exposed to simulated dynamic temperature and humidity disaster conditions for ≤72 hours (A) vs >72 hours (B). The SEG plot presents deviation of stressed from nonstressed tests strips plotted on the reference BG axis.

Abbreviations

BGM, blood glucose monitoring; SEG = surveillance error grid.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: BK: advisory panel: Animas, Sanofi-Aventis; research grant/study material support: Animas, BD, DexCom, Insulet, Roche Diagnostics, Sanofi-Aventis, and Tandem; patent royalties: Johnson&Johnson, Sanofi Aventis. MB: consultant/advisor: Roche Diagnostics; research grant/material support: BD, Dexcom, Insulet, Roche Diagnostics, Sanofi-Aventis, and Tandem; patent royalties: Johnson&Johnson, Sanofi Aventis. GK none. RL: none. NT: material support from Nova Biomedical. DK: consultant to Google, Insuline, LifeCare, Roche, Sanofi, and Voluntis.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the Diabetes Technology Society. The computer simulation environment used for all experiments was developed with the support by the JDRF Artificial Pancreas Project and by grant R01 DK 085623 from the National Institutes of Health. The POCT•CTR supported research that produced data sets contributed from references 7, 8, and 9. Research that produced data sets from references 10, 11, and 12 was supported by a grant from the National Institute of Biomedical Imaging and Bioengineering (NIBIB), National Institutes of Health (U54EB007959, principal investigator, GK). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIBIB or the National Institutes of Health.

References

1. Klonoff DC, Lias C, Vigersky R, et al. The surveillance error grid. *J Diabetes Sci Technol*. 2014. doi: 10.1177/1932296814539589.
2. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating the clinical accuracy of self-blood glucose monitoring systems. *Diabetes Care*. 1987;10:622-628.
3. 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;23:1143-1148.
4. Freckmann G, Schmid C, Baumstark A, Pleus S, Link M, Haug C. System accuracy evaluation of 43 blood glucose monitoring systems for self-monitoring of blood glucose according to DIN EN ISO 15197. *J Diabetes Sci Technol*. 2012;6:1060-1075.
5. Breton MD, Kovatchev BP. Impact of blood glucose self-monitoring errors on glucose variability, risk for hypoglycemia, and average glucose control in type 1 diabetes: an in silico study. *J Diabetes Sci Technol*. 2010;4:562-570.
6. Clarke WL, Cox DJ, Gonder-Frederick LA, Julian DM, Kovatchev BP, Young-Hyman D. The bio-psycho-behavioral model of severe hypoglycemia II: self-management behaviors. *Diabetes Care*. 1999;22:580-584.

7. Kost GJ, Tran NK, Abad VJ, Louie RF. Evaluation of point-of-care glucose testing accuracy using locally-smoothed median absolute difference curves. *Clinica Chimica Acta*. 2008;389:31-39.
8. Kost GJ, Tran NK, Louie RF, Gentile NL, Abad VJ. Assessing the performance of handheld glucose testing for critical care. *Diabetes Technol Ther*. 2008;10:445-451.
9. Kost GJ, Tran NK, Singh H. Mapping point-of-care performance using locally-smoothed median and maximum absolute difference curves. *Clin Chem Lab Med*. 2011;49:1637-1646.
10. Truong AT, Louie RF, Vy JH, et al. Effects of humidity on foil and vial packaging to preserve glucose and lactate test strips for disaster readiness. *Disaster Med Public Health Prep*. 2014;8:51-57.
11. Lam M, Louie RF, Curtis CM, et al. Short-term thermal-humidity shock affects point-of-care glucose testing: implications for health professionals and patients. *J Diabetes Sci Technol*. 2014;8:83-88.
12. Louie RF, Ferguson WJ, Sumner SL, Yu JN, Curtis CM, Kost GJ. Effects of dynamic temperature and humidity stresses on point-of-care glucose testing for disaster care. *Disaster Med Public Health Prep*. 2012;6:232-240.
13. International Organization for Standardization. Requirements for blood glucose monitoring systems for self-testing in managing diabetes mellitus. 15197:2003.
14. Boyd JC, Bruns DE. Quality specifications for glucose meters: assessment by simulation modeling of errors in insulin dose. *Clin Chem*. 2001;4:209-214.
15. Boyd JC, Bruns DE. Monte Carlo simulation in establishing analytical quality requirements for clinical laboratory tests: meeting clinical needs. *Methods Enzymol*. 2009;467:411-432.
16. Karon BS, Boyd JC, Klee GG. Glucose meter performance criteria for tight glycemic control estimated by simulation modeling. *Clin Chem*. 2010;56:1091-1097.