Future Science
OA

Impact of measurement error on predicting population-based inpatient glucose control

George E Saulnier[1], Janna C Castro[1] & Curtiss B Cook*,[2]

[1]Department of Information Technology, Mayo Clinic, Phoenix, AZ 85054, USA
[2]Mayo Clinic Hospital, Phoenix, Arizona, & Division of Endocrinology, Mayo Clinic, Scottsdale, Arizona, AZ 85259, USA
*Author for correspondence: cook.curtiss@mayo.edu

Aim: Instrument measurement error (ME) may affect ability of damped trend analysis to forecast inpatient glycemic control. **Materials & methods:** A statistical approach was developed to introduce ME into damped trend analysis algorithm. Point-of-care blood glucose device data were extracted from the laboratory system. Forecasts were generated from various inpatient subgroups and time intervals. **Results:** ME produced differences in damped trend model during the forecast learning cycle. However, forecast trajectory stayed identical regardless of ME in 85% (119/140) of studied scenarios. Forecasts did not change with greater ME. **Conclusion:** ME inherent in the point-of-care blood glucose device had little effect on trajectory of damped trend exponential forecasts and apparently would not influence decision making in inpatient glycemic control algorithms.

Lay abstract: High blood glucose (sugar) levels can lead to complications for hospitalized patients, including more surgical infections or longer hospital stays. The ability to forecast glucose control through trend analysis could identify problems sooner and allow earlier care to keep levels in the recommended range. However, measurement error (ME) is inherent in the glucometer used to check point-of-care glucose values and could limit the usefulness of forecasting methods. This study examined how ME affects forecasting. It showed little effect on glucose forecasts and showed potential robustness of trend analysis in assessment of inpatient glucose control.

Current guidelines recommend optimization of glucose levels in the hospital. This consensus is based on data showing that inpatient hyperglycemia and hypoglycemia are associated with worse patient outcomes, such as additional wound infections, longer length of hospital stay and higher mortality rates. In most cases, an inpatient target glucose range of 140–180 mg/dl is believed to represent the best balance for avoidance of complications associated with extreme high- and low-glucose events [1–3]. Management of inpatient blood glucose levels within this clinically recommended range reduces the possibility of adverse events associated with glycemia outside the range. Increased appreciation of the role of inpatient glucose control on patient outcomes has driven extensive study and discussion about which methodologies should be used to analyze and report the quality of hospital glycemic control [4].

Current approaches to analysis and report of inpatient glucose data – so-called glucometrics – are limited because the analyses rely on retrospective data [5]. Results of these analyses inform a hospital of where the data have been but not necessarily their future state. At the population health level, ability to forecast inpatient glycemic control could provide an opportunity to anticipate unfavorable changes before they become a problem. For the purposes of this study, the unfavorable changes would mean a drift toward glycemic control outside the accepted bounds. Unfavorable predictions could be forecast from historical data potentially, and interventions could be introduced earlier.

Operational research reports have described well the damped trend exponential smoothing statistical models that forecast trends derived from time series data [6–10]. These models have been used in commerce. The authors have

FUTURE
SCIENCE

> **Box 1. Overview of study methodology.**
>
> **Step 1. Optimize the dataset**
> - Define a forecast interval
> - Identify a learning dataset
> - Run a damped trend model
>
> **Step 2. Simulate measurement error**
> - Identify error of the measuring device
> - Calculate minimum simulation run
> - Conduct the simulation
> - Select the simulation run with the greatest AVE
>
> **Step 3. Assess the impact of error**
> - Solve the damped trend model with AVE upper bound as goal
> - Aggregate and graph the results
> - Superimpose the results
> - Assess for visual discrepancies
>
> AVE: Absolute value error.

reported previously on the feasibility of forecasting inpatient glucose data with use of this statistical approach [11]. In their report, they described the potential superiority over simple linear models in the projection of a future glycemic state.

The data used to perform the previous damped trend analysis were derived from point-of-care blood glucose (POC-BG) data [11]. In addition to measurement error (ME) inherent in the glucometer, the inpatient POC-BG results are potentially affected by other factors, such as substances that interfere with accuracy [4]. On an individual patient level, POC-BG ME in the hospital has the potential to lead to inappropriate changes in medical therapy (e.g., increased or decreased insulin value when not indicated). ME in POC-BG testing may have implications at the population health level. How can this error be generalized from a sole patient or result to decisions about hundreds of patients over long periods? In their previous forecasting paper, the authors did not evaluate the effect of ME on the outcomes of damped trend analyses [11]. Therefore, the objectives of the present analysis were twofold: determination of the possible effect of POC-BG ME on inpatient glycemic control forecasts and determination of whether the error affects a forecast enough that a decision to act (or not act) is influenced.

Methods

Statistical approach

A three-step process was used to assess the effect of ME on inpatient glucose forecasting ability (Box 1). The first step is the optimization of the original inpatient glucose dataset. The parameters that minimize absolute value error (AVE) are calculated, the forecast determined, and graphs generated accordingly. These results represent the non-ME optimal solution, presented in a previous paper [11].

The second step requires that the ME be simulated. In the third and final step, impact of error is evaluated. Every parameter combination is recorded that yields an AVE less than or equal to the upper bound. An assessment for visual discrepancies in the forecast interval is run, and where discrepancies exist, the ME affects the forecast.

Description of ME types

POC-BG measurements were performed with approved standard glucometers. Per vendor specifications, results must be ±15 mg for glucose values less than 75 mg/dl or ±15% when more than 75 mg/dl. Accuracy, linearity and precision tests are verified on new meters and revalidated twice yearly. For determination of the ME in the present study, the smallest error interval (±15%) was selected such that at least 95% of the acquisition method results are within the range. This error was determined from the vendor information on the glucometer used during the period analyzed. An average accuracy was then calculated and converted to a z-score derived from a standard normal distribution. In addition to the ME, two other error types were relevant in this analysis: the AVE and the mean absolute percent error (MAPE). Both types measure the cumulative distance between each point on the model and the corresponding observed result. AVE evaluates the parameter combinations that optimize the model against a learning dataset; MAPE measures after the fact how closely the observations matched the forecast.

The largest AVE, obtained for a given scenario across multiple simulations, represented an upper bound of error between the observed measurements and the model. Assessment of the effect of ME on forecasts involved a rerun of the non-ME model and identification of all parameter combinations that yielded an AVE less than or equal to the upper bound.

Two characteristics define the results of a model: the optimized model parameters (α, γ and φ) and the minimized AVE that identified them. The optimized parameters are unique to the set of random values (RVs) used in the simulation and therefore are relevant only to that simulation. The AVE, by comparison, is a relative measure of the efficacy of the model to represent the underlying data (i.e., the smaller the AVE, the closer the fit). Over a set of simulations, the largest AVE defines the least effective optimal solution for that ME level and depicts a worst case of model error. Investigators have proposed that relaxation of the constraints of the deterministic model to assume suboptimal solutions – to the degree defined by the worst case AVE – represents a unique depiction of the effect of ME on forecasts.

Data acquisition

The authors extracted POC-BG data from the laboratory information system, as previously described [11]. To more comprehensively assess the impact of ME impact, the authors expanded the analytic time frames to encompass two additional years' of data (2009 through 2015 vs 2010 through 2014). A total of 91,928 patient-day-weighted mean (PDWM) measures were input to the model and were grouped and averaged by week. Datasets were defined such that 48-week damped trend exponential forecasts began at the start of every quarter. In addition to the overall patient population not in the intensive care unit (ICU) previously studied (55,328 PDWM data points), four smaller populations were considered in the present analysis: non-ICU medical-surgical (n = 42,060 PDWM data points), non-ICU telemetry (n = 30,075), intermediate care (n = 8568) and ICU (n = 6214). The study of ICU patients and non-ICU patients, as well as subpopulations of non-ICU patients, over the longer time frame represented an expansion of scope over the authors' previous work. It also permitted a more detailed investigation into the effects of measurement error. Quarterly forecasts from data across 7 years in five populations resulted in assessment of 140 scenarios.

Analysis

To calculate each forecast, a learning dataset was compiled. It consisted of between 58 and 63 weekly data points that directly preceded the forecast. The size of the learning dataset was determined by data availability. As in the previous study [11], AVE was the decision criterion used to evaluate the fit of the model to the underlying data. The α, γ and φ parameters combination that yielded the smallest AVE between observations and the model was selected as the optimal solution.

After the optimal solution was identified, ME was introduced into the PDWM in two arbitrary ways. First, it was introduced as a uniformly distributed RV in an interval of ±15% of the PDWM. Second, it was introduced as a normally distributed RV with a PDWM mean and standard deviation extracted from the glucometer documentation. Each PDWM result was allowed to vary randomly within these established limits.

The number of simulation runs (n) required to achieve a desired confidence level was calculated using the following equation [12]:

$$n = \left(\frac{100 * z_{95} * S_x}{E * \overline{x}} \right)^2$$

where,

$z_{95} = standard\ normal\ z-score\ for\ 95\%\ confidence$

$S_x = standard\ deviation\ of\ the\ learning\ data\ set$

$$E = half - interval\ width$$

$$\bar{x} = mean\ of\ the\ learning\ data\ set$$

Glucose data at the authors' hospital, especially in large populations with tightly managed glucose control, are subject to small variability. Additionally, the general trending nature of the data over the long term is flat (neither increasing nor decreasing) [11]. For this reason, a large half-interval width (half of ±15% is 15) combined with a small dataset standard deviation (~6.29) and a large sample mean (159.3) yielded a remarkably small number of simulations (typically between 1 and 8 runs) for a confidence limit of 95%.

Despite the small number of required simulations, an initial arbitrary decision was made to run the 140 scenarios with an n of 30 (15 uniformly distributed runs and 15 normally distributed runs). The average duration of each simulation was approximately 4 min. With a 140-scenario run time of 280 h, the n of 30 appeared to strike a balance between a reasonably short processing time and a fully representative large-run simulation. For the purposes of the present analysis, three of the quarters analyzed in the previous paper were revisited, this time assessing the impact of ME [11].

A computer program was written to incorporate a multiple-run Monte Carlo simulation into the damped trend forecasting model. Each simulation generated either a uniformly or a normally distributed RV dataset. The simulated datasets were constructed within the limits of error of the glucometer. Hence, each simulated solution, with its own minimized AVE, represented a potential outcome in the overall set of possible solutions.

Results

To understand the impact of ME on forecasts, the authors sought discrepancies between the non-ME solution and the results from the various simulations. Of the 140 scenarios observed, 119 (85%) showed no visible differences: Both the non-ME solution and the simulation depicted identical results. In the few scenarios where discrepancies were evident, root causes other than ME, such as sparse data or poor model fit, were attributed.

To evaluate the initial hypotheses, three examples from the authors' previous paper were revisited [11]. In that analysis, the examples highlighted the differences between use of a damped trend model and use of a linear regression to forecast future glucose values. In this analysis, the authors retained the damped trend and replaced the linear regression with a second forecast that incorporated the simulated ME. The results section concludes with a more detailed analysis of the effects of an increase in the percent ME.

Example 1

The authors report the results of a forecasting analysis for POC-BG data derived from observations from 2010 (Figure 1). They show the observed (actual) preforecasted PDWM glucose values, the damped trend prediction intervals and a simulated forecast based on the error of the measuring device (Figure 1A). In particular, the graph depicts the forecast at the start of the 48-week period without the benefit of observed values from periods 63 through 111. This illustration represents information that one might receive at the start of the forecasting period. A retrospective view of the forecast was obtained after the observed values were known (Figure 1B). The view of the 48-week forecast was expanded where the overlap of the damped trend and the simulated ME forecast was evident (Figure 1C). During the learning cycle, the damped trend not based on the ME depicted a slightly increasing progression that, beyond period 63, extended into a flat forecast. During the learning cycle, the ME-based simulation more closely followed the increasing or decreasing, or both, patterns of the underlying data, and it, too, projected the same flat forecast as the non-ME results. For both ME and non-ME forecasts, the MAPE (3.140%) was the same as reported previously [11].

Example 2

The POC-BG results from the second quarter of 2012 are illustrated, showing data forecasted for 48 weeks (Figure 2). Observed preforecasted PDWM values were known (Figure 2A). The graph depicts the forecast at the start of the 48-week period without the benefit of observed values from periods 63 to 111. The retrospective view after the observed values was also known (Figure 2B), and the authors created an expanded view of the 48-week forecast (Figure 2C). The difference in this example was in the similarities between the ME and the non-ME damped
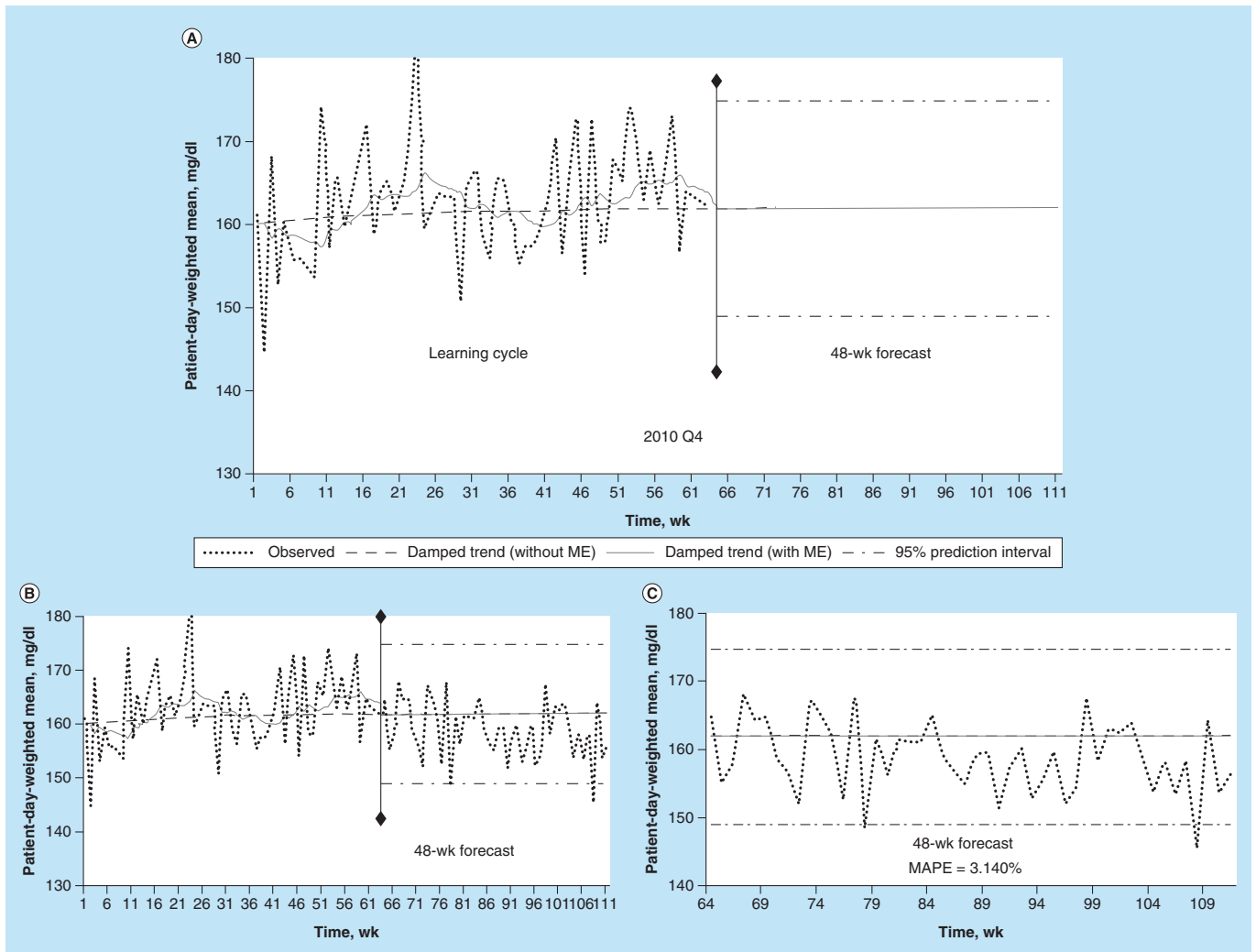
Figure 1.   Damped trend forecast for 2010 with measurement error. (A) The 48-week damped trend forecast with ME not in the intensive care unit, without observed data for the time period. **(B)** The same comparisons as shown in A with the overlay of observed values. **(C)** The 48-week forecast depicting the overlap of damped trend and ME forecasts at a glucose value of 161.9 mg/dl. Each observation is the aggregation of a 1-week period.
MAPE: Mean absolute percent error; ME: Measurement error.

trend learning cycles. Whereas the non-ME trend depicted a smoother progression during weeks 1 through 62, both cycles nonetheless generally followed the patterns present in the underlying data. Once again, the forecasts were identical (MAPE = 2.667%), with 1 forecast superimposed over the other (Figure 2C).

Example 3

The final example (Figure 3) illustrates the results of a trending analysis for second quarter 2014. Observed preforecasted PDWM values are shown (Figure 3A). The graph depicts the forecast at the start of the 48-week period without the benefit of observed values from periods 63 to 111. The retrospective view was created after the observed values were known (Figure 3B), with an expanded view of the 48-week forecast (Figure 3C). Similar to Example 2, the non-ME and ME learning cycles mimicked the underlying data, again with a less extreme progression evident in the non-ME trend. As anticipated, the 48-week ME and non-ME forecasts were the same. As reported previously, the MAPE for both forecasts is 3.423%.

Figures 1 through 3 are representative of what was observed most frequently in the present analysis. Despite a degree of variability in the learning cycle between the ME and non-ME models, the forecast trajectories were most often identical.
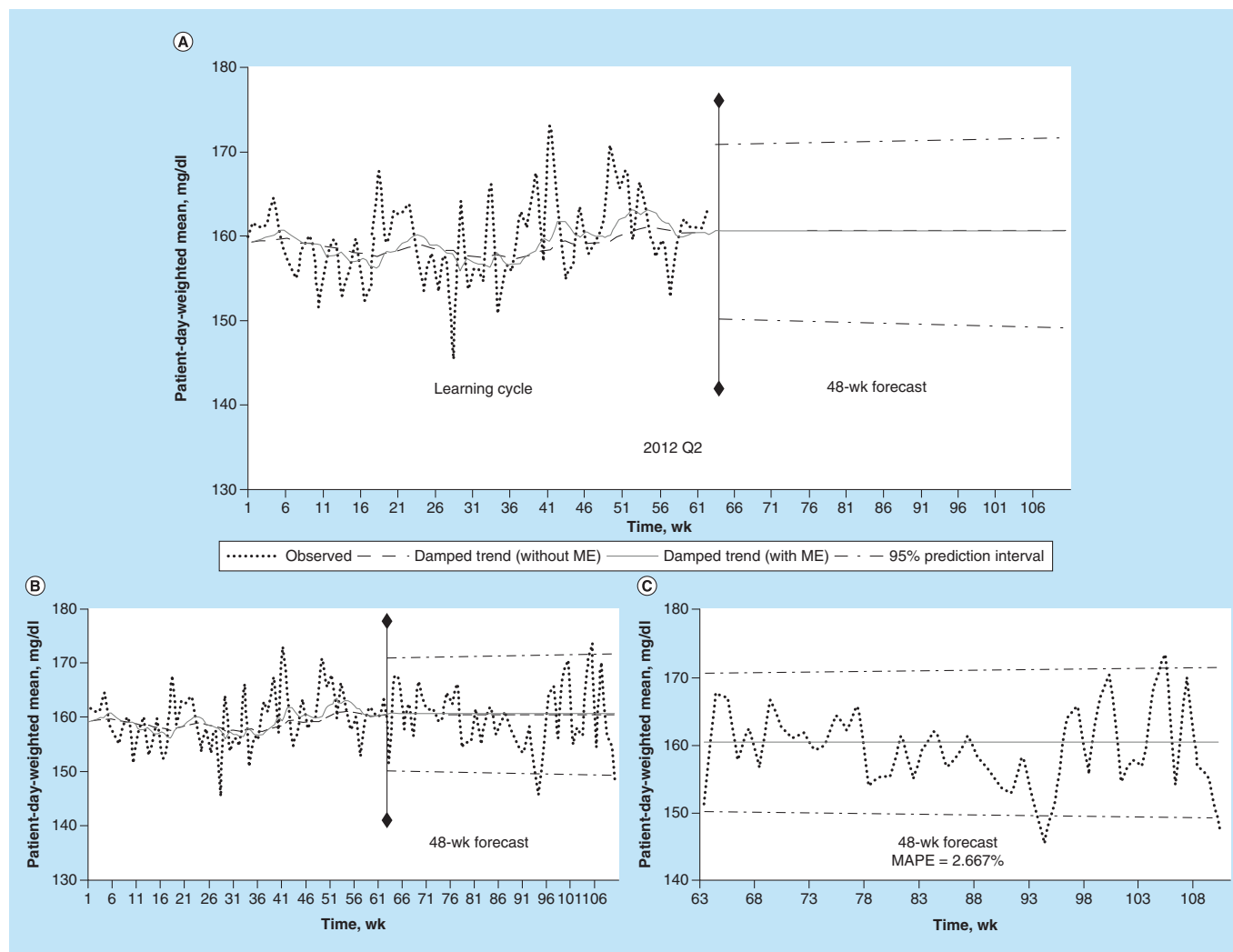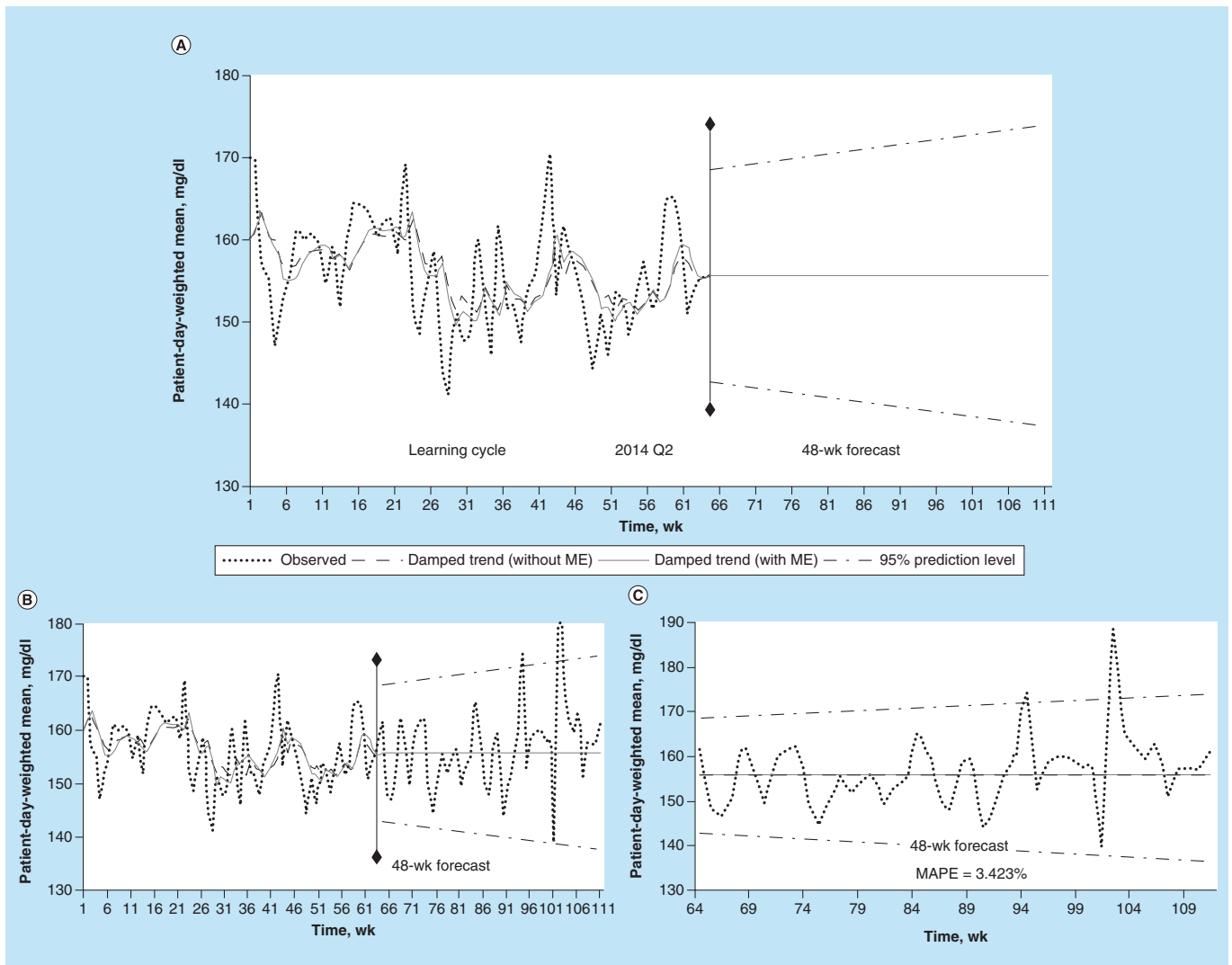
Figure 2.   Damped trend forecast for 2012 with measurement error. (A) The 48-week damped trend forecast with ME not in the intensive care unit, without observed data for the time period. **(B)** The same comparisons as shown in A with overlaying of observed values. **(C)** The 48-week forecast depicting the overlap of damped trend and ME forecasts at a glucose value of 160.4 mg/dl. Each observation is the aggregation of a 1-week period.
MAPE: Mean absolute percent error; ME: Measurement error.

Impact of an increasing ME

To better show the degree to which larger MEs might affect predictions, the magnitude of error for the second quarter 2014 scenario (Figure 3) was arbitrarily increased from 15 to 45% and 75% (Figure 4). The thick black line is the 15% ME result depicted in Figure 3. At an ME of 45% (the medium black line), the variability of the results during the learning cycle increased such that the peaks and valleys were visibly more pronounced. At 75%, the variability of results (the thin black line) during the learning cycle was even greater than at an ME of 45%. However, an increased ME and a rerun of the simulation had no effect on the 48-week forecast.

Discussion

Concern is warranted about the importance of ME in glucose results. Since too high or too low a value might steer quality improvement efforts in an unnecessary (or even incorrect) direction, ensurance that a measurement is as accurate as possible is the goal. Frequently, however, in cases where a result is suspect, retesting ensures the use of the most accurate information possible. For that reason, at the patient level, ME most likely presents itself as an opportunity to validate a previously recorded result. In other words, when in doubt, try again.

Figure 3.   Damped trend forecast for 2Q 2014 with measurement error. (A) The 48-week damped trend forecast with ME not in the intensive care unit, without observed data for the time period. **(B)** The same comparisons as A with overlaying of observed values. **(C)** The 48-week forecast depicting the overlap of damped trend and ME forecasts at a glucose value of 155.6 mg/dl. Each observation is the aggregation of a 1-week period.
MAPE: Mean absolute percent error; ME: Measurement error; Q2: Second quarter.

The forecast trajectory may be sufficient to influence the decision to extend or to cease a treatment protocol. Inpatient glycemic control is a result of a complex interaction between many variables. These include medications (e.g., use of glucocorticoids), varying types and states of nutritional status, and alterations in nutritional status. How an institution and its staff view the priority inpatient glycemic control determines the resources dedicated to its management (e.g., employment of dietitians or diabetes nurse educators). The presence or absence of glycemic control protocols, including the recognition and the treatment of hypoglycemia, and the staff experience in implementing them could influence inpatient glycemic outcomes. Either intentional or unrecognized alterations in these or in any variable could have an effect on glucose control that could be detected with implementation of forecasting methodologies. Added to this list of factors that could influence the trajectory of a forecast is the accuracy of tests used to derive the forecast. At a population level, the impact of ME is less clear.

To learn more about how ME affects the forecast of population-based glucose results, the authors studied 140 scenarios spanning 7 years of data and encompassing more than 90,000 PDWM values. A 15% ME was introduced into the data. Optimization models were run with the simulated results, applying a worst-case error methodology
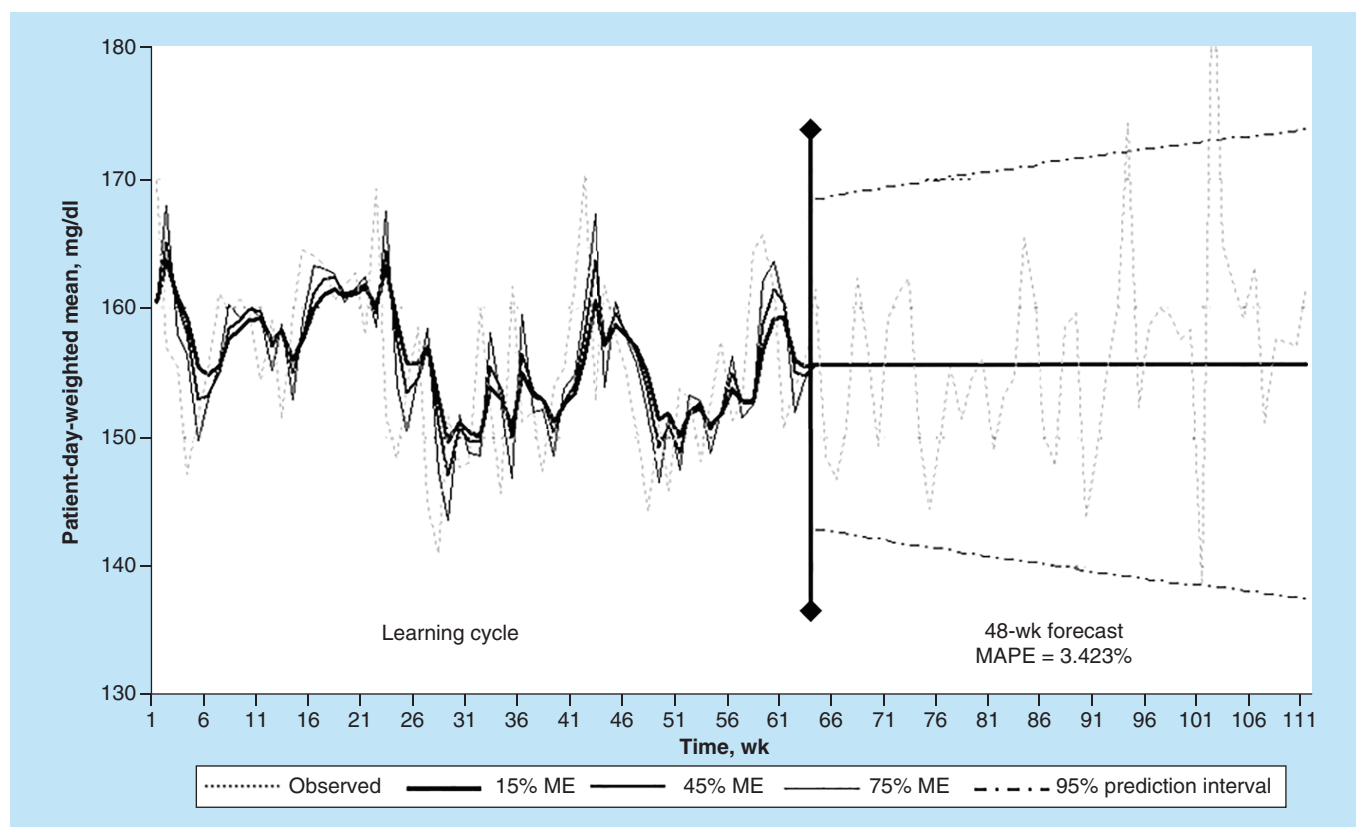
Figure 4.   The effect on forecasting of increasing the measurement error from 15 to 45% and 75%. Despite a greater fluctuation of values during the learning cycle, the 48-week forecasts stayed the same at a glucose value of 155.6 mg/dl.
MAPE: Mean absolute percent error; ME: Measurement error.

to recalculate forecasts. In the end, however, the authors saw no consistent evidence to suggest that forecasts were influenced by the ME defined in their analysis.

The authors incorporated ME into three examples from an earlier analysis. This study included data from a population of patients from the ICU. These patients, by definition, are more acutely ill. Management of hyperglycemia in the authors' institution typically differs for the ICU patients compared with non-ICU patients. Glucose management in the ICU most often involves use of an insulin infusion versus the subcutaneous insulin administered in the non-ICU setting. Despite differences in these two patient population types, the forecasting methodology showed comparable conclusions for both populations. These examples reflected what the authors observed for 85% of the scenarios analyzed: Trajectories for ME and non-ME forecasts were the same.

The authors also observed that while an increase in ME appeared to increase learning cycle variability, the forecasts themselves were not influenced by the greater degree of error, a result indicating the robustness of this approach to predict the future state of glycemic control. Conceivably, the reason for these conclusions may be evident in Figure 4. The arbitrary selection of uniform and normal distributions to model the randomness of error may have affected the analysis. Both were symmetrical about their means, and this factor may have been limiting. With either distribution, an increase in ME produced an increase in the variability of the results. However, the variability did not fundamentally change the underlying structure of the data, at least insofar as the damped trend model was concerned. Even when peaks and valleys were more pronounced, they were more pronounced equally on either side of their means. If the general patterns within the data are the same, then the optimal parameters of the model are the same, and the forecasts will be the same.

One limitation of this analysis is that the data were from a single institution using 1 type of POC-BG device. Before the generalizability and usefulness of the approach can be established, data should be evaluated from other inpatient facilities that might use a different vendor for POC-BG testing. In another limitation, although every glucose measure cannot be regarded as independent (i.e., each glucose result may depend on how the previous result

was treated), any autocorrelation that may exist in the data would not adversely affect the modeling of trends [6]. Another reason that also may account for these results is the smoothing effect of grouping PDWM values by week, which may have masked a larger impact of error. Finally, it should be understood that this is a population-based analysis. Patient level data were not considered, so conclusions regarding individual behavior cannot be made. More relevant factors of influence may exist in the trajectory of forecasts than ME alone.

Conclusion

The results are suggestive that the damped trend forecasting model may be robust against the ME inherent in the POC-BG device used by the authors' institution. Evidence indicated that this ME had little effect on the trajectory of damped trend exponential forecasts and has no apparent influence on decision making with regard to inpatient glycemic control algorithms. Although the magnitude of ME increased the variability of results, the analysis suggests that this variability did not visibly affect forecasts.

Future perspective

The greatest usefulness of damped trend analysis may be to assess the impact of an intervention geared to affect inpatient glycemic control processes. Next steps would be to identify institutional practices that might alter inpatient hyperglycemia management. For instance, the authors' institution transitioned to a new electronic health record platform in October 2018, which required staff retraining on all processes, including how insulin treatment algorithms are presented. Forecasting methods could be used to determine whether this new system is projected to affect glycemic control. Future research focused on the details that were overgeneralized in the present analysis may glean new insight. Refinement of analyses to specific geographic locations in the hospital (e.g., examination of different floors, examination of areas adjacent to one another) could assist in the determination of whether 1 area is forecasted to have better or worse glycemic control. Additionally, statistical questions remain. For example, should ME be modeled symmetrically about the mean? Statistical distributions may be worth consideration with skew parameters that could model bias. Finally, should PDWM values be grouped in intervals smaller than weekly? Did data centering occur because of the authors' choice of grouping?

Summary points

- Damped trend analysis may be useful in predicting future inpatient glucose control.
- The authors have previously published an analysis showing the feasibility of forecasting inpatient glycemic control.
- The previous study did not account for measurement error (ME).
- The goals of this analysis were the following:
    - Determine the possible impact of the point-of-care blood glucose (POC-BG) ME on inpatient glycemic control forecasts.
    - Establish whether this error effects a prediction to the extent that a decision to act is influenced.

Methods
- An algorithm was developed to assess the impact of ME on forecasting inpatient glucose control.
- POC-BG data were extracted from laboratory information system.
- Forecasts with and without the presence of ME were constructed and compared.

Results
- Forecasts were identical regardless of the presence of ME.
- Forecasts were not influenced with greater degrees of error.

Discussion & conclusion
- The results suggest that the damped trend forecasting model is robust against the ME inherent in the institutional POC-BG device.
- Evidence indicates that ME inherent in the institutional POC-BG device has little effect on the trajectory of damped trend exponential forecasts and would have no apparent influence on decision making with regard to inpatient glycemic control algorithms.
- Although the magnitude of ME increases the variability of results, the analysis suggests that this variability does not visibly impact forecasts.

Author contributions

Financial & competing interests disclosure

Open Access

References

Papers of special note have been highlighted as: ● of interest

1.   Umpierrez GE, Hellman R, Korytkowski MT *et al.* Management of hyperglycemia in hospitalized patients in non-critical care setting: an endocrine society clinical practice guideline. *J. Clin. Endocrinol. Metab.* 97(1), 16–38 (2012).

2.   American Association of Clinical Endocrinologists. AACE Diabetes Resource Center. (2019) http://resources.aace.com/

●   **Reviews current guidelines for inpatient glycemic control and management.**

3.   American Diabetes A. 14. diabetes care in the hospital. *Diabetes Care* 40(Suppl.1), S120–S127 (2017).

4.   Thompson BM, Cook CB. Glucometrics and insulinometrics. *Curr. Diab. Rep.,* 17(12), 121 (2017).

5.   Goldberg PA, Bozzo JE, Thomas PG *et al.* 'Glucometrics' – assessing the quality of inpatient glucose management. *Diabetes Technol. Ther.,* 8(5), 560–569 (2006).

6.   Everette S, Gardner J, Mckenzie E. Forecasting trends in time series. *Manage. Sci.* 31(10), 1237–1246 (1985).

7.   Makridakis S, Hibon M. The M3-Competition: results, conclusions and implications. *Int. J. Forecast.* 16(4), 451–476 (2000).

8.   Bowerman BL, O'Connell RT, Koehler AB. *Forecasting, Time Series, and Regression: An Applied Approach* Thomson Brooks/Cole, CA, USA (2005).

9.   Fildes R, Nikolopoulos K, Crone SF, Syntetos AA. Forecasting and operational research: a review. *J. Oper. Res. Soc.* 59(9), 1150–1172 (2008).

10.   McKenzie E, Gardner ES. Damped trend exponential smoothing: a modelling viewpoint. *Int. J. Forecast.* 26(4), 661–665 (2010).

11.   Saulnier GE, Castro JC, Cook CB, Thompson BM. How is the weather? Forecasting inpatient glycemic control. *Future Sci. OA* 3(4), FSO241 (2017).

●   **First paper demonstrating application of damped trend analysis in forecasting inpatient glycemic control.**

12.   Driels MR, Shin YS. Determining the number of iterations for Monte Carlo simulations of weapon effectiveness. (2004). https://calhoun.nps.edu/handle/10945/798