An important question about blood glucose control in diabetes is, Can present and future blood glucose values be predicted from recent blood glucose history? If this is possible, new continuous glucose monitoring technologies under development may lead to qualitatively better therapeutic capabilities. Not only could continuous monitoring technologies alert a user when a hypoglycemic episode or other blood glucose excursion is underway, but measurements may also provide sufficient information to predict near-future blood glucose values. A predictive capability based only on recent blood glucose history would be advantageous because there would be no need to involve models of glucose and insulin distribution, with their inherent requirement for detailed accounting of vascular glucose loads and insulin availability. Published data analyzed here indicate that blood glucose dynamics are not random, and that blood glucose values can be predicted, at least for the near future, from frequently sampled previous values. Data useful in further exploring this concept are limited, however, and an appeal is made for collection of more. Diabetes 48:445–451, 1999

To a great extent, the current treatment for diabetes is based on a programmed open-loop strategy for management of glycemia. Medications are administered largely on the basis of best estimates of the amount and timing of anticipated caloric absorption, activity levels, and medication effects. The various difficulties associated with this programmed open-loop approach are well known to individuals with diabetes, including variable rates and amount of caloric absorption, interruption of schedules and unanticipated activity demands, unpredictable rates of insulin absorption from injection sites, miscalculation of needs, dosages, and caloric content of foods, etc. There is rarely an opportunity to apply true closed-loop feedback control, in which action to counter glycemic imbalances is taken only after the initial stage of a blood glucose excursion is detected, even though experimental closed-loop feedback control has been demonstrated repeatedly to maintain blood glucose within normal physiological levels (1,2).

The obvious reason that programmed open-loop control is the current treatment modality is that present blood glucose monitoring methods based on blood collection by finger sticking are not automatic, and cannot realistically be performed frequently enough to reliably detect the early stages of blood glucose excursions in time for immediate corrective action. As a result, glycemic disturbances often go undetected, including severe hypoglycemia that occurs during the day before symptoms are recognized (3,4) or at night without recognition of symptoms (5,6).

There is much interest in the development of new sensing technologies that can continuously indicate glucose concentration (7). Sensors based on various electrochemical, optical, and other principles are being developed that may provide a continuous or very frequent initiative-independent indication of blood glucose. When successfully implemented, new glucose sensors with these capabilities should lead to qualitatively improved therapeutic strategies. An obvious example is the application of new sensor technology to alert the user at the early stages of hypoglycemia. This would allow immediate corrective feedback action to preempt severe hypoglycemia. In addition, however, another operational mode may be feasible. If the recent blood glucose history is not random but has an exploitable structure, it may be possible to anticipate blood glucose values in the near future based only on previous values. This could lead to even further improvements in blood glucose regulation. In this Perspective, we examine the intrinsic short-term structure of blood glucose dynamics from data in the literature collected with conventional methods of glycemia determination and demonstrate the predictive capability of blood glucose monitoring.

BACKGROUND

Experimental electrochemical glucose sensors implanted in human subcutaneous tissues have been reported to occasionally produce transient signals that precede blood glucose values by several minutes (8,9). These apparent anticipatory responses were observed in the downward phase only of certain recorded hypoglycemia transients and were retrospectively distinguished from more common transient signals that lagged blood glucose values. The possibility that signals from implanted subcutaneous sensors may slightly precede hypoglycemic blood glucose values was considered a potential advantage of the monitoring approach. However, this phenomenon may be only indirectly related to blood glucose concentration, and there is a variety of
Glycemic dynamics can also be nonstationary as a result of either extrinsic glycemic disturbances or physical changes in the intrinsic governing processes. An example of an extrinsic disturbance that may lead to nonstationary dynamics is typical eating behavior observed for a 1- to 2-day period. Physical changes in the intrinsic processes governing glycemic dynamics may be caused by sickness, exercise, stress, pregnancy, hormonal changes associated with menstruation, etc. If the observed process dynamics are nonstationary, the structure may vary in a complex fashion over time and may not be useful for making predictions. Observations indicating stationary processes were therefore sought.

By a widely accepted definition (22), a process is considered stationary when the first- and second-order averages of temporal measurements are invariant to a shift of the time origin. First-order averages include mean and variance, and second-order averages include the autocorrelation and autocovariance functions. These functions generate a sequence of coefficients that indicate the correlation between measurements separated by time (23). An autocorrelation function (ACF) is defined as the autocovariance function of a sequence normalized by the variance of the process (21,23). This normalization allows processes with different variances to be readily compared by respective ACF estimates. Using these definitions, a process can be considered stationary when the sample mean and variance of process measurements are constant and the ACF is independent of absolute time but dependent on the relative time between individual measurements (22).

**Interpretation of the ACF estimate.** An ACF estimate provides a statistical measure of the dependence between individual measurements of a process at different times. If the process is stationary, the sequence of ACF coefficients describes the statistical dependence between pairs of measurements separated by fixed time intervals throughout the recorded observations. Consequently, an ACF analysis can be used to show whether the observed process measurements are mutually dependent or independent (21). Identified statistical dependence implies that some predictable process structure exists. If the process observations have been and remain stationary, this statistical dependency can be exploited to predict future measurements.

To illustrate the use of an ACF, cases of contrived sequences of dependent and independent measurements are presented in Fig. 1. The sequence shown in Fig. 1A is comprised of mutually dependent measurements generated by a simple third-order difference equation (21), whereas the sequence shown in Fig. 1B is comprised of identically distributed random values that are mutually independent. Both sequences are offset to a mean of zero, as are subsequent process measurements. The ACF estimates for the dependent and independent data sequences are shown in Fig. 1C and D, respectively. For simplicity, only the right half of the complete ACF estimates is given. All ACF coefficient estimates for a stationary process have a value of unity at the relative lag or displacement of zero. The ACF estimate includes errors that depend on the number of observations included in the calculation of each ACF coefficient. A smaller number of total measurements leads to an increase in the variance of ACF coefficient estimates and subsequent predicted values. These estimation errors should be considered when ACF estimates are analyzed. To help assess estimation errors, the broken lines in all ACF plots represent the 95% CI bounds around zero. These bounds are commonly
An ACF estimate can be used to determine whether measurements in a sequence are mutually independent. For mutually independent measurements, the expected value of the product of two different zero-mean measurements is, by definition, zero (21). The sequence of ACF coefficient estimates in Fig. 1D demonstrates that the sequence in Fig. 1B is random, since no ACF coefficients, other than the zero lag, are significantly different from zero (22). (The single ACF coefficient outside the confidence interval in Fig. 1D is expected, since the 95% CI should allow ~1 in 20 coefficients that have a value not significantly different from zero to fall outside the confidence lines.) In contrast to Fig. 1D, several ACF coefficients at the early lags in Fig. 1C are significantly different than zero. The data in Fig. 1A may appear independent by simple inspection, but the ACF coefficient sequence demonstrates dependency and structure in the data.

In the present case, where the goal is to determine whether glycemic measurements are predictable, it is desirable not only to determine whether structure is directly observable in complete observation sets, but also to determine the level of predictability achievable with specific data analysis models. In conjunction, the applicability of the identified structure needs to be assessed. Analyzing data subsets may also be appropriate, as in the case of nocturnal glycemia, but this limits the data available for analysis and increases the variance of ACF estimates.

**METHODS**

Twenty-two data sets representing various states of glycemia were obtained from the literature. The sets contained frequently sampled nonaveraged blood glucose measurements. The data included measurements from nondiabetic subjects collected during continuous enteral feeding over an extended period.
A continuous Enteral Feeding, Moving Average B Non-Diabetic, Moving Average

FIG. 3. The 200-min moving averages of the respective blood glucose measurements in Fig. 2A–D. The shaded bands indicate the standard deviation of the individual data sets. A: Continuous enteral feeding (26). B: Nondiabetic, ambulatory, and eating on a typical schedule (27). C: Type 1 diabetic individual (IDDM), ambulatory and eating on a typical schedule (27). D: Type 1 diabetic individual (IDDM), ambulatory and eating on a typical schedule (27).

period (26) and from type 1 diabetic and nondiabetic subjects collected under conditions typifying daily activities (27). Blood glucose measurements representative of those obtained during continuous enteral feeding are shown in Fig. 2A. As before, data are offset to a mean of zero. Representative data sets of larger glycemia excursions caused by typical eating and exercise schedules obtained from normal and diabetic subjects are shown in Fig. 2B–D.

ACF coefficient sequences were estimated for each data set to determine whether there was structure in the measurements. The respective plots of normalized ACF coefficients are shown in Fig. 2E–H, with the 95% CIs around zero indicated by the broken lines and ± 2 SD indicated by the shaded region, as before. The sample means, variances, and ACF coefficient estimates were calculated from subsets of each complete data set and were assessed for time-invariance to determine whether identified structure in the glycemic measurements could be easily exploited. Simple constant coefficient linear autoregressive integrated models of varying system order were constructed from the first half of each data set (22). The coefficients of the predictor were selected by minimizing the squared error between the actual and predicted glucose values. Predictions of glycemic measurements were then made for the second half of each data set, with prediction horizons of one, two, and three steps ahead (i.e., 10, 20, and 30 min, respectively).

RESULTS

Do glycemic measurements have inherent structure? Estimates of ACF coefficients for each of the representative data sets of Fig. 2A–D are shown in Fig. 2E–H. Comparing the ACF estimates in Fig. 2 with the ACF estimate for the independent noise process in Fig. 1D demonstrates that there is significant statistical dependence between the individual glycemic measurements. The ACF estimates for all of the glycemic data analyzed showed similar marked statistical dependence between the individual measurements. However, the ACF coefficient sequences for many of the data sets showed a very slow decay to zero, as seen in Fig. 2G and H. This is an indication of a trend in the data sequence that results in a nonstationarity. If the data are nonstationary, then the statistical dependence implied by ACF estimates will be difficult to meaningfully interpret or exploit for prediction.

Which glycemic observations are stationary? For glycemic measurements to be predictable with a simple constant-coefficient model, the process measurements need to be stationary. Figure 3 shows a 200-min moving average of the glycemic measurements in Fig. 2. The shaded bands indicate the standard deviation of the individual data sets. Ideally, the resulting averages would be constant, confirming that the process measurements were stationary in the mean. However, the representations in Fig. 3B–D display substantial variation in the mean, with the degree of nonstationarity in the mean increasing from Fig. 3B to Fig. 3D.
The normalized ACF coefficients for subsets of the continuous enteral feeding and normal glycemic data shown in Fig. 3A and B are illustrated in Fig. 4A and B, respectively. In Fig. 4A, the complete data set was arbitrarily divided into two sequential subsets of equal duration to assess stationarity, and the ACF coefficients are shown as the filled circles. In Fig. 4B, the data set was divided into fed and nocturnal fasting periods. The ACF coefficients for fasting subsets are indicated as empty circles, and filled circles indicate the remaining nonfasting subsets. In Fig. 4A and B, the shaded band indicates ±2 SD around the ACF estimates computed from the complete nonfasting data set, which is the complete data set for Fig. 4A, but only a subset of the data in Fig. 4B. The ACF estimates of the two subsets of the continuous enteral feeding observations are not significantly different from the ACF estimate for the complete data set, indicating that the process is likely stationary. In contrast, the ACF estimates in Fig. 4B for the nocturnal subsets differ from those for the nonfasting subsets, indicating that the glycemic dynamics are nonstationary under conditions of typical meal ingestion over 2 days.

Can a simple linear model be used to make glycemic predictions? One and three step ahead (10- and 30-min) predictions and blood glucose measurements during continuous enteral feeding are shown in Fig. 5. The solid lines in Fig. 5 represent observations, and the broken line with the symbol indicates the predictions. Models were computed from the first half of the process measurements during the period of 0–10 h (not shown), and predictions were made in the second half of the data, during the period of 10–20 h. The difference between consecutive measurements was used to ameliorate the apparent nonstationarity in the mean for typical glycemic data and to develop all of the autoregressive integrated models for linear prediction. Figure 5 indicates that both the 10- and 30-min predictions follow the general course of the data, but that the 10-min prediction is clearly the better of the two. Further representative examples of 10-min one step ahead prediction based on data from nondiabetic and type 1 diabetic (group B) subjects are shown in Fig. 6. As in Fig. 5, the solid line corresponds to observations, and the dashed line with the symbol indicates predictions. This demonstrates that a simple linear model can be used to obtain close 10-min predictions from data representing a variety of glycemic states. For certain data (not shown), 20- or 30-min predictions may also be acceptable.

To quantitatively evaluate prediction performance, the root mean squared error is plotted in Fig. 7 as a function of the prediction horizon. The solid circles correspond to prediction horizons of 10–30 min. The dashed line indicates the mean value estimate, which can be interpreted as a limit of prediction horizon. As estimates are computed for increasing prediction horizons, the prediction error rapidly approaches the error that would be incurred by assignment of the mean value of past measurements as the estimate of future measurements. Certain other models (not shown) were less effective at accurate prediction, as indicated by higher error values at the respective horizons.

DISCUSSION

Various methods were used to assess process stationarity. The means for the entire observation period, as well as subsets of the observation period, were used to show whether the first-order averages of the analyzed processes were constant. The results shown in Fig. 4 indicate that the continuous enteral feeding observations had an approximately constant mean, whereas the data collected during more typical daily eating conditions did not have a constant mean. ACF coefficient sequences were estimated for subsets of the measurements
FIG. 6. Comparison of predictions and actual measurements of blood glucose from fed nondiabetic and type 1 diabetic subjects under ambulatory conditions. A: One step ahead (10-min) predictions for nondiabetic subject (Fig. 2B). B: One step ahead (10-min) predictions for subject with type 1 diabetes (Fig. 2D). Predictions are indicated by broken lines with symbols, actual measurements by solid lines.

collected during continuous enteral feeding and throughout the day of a nondiabetic subject with typical eating routine. The ACF estimates for the subsets of the data for continuous enteral feeding were not significantly different from the complete data set, as shown in Fig. 5. This indicates that the second-order averages are independent of the time origin. This fact, in conjunction with the observation that first-order averages were approximately constant, indicates that the continuous enteral feeding data are stationary. However, there were significant differences between the ACF estimates for the nocturnal and fed subsets during a typical day, confirming that the glycemic dynamics observed during a day with a typical eating pattern are nonstationary in the mean. The data collected from subjects with type 1 diabetes under ambulatory conditions with a typical eating pattern, such as that shown in Fig. 2C and D, are also nonstationary. This is evidenced by a nonconstant mean (Fig. 3C and D) and variance (not shown), as well as by the very slow decay of the respective ACF estimates, as illustrated in Fig. 2G and H. Using the first difference of the measurements ameliorated this degree of nonstationarity and allowed satisfactory prediction performance for the data sets with more typical extrinsic glycemic disturbances.

The statistical dependence in the glycemic measurements identified by the ACF estimates shown in Fig. 2 was exploited to construct linear models of varying system order. Predictions of the glycemic data 10-min in the future were computed to assess the linear predictability of the glycemic dynamics during small stationary disturbances and more typical glycemic disturbances. The average root mean squared error for predicting glycemic dynamics for a 10-min prediction horizon under small stationary disturbances and more typical conditions was 0.2 mmol/l. This degree of error is comparable to typical measurement error, and is considerably less than the error incurred by using the last measurement as an estimate of future measurements, which is a common basis for comparison.

Because of the limited number and size of data sets, the models selected to assess predictability were simple and linear. The success of linear models in predicting glycemic dynamics indicates that such models may be used in conjunction with frequent blood glucose measurements to provide an earlier warning of impending hypoglycemia than would be attainable by measurement alone. However, a simple linear model created...
from a single day of blood glucose measurements is not expected to be applicable for predicting blood glucose dynamics without being updated. Since the biological processes responsible for glycemic dynamics are nonlinear and nonstationary under many conditions, models that are more complex are expected to improve the predictor performance and extend the prediction horizon. However, these models will require more data collected over longer periods of time from more individuals than is currently available in the literature.

These methods have potential applicability in conjunction with new continuous or very frequent glucose sensors under development. If glucose sensor signals are confirmed to consistently lead (rather than lag) blood glucose, predictive methods may provide an even greater prediction horizon. If sensor signals are found to consistently lag blood glucose values, which is expected to be more common, predictive methods may advance the reported measurements closer to real time. It is of note that the methods described here may not be needed when continuous sensors are used as part of automatic insulin feedback systems, since equivalent features can be included in the controller.

It is clear that more data are needed. The present conclusions are based on a very limited number of individual experiments. Simple extrapolation of the methods to more general circumstances without greater experimental verification may not be prudent. Individual glycemia records, including information about extrinsic influences, should be collected from many diabetic subjects under a variety of conditions with frequent blood glucose measurements (every 2-20 min) over a period of several days or more to capture details of glycemic excursions. Repeated studies on individuals over a period of months may be useful to assess intrinsic physiologic changes. Investigators are asked to modify experimental protocols to provide the type of data that is useful here.

In conclusion, this study demonstrates new ways of analyzing glycemic data that may lead to insights for improved forms of treatment. However, the conclusions here are based on limited sets of data, since the type of data and data collection formats that are most useful for these studies are not commonly used. Information extracted from appropriate data may reveal new aspects of glycemia control.

ACKNOWLEDGMENTS
This work was supported by grants from the National Institutes of Health and the Juvenile Diabetes Foundation International.

REFERENCES