Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Biomed Inform. Author manuscript; available in PMC 2016 October 24.

Published in final edited form as:

J Biomed Inform. 2016 October ; 63: 259–268. doi:10.1016/j.jbi.2016.08.022.

Using uncertain data from body-worn sensors to gain insight into type 1 diabetes

Nathaniel Heintzman[a] and **Samantha Kleinberg**[b]

Nathaniel Heintzman: nheintzman@dexcom.com; Samantha Kleinberg: samantha.kleinberg@stevens.edu

[a]Dexcom Inc., San Diego, CA, United States

[b]Department of Computer Science, Stevens Institute of Technology, NJ, United States

Abstract

The amount of observational data available for research is growing rapidly with the rise of electronic health records and patient-generated data. However, these data bring new challenges, as data collected outside controlled environments and generated for purposes other than research may be error-prone, biased, or systematically missing. Analysis of these data requires methods that are robust to such challenges, yet methods for causal inference currently only handle uncertainty at the level of causal relationships – rather than variables or specific observations. In contrast, we develop a new approach for causal inference from time series data that allows uncertainty at the level of individual data points, so that inferences depend more strongly on variables and individual observations that are more certain. In the limit, a completely uncertain variable will be treated as if it were not measured. Using simulated data we demonstrate that the approach is more accurate than the state of the art, making substantially fewer false discoveries. Finally, we apply the method to a unique set of data collected from 17 individuals with type 1 diabetes mellitus (T1DM) in free-living conditions over 72 h where glucose levels, insulin dosing, physical activity and sleep are measured using body-worn sensors. These data often have high rates of error that vary across time, but we are able to uncover the relationships such as that between anaerobic activity and hyperglycemia. Ultimately, better modeling of uncertainty may enable better translation of methods to free-living conditions, as well as better use of noisy and uncertain EHR data.

Keywords

Diabetes Mellitus, Type 1; Monitoring, Ambulatory; Uncertainty; Causality

1. Introduction

Chronic diseases like diabetes, hypertension, and heart disease account for most U.S. deaths each year [1], and unlike acute illnesses they are primarily managed not by clinicians but by

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Correspondence to: Samantha Kleinberg, samantha.kleinberg@stevens.edu.

Conflict of interest

SK has no conflicts of interest to report. NH had no conflicts of interest at the time of research (i.e. study design, data collection), but now works for and holds stock in Dexcom, Inc., whose CGM devices were used in the research.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

patients themselves. Blood glucose management in people with type 1 diabetes mellitus (T1DM) imposes a particularly significant cost on time and attention. Patients make frequent (~hourly) decisions about the timing and dosing of insulin to manage their blood glucose with infrequent (~semiannually) feedback from clinicians while having to account for many factors that affect glucose including physical activity, meals, and illness. Managing blood glucose is key to avoiding secondary complications like stroke, but requires constant vigilance from patients with little external feedback and support.

Yet, patients now generate massive amounts of health data through tracking symptoms and disease progression, uploading sensor data into patient portals, and sharing information with one another through social networks like PatientsLikeMe [2]. These patient-generated data may fill in information gaps between medical visits and help engage patients but there have been concerns about data quality and how to integrate patient-generated data into care without overloading clinicians [3]. Tests with providers, though, found the primary question was whether the data provided actionable information [4].

Most work has focused on how these data can improve encounters with providers, but continuously-collected data from body-worn sensors and devices such as continuous glucose monitors (CGMs), insulin pumps, and activity monitors could be used as input to a patient-centered decision support system, which would reduce the burden placed on patients with diabetes and other chronic diseases. While CGMs allow continuous recording of glucose data, these and other sensors can have errors, and imperfect use in real-world settings leads to new challenges (e.g. samples contaminated by inadequate hand-washing, loss of connectivity). Further, not all sensors can be worn during all activities (e.g. swimming) or may be removed by a patient for other reasons, leading to missing data. These issues are present in many types of observational biomedical data, yet this uncertainty is rarely taken into account when analyzing the data (e.g. ICD9 codes, from the International Classification of Diseases, Ninth Revision, are often used as proxies for diagnoses), impeding the development of patient-centered technologies.

We aim to demonstrate the utility of data collected in free-living environments from consumer-grade sensors for finding causal relationships. We propose that these data may be able to reproduce findings from highly controlled studies (at lower cost), while also enabling observation of a wider variety of scenarios than observed in the lab (e.g. activity from running to catch a bus or athletic competition). While we aim specifically to infer causes to enable accurate prediction of glucose trends and effective interventions to correct unhealthy highs and lows, there is not yet a method that can weight different observations of a variable differently during inference (based on the certainty of measurements). We hypothesize, though, that this is critical for robust inferences from such data. That is, if we are sure that a particular measurement of glucose indicates hypoglycemia, that measurement should be weighted more strongly in finding causes of hypoglycemia than a measurement that has an equal chance of being within the normal range. Since errors and the values of glucose measurements are not evenly distributed throughout the range, and devices have significant error rates, ignoring this underlying uncertainty of the data can lead to false positives and false negatives. In many cases information on error rates is available, but is not routinely incorporated into inference. For example, studies evaluate accuracy of continuous glucose

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

monitors in various scenarios (e.g. [5,6]) and models of this error have been developed (e.g. [7,8]), leading to the ability to determine which data points may be untrustworthy but not yet the ability to use that information in causal inference.

To handle this, we extend a causal inference method to handle uncertain data, by separating the observation of an event or measurement of a variable's value from the underlying truth of what occurred. This ensures that conclusions drawn from highly uncertain measurements are given less weight than those drawn from reliable measurements and a completely uninformative variable will be treated as missing, instead of propagating errors. The methods are applied to data collected from 17 individuals with T1DM in free-living conditions, using a variety of body-worn sensors. We demonstrate that these patient-generated data can be successfully used to uncover some causes of unhealthy glucose excursions in people with diabetes – but only when uncertainty of the data is properly accounted for.

This paper makes two main contributions: (1) we extend causal inference methods to better handle uncertainty in observational data; (2) we present a novel set of free-living data from people with diabetes that is available for research use (https://idash.ucsd.edu). We demonstrate the utility of the method through rigorous comparison on simulated data and its successful application to the free-living data.

2. Background

2.1. Diabetes

Chronic diseases are rising in prevalence, necessitating tools that can provide feedback directly to individuals, in contrast to the clinician-centered decision-support paradigm. Diabetes in particular affects over 29 million people in the U.S. (over 9% of the population) [9] and the CDC estimates that if current trends continue, this will grow to 33% of the U.S. population by 2050 [10]. The annual cost of diabetes in the U.S., including medical costs and loss of work, was estimated at over $245 billion in 2012, a 41% increase over the 2007 estimate of $174 billion [11]. People with diabetes incur substantial out-of-pocket costs, impeding access to preventative medical services [12]. Diabetes puts patients at risk for many other diseases, and is the primary cause of chronic kidney disease [13], a leading cause of amputations and blindness [14], and a risk factor for heart disease and stroke [15].

T1DM is a chronic autoimmune condition characterized by an inability to produce insulin and is currently an unpreventable, incurable, disease requiring life-long insulin therapy. Since complications are primarily from long-term high or low blood sugar (hyper- or hypoglycemia), they may be preventable with better glucose management [16]. However this requires patients to frequently test their blood glucose and determine whether to administer insulin (to correct high glucose levels) or ingest carbohydrates (to counter falling glucose levels). While glucose has traditionally been measured before and after meals with fingerstick monitors, these discrete samples cannot provide feedback on trends. CGMs instead provide constant feedback on glucose levels, potentially enabling better management [17].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Artificial pancreas systems aim to create closed-loop systems, linking the CGM and insulin pump with control algorithms that regulate glucose without a patient's intervention. These systems have demonstrated increases in glucose within normal ranges (euglycemia) and reductions in hypoglycemia [18], but most tests are conducted overnight in hospitals when glucose control is simpler and they have not been routinely used in ambulatory circumstances where meals, physical activity, and daily life can confound results. Outpatient studies have primarily focused on demonstrating the safety of these systems [19]. The problem is further complicated by the input to the control algorithms: CGMs measure glucose in the interstitial fluid between cells, rather than in blood. This can lead to a delayed signal that is a shifted and transformed version of blood glucose [20], requiring strategies to accurately reconstruct it [21]. Further, while sensors do not have perfect accuracy (due to factors such as noise or calibration errors), understanding real-world performance and factors affecting it has been a significant area of work [5,6]. However, it remains to incorporate what is known about error rates into inference.

Data from body-worn sensors (e.g. heart rate and activity monitors) and mobile apps may provide a more complete picture of the causes of changes in glucose and fill in information gaps. Physical activity can improve glycemic control [22] and increase insulin sensitivity [23], but creates challenges in glucose management since its effects are a function of the activity context (training vs. competition), duration and intensity, and even time of day [24], and physical activity includes both formal exercise and things like running to catch a bus [25]. Recent articles have assessed activity monitors and phone apps in laboratory and free-living conditions, finding that step counts are accurate, but other activity measures may differ from research-grade devices [26,27]. However, it is unclear whether devices are more accurate at tracking changes from a given individual's baseline. More fundamentally, little work has been done to assess causes of hypo- and hyperglycemia in an integrated way in realistic settings (rather than understanding one piece of the puzzle at a time in controlled experiments). To this end, we show that consumer-grade sensors can be used to monitor factors that affect glucose, and with proper analysis can do so accurately enough to find causes of changes in glycemia and eventually, to better inform patients. Further, our publicly available dataset, collected in real-world environments, may enable researchers conducting lab-based studies to understand how these relate to uncontrolled environments and may facilitate computational researchers developing new algorithms to handle the challenges of these data.

2.2. Causal inference

Many types of uncertain observational data are used for research, including EHRs and patient generated data streams, but there have been few approaches that address the specific challenges of using these uncertain data to find causal relationships. This is critical for ensuring that interventions, such as to lower blood glucose, will be effective. Data mining methods have addressed uncertainty due to missing data with imputation of missing values [28,29]. Multiple imputation handles this by imputing multiple values for each missing instance, and aggregating results. However, this is addressing a somewhat different problem than the one we discuss here, which is uncertainty in observed values, and incorporating that information into causal inference. Other approaches exist for weighting multiple observed

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

measurements, such as when combining the results of many studies in a meta-analysis. Inverse-variance weighting, for example, gives less weight to variables or studies as their variance increases [30]. However, in our application variance in a measure is not necessarily a proxy for certainty, and this approach can't be used to assign different weights to different observations of a single variable within an individual (as variance cannot be calculated for each individual observation). Other methods exist for directly representing uncertainty, including probability intervals [31], possibility theory [32], and belief functions [33], but these have yet to be incorporated into causal inference from time series and generally assume one begins with a known structure. Thus, these methods do not yet address our task: causal inference from uncertain data.

While causal inference is necessary, current methods do not account for the types of uncertainty we face. We need be able to use that a variable's measurement at, say, 2:00 pm on a particular day is less certain than that at 8:00 pm when inferring causal structures. Variable and time-dependent error rates are available in many forms. In the case of glucose measurement, comparison of a CGM to a fingerstick monitor can uncover mismatches between values at specific times, and device error rates can provide prior information on both overall accuracy of a variable's measurement and factors affecting the accuracy of specific measurements (e.g. sensor used past recommended duration). We also cannot assume that overall the incorrect observations will average out, as errors may not be evenly distributed across either a variable's range or throughout time. For example, if a new sensor for a CGM is always inserted on Sunday, values on Sunday will always be less accurate than those on Monday, which will be an even greater problem if this coincides with the day a user always engages in vigorous exercise. Graphical model-based methods such as Bayesian networks [34,35] or dynamic Bayesian networks (DBNs) [36] aim to find probabilistic models representing a system's causal structure. Efforts to incorporate uncertainty into these methods mainly address uncertainty at the level of independence relations in static networks [37,38] or in other types of prior knowledge about a structure [39] rather than at the level of variables and individual observations during structure inference. The primary type of uncertainty faced here is when constraints are in conflict, and the methods enable inference to use the more reliable constraints in such cases. Thus, the uncertainty is in the model as a whole rather than in a specific data point. This is the difference between representing uncertainty in whether a particular common cause screens off its effects and whether specific instances of the common cause are trustworthy observations. These approaches also face challenges in inferring complex relationships and their timing. Granger causality [40] aims to determine whether one time series is predictive of another, but its more accurate multivariate form is too computationally complex to be used with many variables over many time lags, while the bivariate form may erroneously find relationships between effects of a common cause.

We are not aware of any methods for finding complex temporal relationships from data where variables have differing levels of uncertainty that may also differ across time. It is critical to incorporate this uncertainty to avoid biasing inference. For example, during periods of high uncertainty, an individual may make more measurements of their blood glucose. In a frequency-based approach, these measurements would then be overrepresented in the data, while incorporating uncertainty in individual observations enables these to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

contribute only a weak signal. Further, that these measurements have higher uncertainty can be automatically inferred based on the unusual measurement density. We build on the approach of [41], where each relationship has an associated time window (that can be inferred from the data) rather than a discrete time lag as with DBNs or Granger causality. In body-worn sensor data, even if the underlying relationship has a single lag, errors and gaps in measurements make it unlikely that this would be observed as such. In the approach we build on, relationships and their timing are inferred directly from the data so that one may begin by testing for relationships between all variables and elevated glucose in 10–100 min and ultimately find that vigorous activity leads to high glucose in 5–30 min with probability at least 0.3. We augment this approach to handle patient-generated data by incorporating uncertainty in measurements, which allows each observation of each variable to have an associated probability, rather than simply being true or false. Ultimately this approach can be used more broadly to represent uncertainty such as that of imputed values for missing data, diagnoses (single ICD9 code versus multiple pieces of evidence), and concepts extracted from free text.

3. Causal inference

Many causal inference methods take a probabilistic approach, with probabilities based on frequencies, so a conditional probability such as $P(e|c)$ is defined as the number of occurrences of $c \wedge e$ divided by the number of times c is true. Each observation has the same impact on the calculation – yet some variables are more error-prone than others and individual measurements may further differ. We now discuss how to use information on observation uncertainty during inference, so that when finding causes of high blood glucose, timepoints that are less likely instances of hypoglycemia will have less of an impact. We begin with a brief overview of the inference method being extended, then discuss the incorporation of uncertainty into the calculations and the computational complexity of the approach.

3.1. Background

The causal inference approach is based on that developed in [41], which separates causes from correlations using a calculation of a cause's average impact on an effect's probability. We briefly review the inference approach here and refer the reader to the prior work for more details. Each relationship is represented by a probabilistic temporal logic formula, enabling inference of relationships involving conjunctions, durations, and sequences of events, along with their timing, without prior knowledge [41]. For example, "vigorous activity (v) leads to ($\rightsquigarrow$) high glucose (g) in 5–30 min with probability at least 0.3," is represented by:

$$v \overset{\geq 5, \leq 30}{\underset{\geq 0.3}{\rightsquigarrow}} g. \quad (1)$$

The significance of cause c for effect e, where X is the set of all variables that raise the effect's probability, is:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\varepsilon_{\text{avg}}(c,e)=\sum_{x\in X\setminus c}\frac{P(e|c\wedge x)-P(e|\neg c\wedge x)}{|X\setminus c|}, \quad (2)$$

where c and x are potential causes of the form $c \rightsquigarrow^{\geq r,\leq s} e$ and $x \rightsquigarrow^{\geq r',\leq s'} e$. After each potential cause occurs, e is thus most likely to occur in the time window $[r, s]$ and $[r', s']$ respectively. Note that $0 \leq r \leq s$, and if $r = s$, this is simply a single lag. Thus if c is only an effect of a common cause, it will make a small difference when the true cause is held fixed. Timing subscripts are omitted here, but the probabilities refer to e occurring at a time after x and c such that either could have caused it (i.e. their windows overlap).

Formally, the (in)significant causes are defined as follows. The terminology is significant/insignificant rather than genuine/spurious, as seemingly significant causes may be due to bias or hidden confounders and insignificant causes may actually just be weak genuine causes.

Definition 1—A potential cause c of an effect e is an ε-insignificant cause of e if $\varepsilon_{avg}(c, e) \leq \varepsilon$.

Definition 2—A potential cause c of an effect e that is not an ε-insignificant cause of e is an ε-significant or just-so cause of e.

This can then be treated as a hypothesis testing problem, where one aims to control the false discovery rate (FDR) or false negative rate (FNR) and can use this rate to choose a threshold for ε.

The key assumptions required are that causal relationships are stationary across time (that is, the system is governed by one underlying causal structure); and to find genuine causes and not simply statistically significant causal hypotheses, all common causes of pairs of variables must be measured.

This approach has been compared against others (BNs, DBNs, Granger) on data from multiple domains (finance, biology) [42,43], with significantly fewer false discoveries on time series data and accurate inference of time windows without prior knowledge. To discover time windows, the approach essentially uses an iterative refinement of the time windows (expanding, shrinking, and shifting them earlier and later), greedily aiming to increase the causal significance score. This has been proven to converge to the true windows, assuming data are sampled regularly, and was shown to recover the correct time windows on simulated data where ground truth is known [41]. The key point for inference of timing is that the method does not simply accept or reject hypotheses proposed by a user, but rather infers both the relationship and its true timing (which might be different than that initially tested) directly from the data.

Inference has two phases (1) generating hypotheses and finding potential causes, and (2) assessing the significance of potential causes. In the simplest case hypotheses are pairwise relationships between all variables across a set of time windows, but one may iteratively test

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

more complex formulas. At a minimum, a potential cause must occur before and raise the probability of an effect. To distinguish between confounding due to a common cause and a potentially significant causal relationship, the difference in probability of the effect is averaged in the presence and absence of the cause, holding fixed other potential causes.

3.2. Intuition behind adding uncertainty

Using the method described in the previous section, probabilities are calculated using the frequencies with which the various conjunctions were observed. Thus $P(e|c \wedge x)$ could be calculated with $\#(c \wedge e \wedge x)/\#(c \wedge x)$, making ε_{avg}:

$$\varepsilon_{\text{avg}}(c,e)=\frac{\sum_{x\in X}\frac{\#(c\wedge x\wedge e)}{\#(c\wedge x)}-\frac{\#(\neg c\wedge x\wedge e)}{\#(\neg c\wedge x)}}{|X\backslash c|}. \quad (3)$$

Each instance of $c \wedge x$ adds equally to the sum, even if some observations are less certain than others. Yet, even though we know that CGM readings after a new sensor is inserted and near the end of a sensor's life are less accurate than others, this is not used in standard frequency-based estimates of probability. We aim to be able to use this type of information during inference, so that when finding causes of high blood glucose, these timepoints will have less of an impact. Note that while probabilities calculated from larger samples (e.g. 500/1000) have higher precision than those from smaller samples (e.g. 1/2) we do not yet incorporate this information. However, it may be possible to do so using the same approach of incorporating probabilities of probabilities.

Instead of each observed instance having the same weight in this count, we propose to sum their probabilities. Intuitively this means that if there is an observation of $c \wedge x$, but it is very uncertain (low probability), then whether or not e is observed after, it will not have as large an impact on the calculation as more reliable observations will. On the other hand, this enables one to not simply throw away data points that may potentially be outliers or errors (since distinguishing between an outlying value and an important extreme value in a critically ill patient can be challenging). To simplify calculations we assume the uncertainty in each observation of each variable is independent. However, this may not hold when they are measured by a single device (such as a 9-axis motion sensor). The relationship between error in subsequent observations of a single variable, though, can be handled by the prior model. For example, our belief in the correctness of a blood glucose measurement may depend on when it was inserted (i.e. if it still needs to be calibrated or is past its lifespan). However, this does not require a model of dependent error in the measurements, but rather prior beliefs that account for time since sensor insertion.

3.3. Notation

Before reformulating the approach described to handle uncertain observations, we briefly introduce some new notation. Where i is a timepoint in a series of observations T, $P(x_i)$ is the posterior probability of proposition x at that specific actual time (and is conditioned on the data and our prior beliefs about the reliability of measurement). For discrete and certain data, this is 1 if it is observed and 0 otherwise.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The probability of x at each time in window $[i, j]$ is:

$$P(x_{i\wedge\ldots j})=P(x_i \wedge x_{i+1} \wedge \ldots \wedge x_j). \quad (4)$$

When uncertainties are independent then this calculation simplifies to:

$$P(x_{i\wedge\ldots j})=\prod_{k\in[i,j]} P(x_k). \quad (5)$$

The probability of x occurring at least once in $[i, j]$ is:

$$P(x_{i\vee\ldots j})=1-\prod_{k\in[i,j]} P(\neg x_k). \quad (6)$$

3.4. Calculating causal impact with uncertainty

To incorporate uncertainty into inference, we now replace the frequency-based counts of Eq. (3) with ones based on sums of probabilities. Due to the time windows associated with each relationship (as shown in Eq. (1)), this is more complex than replacing each event with a single probability. When calculating $P(e|c \wedge x)$ in the certain case, we iterate over instances of c and x such that the time windows where they may cause e overlap, summing how often this occurs. Previously e either occurred at least once in the window or did not. Now we may have multiple observations of e with varying probabilities. If e is observed twice, once with $P(e) = 0.1$ and once with $P(e) = 0.6$, our belief in the occurrence of e during that window should then be somewhat greater than 0.6. We now aim to calculate the *probability* of e happening at least once during the overlap of time windows.

In the uncertain case, we must ask what counts as an instance of $c \wedge x$? Instead of determining whether there merely exists an instance of x such that its time window overlaps c's, we now calculate the probability of this happening at least once. Thus if the potential relationships are $c \rightsquigarrow^{\geq 1,\leq 2} e$, and $x \rightsquigarrow^{\geq 2,\leq 2} e$, and c is observed at $t = 2$, then the constraints are as shown in Fig. 1. First, the window for e to be caused by c is [3, 4], as shown with the shaded bar. Then, for x to cause an instance of e (which it can only do in exactly two time units) in conjunction with c_2 it must occur at either time 1 ($t + 1 - 2$) or time 2 ($t + 2 - 2$). If for example $P(x_1) = 0$ while $P(x_2) \neq 0$, then the shaded window would be solely the timepoint $t = 4$.

More generally, the conditional probabilities in Eq. (3) are now replaced with the following calculations. First, $P(e|c \wedge x)$ is updated:

$$P(e|c \wedge x)=\frac{\sum_{t\in T} P(c_t, x_{i\vee\ldots j}, e_{k\vee\ldots l})}{\sum_{t\in T} P(c_t, x_{i\vee\ldots j})}. \quad (7)$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The window $[i, j]$ is the set of times where x may occur and have a time window that overlaps c's, and is given by:

$$[i, j]=[(r+t) - s', (s+t) - r']. \quad (8)$$

Analogously, the window $[k, l]$ is the set of times after $c \wedge x$ where either could have caused e and is defined by:

$$[k, l]=[\max(x_s+r', r+t), \min(x_e+s', s+t)], \quad (9)$$

where x_s is the first time in $[i, j]$ where $P(x_s) \neq 0$ and x_e is the last.

In calculating the significance score in Eq. (7) we calculate the probability of an effect when a cause and other condition are present. We compare this to the probability of the effect when the cause is absent and other condition is held fixed, $P(e|\neg c \wedge x)$, modifying the equation to incorporate uncertainty as follows:

$$P(e|\neg c \wedge x)=\frac{\sum_{t\in T} P(\neg c_{g\wedge\ldots h}, x_t, e_{k'\vee\ldots l'})}{\sum_{t\in T} P(\neg c_{g\wedge\ldots h}, x_t)}. \quad (10)$$

The window $[g, h]$ is then the set of times such that c occurring at these would lead to its window overlapping x's:

$$[g, h]=[(r'+t') - s, (s'+t') - r]. \quad (11)$$

The window for e is now just the window of x after its occurrence at time t':

$$[k', l']=[t'+r', t'+s']. \quad (12)$$

We now iterate over instances of x, finding the probability of there being no instance of c that overlaps with x. Unlike the previous case we have a conjunction rather than a disjunction (i.e. $P(\neg c_g \wedge \neg c_{g+1} \wedge \ldots \wedge \neg c_h)$).

In Fig. 2 is an illustration of the timing constraints for one observation of c at time t, with potential causal relationships $c \overset{\geq r, \leq s}{\rightsquigarrow} e$, and $x \overset{\geq r', \leq s'}{\rightsquigarrow} e$. Shown in grey to the right of c is the time window $[t + r, t + s]$. Then, for $r = r' = 2$ and $s = s' = 4$, the corresponding window $[i, j]$ where x's window may overlap c's is shown. Diagonal lines indicate the timespan for which x is observed with nonzero probability. Finally, the window $[k, l]$ for e is indicated below these.

Assuming uncertainties are independent across variables and observations of each variable, Eq. (7) simplifies to:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$P(e|c \wedge x) = \frac{\sum_{t \in T} P(c_t) P(x_{i \vee \ldots j}) P(e_{k \vee \ldots l})}{\sum_{t \in T} P(c_t) P(x_{i \vee \ldots j})} \quad (13)$$

and Eq. (10) to:

$$P(e|\neg c \wedge x) = \frac{\sum_{t \in T} P(\neg c_{g \wedge \ldots h}) P(x_t) P(e_{k \vee \ldots l})}{\sum_{t \in T} P(\neg c_{g \wedge \ldots h}) P(x_t)}. \quad (14)$$

Definition 3—The significance of cause c for effect e, where X is the set of all factors that raise the probability of e is:

$$\varepsilon_{\text{avg}}(c, e) = \frac{\sum_{x \in X} \varepsilon_x(c, e)}{|X \backslash c|} \quad (15)$$

where

$$\varepsilon_x(c, e) = \frac{\sum_{t \in T} P(c_t) P(x_{i \vee \ldots j}) P(e_{k \vee \ldots l})}{\sum_{t \in T} P(c_t) P(x_{i \vee \ldots j})} - \frac{\sum_{t \in T} P(\neg c_{g \wedge \ldots h}) P(x_t) P(e_{k \vee \ldots l})}{\sum_{t \in T} P(\neg c_{g \wedge \ldots h}) P(x_t)}. \quad (16)$$

In the certain case, where a variable's probability is one when observed and zero otherwise, this reduces to the frequency-based calculation.

3.4.1. Complexity—The complexity of testing pairwise relationships between N variables in a time series of length T in the certain case (once all logical sub-formulas have been checked) is $O(N^3T)$. Adding uncertainty does not change the theoretical bounds, but increases the computation time in practice. Usually an event is not occurring at every timepoint, and the various calculations can be done for only the times where the event is true. However, a proposition now has a probability at each timepoint and all of these times will be incorporated in the calculation so we are always iterating over the T timepoints. Note that N^2 of the computations are independent and can be done in parallel.

3.5. Uses for uncertainty

We can now enable probabilistic discretization, by defining distributions, rather than bins, mapping values to states. This is critical, as many methods cannot use continuous data and often the boundaries between states are not precisely known (or there is a transition that is not captured by a binary mapping). Incorporating probabilities captures this uncertainty while allowing the benefits of discretization. Now instead of mapping each value of a variable to one of a set of discrete states, one can map each value to the probability of a particular state. For instance, when glucose moves from 69 to 70 mg/dl there is not a sudden shift from hypo- to euglycemia. Instead, a set of three probability distributions (one each for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

low, normal, and high glucose) can be created so that each value of blood glucose has a probability of being in each of these states.

Probabilities can be used more generally to incorporate prior knowledge (such as about device accuracy), deal with missing data, and account for error. Probabilities can indicate the likelihood of a variable taking a particular value given that it is recorded to represent measurement error. Thus one can incorporate device uncertainty and bias (e.g. a device is more likely to give a false positive than a false negative). For missing data, imputation of a single value can lead to many errors, while multiple imputation (finding a set of values for a variable and averaging results across inferences with each) increases computational complexity [44]. Instead one can determine for each missing value the probability of a variable being in each of its possible states. This enables multiple imputation while not increasing computational complexity, as inference is still done only once (rather than for each imputed value).

When variables are measured at different timescales (but all theoretically measurable values are present), gaps are often too large for imputation. One can determine in a variable-specific way the probability of the last value being accurate as a function of time after its measurement. For instance, measurements of weight may be very reliable for a period of days while heart rate quickly becomes uninformative.

Finally, ICD9 and other diagnosis codes are often used as indicators for whether a patient has a particular condition, but errors and omissions are common [45] and it can be difficult to figure out which patients have a chronic disease and when it started based on the EHR alone [46]. By incorporating uncertainty, though, we can represent that a patient with multiple pieces of evidence (ICD9 code, medications, symptoms in text) is more likely to have, say, heart failure, than a patient with only a single diagnosis code. Further, as new evidence is amassed, this probability can increase or decrease, ultimately letting us represent diagnoses as probability trajectories based on the evidence at each timepoint.

3.6. Example

We now illustrate the approach with a simple example where the observations themselves are considered trustworthy, but the mapping from continuous values to discrete states is uncertain. This example demonstrates how the calculations proceed and how proper handling of uncertainty can improve causal inference.

Fig. 3 shows a subset of a time series with three variables: c (a meal high in carbohydrates), e (moderate exercise) and g (blood glucose). The variables c and e are discrete, while g is continuous. In bold are values of g that would be considered instances of euglycemia (normal blood glucose) using the traditional range of [70, 120]. The values for $P(g_n)$ show a mapping of glucose values to the probability of euglycemia (glucose in normal range), which captures the smooth transition between states, allowing that a value of 135 is still much more likely to be an instance of euglycemia than one of 158.

First, using the traditional discretization and calculating the impact of e on g_n at one time unit holding fixed c, we find:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$P(g_n|e,c) - P(g_n|\neg e,c) = \frac{0}{2} - \frac{0}{2} = 0. \quad (17)$$

Here it seems e has no impact at all on g_n. Incorporating uncertainty into the calculation, using $P(g_n)$, we instead have the following. Note that $P(c)$ and $P(e)$ are either one or zero at each time, and the window is exactly one time unit, so we simply use $P(g_n)$ at each time, instead of calculating the probability across a time window.

$$P(g_n|e,c) - P(g_n|\neg e,c) = \frac{0.99+0.95}{1+1} - \frac{0.09+0.01}{1+1} = 0.92.$$

Instead of no impact, we now find that exercise combined with a meal has a significant impact on glucose as compared to a meal alone. In contrast, values close to normal (135 and 140) following "c and e" were indistinguishable from the higher values following "c and not e" (158 and 165) when strictly discretizing according to a fixed window. By assigning probabilities, we can capture this distinction in the continuous values.

4. Data collection

4.1. Simulated data

To evaluate the inference method we simulated data with known ground truth and varied the uncertainty in measurements. The goal of these experiments is to determine whether causal relationships can be accurately inferred from uncertain data with the proposed method, and determine how the level of uncertainty affects inference accuracy. Our primary evaluation is the false discovery rate and false non-discovery rate (FDR and FNR).

We randomly generated 5 causal structures with 10 or 20 relationships among 25 variables (including chains, cycles, self loops, etc.), with each relationship having a lag of one time unit. Fig. 4 shows one example structure. We did not generate relationships with longer lags or windows to isolate the effect of uncertainty and noise.

At each time an event (variable) may occur spontaneously (probability = 0.1) or may occur if one of its causes occurred at the previous time (probability = 0.9). This yields strong relationships, enabling us to distinguish between uncertain observations and weak relationships.

After generating the ground-truth data (20,000 timepoints), we added uncertainty in two ways.

4.1.1. Static uncertainty—Here, each variable in each dataset was independently assigned a randomly generated probability of being correctly reported (i.e. output says true if it occurred). Probabilities were in the range [p, 1] with $p \in \{0.55, 0.75, 0.85, 0.9, 0.95, 1\}$. Thus there may be certain observations of only cause or effect and not both. In all, 60 datasets of this type were generated (5 structures, 6 probability ranges, 2 runs for each combination).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

4.1.2. Varied uncertainty—A key feature of our method is incorporating observation-specific uncertainty. To test this, we generated 8 datasets (2 structures, 2 probability ranges, 2 runs each) where the probability of each observation (rather than variable) being correctly reported was randomly chosen within [0.55, 1] and [0.9, 1]. That is, if a variable is actually true at a given time (e.g. individual has hyperglycemia), we flip a weighted coin to determine whether the output reports true or false. With probabilities in [0.9, 1], most observations are correct, while in the wider probability range there will be many more instances of erroneous output.

4.2. Real-world data

We aim to demonstrate both the feasibility of continuous physiologic monitoring in individuals with T1DM in free-living conditions and that these data can be used to gain insight into causes of changes in glycemia. Our primary evaluation is face validity of the relationships, assessing how they relate to prior knowledge. The Diabetes Management Integrated Technology Research Initiative (DMITRI) study collected data from 17 participants (10 male, 7 female) ages 19–61 with T1DM. Participants were active, with all exercising at least 2–4 times per week, and most (13 of 17) >4 times per week. Average duration of diabetes and HbA1c were 14.9 ± 11.0 years and 7.3 ± 1.3% respectively. Data are available through iDASH (https://idash.ucsd.edu). Participants were monitored over approximately 72 h plus a baseline assessment, though we focus on the sensor and device data. The continuously collected data and body-worn sensors include: glucose (Dexcom 7+ CGM), insulin dosing (insulin pump), activity status (BodyMedia SenseWear, Respironics Actiwatch), heart rate (Polar chest strap), temperature (SenseWear), and sleep (Zeo Personal Sleep Coach). Data collection frequencies differed between devices, so all were synced to the 5-min intervals of the CGM.

Each continuous-valued variable was discretized, with a probabilistic approach used when possible. Data were prepared as follows:

Activity: The SenseWear activity monitor outputs the fraction of each five-minute interval spent in sedentary, moderate, vigorous, or very vigorous activity based on METs (metabolic equivalents),[1] using ranges of under 3.0 (sedentary), 3.0–6.0 (moderate), 6.0–9.0 (vigorous) and above 9.0 (very vigorous). Since the values range from 0 to 1 and the values for a given interval sum to 1, they were used as probabilities of the corresponding activity during that five-minute interval.

Glucose measurements are often volatile and the usual approach of calling values in the range [70, 120] (mg/dL) normal (euglycemia), values below hypoglycemia, and values above hyperglycemia may not accurately distinguish true hypo/hyperglycemic episodes in people with diabetes. So discretizing 68 mg/dL as “low glucose” and 70 mg/dL as “normal” overstates the confidence in these measurements [49]. Instead, we defined a probability distribution based on this range that ensures a value of, say, 68 mg/dL, still has a high probability of corresponding to euglycemia. Distributions corresponding to hypo- and hyperglycemia were created by subtracting the euglycemia distribution from 1, as shown in

[1]These measurements correlate well with actual energy expenditure, though studies have been limited in scope [47,48].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Fig. 5. A data point may correspond to multiple categories, with different probabilities, and a possible but unlikely hyperglycemic episode will not overly influence determination of factors affecting hyperglycemia.

Heart rate (HR) zones, to determine intensity of activity, were calculated for each subject using age and baseline resting heart rate. Maximum HR (HRmax) was estimated, using a standard approach, as 220-age, making the zones: X * (HRmax − HRrest)+HRrest, where X was in the interval [.5, .85, 1]. HR below 90 was defined as resting, and between 90 beats per minute and 50% of HRmax, elevated.

Insulin pump data is broken into two types: basal rate (continuous insulin dosing) and bolus (discrete infusion). We used presence or absence of a bolus at each time.

Sleep: From the Zeo sleep monitor, we recorded mode sleep stage during each interval. The SenseWear activity monitor also recorded percentage of each interval spent in sleep and percentage lying down (each was treated as the probability of each activity during that interval).

Temperature was treated similarly to glucose, with a probability distribution of euthermia centered at 33 °C (as temperature was measured on the surface of the skin), and declining rapidly above and below. The result is three distributions corresponding to hypo-, hyper-, and euthermia.

5. Results

5.1. Simulated data

We first validate the method using the simulated time series. The proposed method used the known time lag for relationships and actual probabilities for each variable. The cleanliness of the synthetic data, meant the usual approach of fitting the null distribution to the data to identify a threshold for ε could not be used (the difference between the insignificant z-values, which were near zero but skewed negative, and the significant ones was too great). We instead used breaks in the distribution (e.g. when a fit to the histogram of significance scores goes to zero), but this increases the false negative rate, so performance can be improved with better methods for choosing the threshold.

We compared results to those of DBNs using Banjo [50]. Using a single lag and relatively small number of variables/timepoints enabled exploration of a large portion of the search space (limiting concerns about identifying a local minima with a large search space). Parameters used were: simulated annealing with random local moves, runtime of 1 h with 6 threads, and max parent count (number of parents of a node) of 6.

Table 1 shows false discovery and false negative rates (FDR and FNR) for each algorithm, by probability range (each being [p, 1] with p shown) for static uncertainty. Our FDR is lower at every probability level, and increases to a max half that of DBNs. Our method yields a higher FNR as noise increases, as highly uncertain measurements will be given little weight. While the tradeoff between FNR and FDR is specific to each problem and the associated costs, in this case the increasing FNR is by design. That is, a completely

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

uncertain variable should not be found as a cause or effect of any other, and should be treated as if it were not measured (i.e. latent). The FDR increases as uncertainty does, as this weakens true relationships while increasing the number of spurious correlations, so false positives make up a larger portion of the inferences (as there are fewer inferences). While DBNs made no false negatives in many cases (the true network was a subset of that inferred), the FDR increased from 0.064 to 0.204 when going from certain data ($p = 1$) to a 5% error rate ($p = 0.95$), and this is higher than the FDR of our approach with uncertainty at 0.55 (.123).

Across all 8 datasets with variable uncertainty (varying by observation), our approach had an FDR and FNR of zero. For DBNs, the FNR was also zero, but the FDR was.048 and.032 for $p = [0.55, 1]$ and $[0.9, 1]$ respectively, as shown in Table 2. Thus by giving less weight to possibly erroneous measurements, we can separate the signal from the noise. This is particularly important for our application to T1DM, as relying more on certain data (e.g. values that are surely hyperglycemic episodes when finding causes of such excursions) may help overcome the substantial noise in glucose measurements. DBNs fared much better with variable uncertainty than with static uncertainty, as a consistently unreliable variable severely reduces their accuracy.

5.2. Real-world diabetes data

Using the DMITRI dataset, we compared uncertain discretization (e.g. each value mapped to probabilities of belonging to each category) to traditional categorical discretization (e.g. glucose values in [70, 120] are considered normal and above and below are hyper- and hypoglycemia respectively). Using a data-driven approach, we tested for causal relationships between all variables and hypo-, hyper- and euglycemia over a series of candidate time windows (5–15, 15–30, 30–45, and 45–60 min). As actual timings may be different, we used the approach of [41] to refine each window for each relationship, before recalculating the causal significance scores with the final set of relationships.

Unsurprisingly, we found that glucose tended to remain high/low/normal at a short timescale (i.e. being low causes glucose to remain low). Using the proposed method we made another finding, though: that very vigorous exercise as a significant cause of hyperglycemia. This finding was made separately using activity estimates based on heart rate in 5–15 min (with estimated anaerobic activity zones) and an activity monitor calculating METs (metabolic equivalents) in 15–30 min, with both having fdr < 0.01. Heart rate was somewhat more significant (ε_{avg} 0.32 vs. 0.29), because of the specificity of the measurement. These relationships between intense activity and hyperglycemia were the most significant at all timescales by a wide margin (aside from those between the glucose states and themselves), but did not reach statistical significance in other time windows. This suggests that we identified when the effect peaks, but it may occur over a longer timescale. Further work with a larger sample is needed to confirm the timing.

While the data contain a limited number of glucose excursions (deviations from euglycemia) and episodes of intense exercise, we were able to find this relationship from only a few days of data collected for 17 individuals. Hyperglycemia after anaerobic activity has been previously identified in individuals with T1DM [51], giving us confidence in the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

methodology and sensor data quality. This finding was not made from data discretized according to the traditional range (only the relationships between hypo/hypo, eu/eu, and hyper/hyperglycemia were identified), demonstrating that incorporating uncertainty into discretization can improve inference power in noisy data.

6. Discussion

Patients themselves make many chronic disease management decisions during daily life rather than in the context of medical treatment. Decision support systems have primarily focused on clinicians, while patient-centered treatment has tried to remove the need for active involvement (e.g. closed-loop glucose control). Yet, many systems fail during translation from controlled environments to the real-world. We propose that (1) patient-generated data are a viable and important data source for researchers and (2) these data require specialized methods due to their specific challenges. We report the development of a unique type 1 diabetes dataset from free-living conditions and introduce a novel extension of a causal inference method to handle uncertain data. These data may potentially be used by researchers to uncover causes of changes in glycemia and understand whether findings from controlled laboratory environments translate to real-world environments.

In contrast to existing work, which weights all data points equally during causal inference we develop an approach that incorporate probabilities into traditionally frequency-based calculations of causal significance. This enables representation of beliefs and information such as error rates specific to each observation of a variable, and was shown to lead to fewer false discoveries than DBNs in data with simulated uncertainty. While the approach may lead to more false negatives in theory (due to treating highly uncertain variables as missing), in practice we showed that it can increase power when dealing with noisy real-world data. Despite the small sample size and few occurrences of hyperglycemia, by incorporating knowledge of the uncertainty inherent in the data, we were able to uncover the causal relationship between intense exercise and hyperglycemia, using both heart rate data and activity measured by METs. With a larger dataset, more causal inferences may be possible.

The primary limitations of this work are the need for priors about uncertainty, as well as the assumptions made about the independence of error across variables. We assumed here that the probability distributions and error rates were given as background knowledge, and while information from manufacturers and research on device accuracy can provide this in some cases, future work is needed on better estimating uncertainty for each variable and observation. In particular, this could potentially be identified in a data-driven way, by triangulating between multiple measurements of a phenomenon (e.g. finding mismatch between heart rate, location, and motion sensing), and combining data with manufacturer-provided accuracy information. Similarly, many devices contain multiple sensors and when one piece of the system fails, others will too. Incorporating dependence in uncertainty will be important for better handling such cases. Finally, one limitation of our experimental work is the relatively small sample size, with 17 individuals over 3 days and lack of ground truth for measurements, which did not allow us to infer or evaluate such models.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Direct representation of uncertainty in biomedical data may enable better use of other data beyond that generated from body-worn sensors. Work extracting phenotypes from EHRs may benefit from allowing probabilistic diagnoses, where researchers can represent the amount of evidence toward a diagnosis rather than treating this as a binary categorization. For instance, a patient with multiple medications for heart failure, multiple visits for the condition, and heart failure on their problem list could be distinguished from patients with only a diagnosis code. Similarly, concepts extracted from text often have ambiguous timings, so instead of assigning a single time to an event, a probability distribution over the likely times of occurrence better captures this uncertainty.

Acknowledgments

NH thanks Prof. Ramesh Rao (UCSD, Qualcomm Institute) for guidance on and provision of monitoring devices and software; Dr. Steven V. Edelman (UCSD) for guidance on clinical study design and diabetes care practices; Giorgio Quer (UCSD, Qualcomm Institute) for guidance on sensor data collection and analysis; Prof. Sara Mednick (UC Riverside) for provision of Actiwatch devices and guidance on sleep data analysis; Zeo, Inc for provision of sleep monitor devices and guidance on sleep data analysis; the DMITRI study team for their extensive work in data collection, retrieval, and curation, especially Allan Mark Asuncion (UCSD), Bryant Chen (UCLA), Tushar Dave (U Maryland Baltimore), and Ashley Hall (Novo Nordisk); Peter Nerothin (Prescott College) for guidance on sensor and device utilization among people with T1DM; and especially the participants in the DMITRI study for their generosity and diligence. SK thanks Yuxiao Huang for assistance with experiments and manuscript preparation.

This publication was supported by the NLM of the NIH under Award Number R01LM011826 (SK). iDASH is supported by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54HL10846. The project described was partially supported by the National Institutes of Health, Grant UL1RR031980 for years 1 & 2 of CTSA funding and/or UL1TR000100 during year 3 and beyond of CTSA funding. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Kung H-C, Hoyert DL, Xu J, Murphy SL. Deaths: final data for 2005. Natl. Vital Stat. Rep. 2008; 56(10):1–120.
2. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int. J. Environ. Res. Publ. Health. 2009; 6(2):492–525.
3. Hull S. Patient-generated health data foundation for personalized collaborative care. Comput. Inform. Nurs. 2015; 33(5):177–180. [PubMed: 25974360]
4. Nundy S, Lu C-YE, Hogan P, Mishra A, Peek ME. Using patient-generated health data from mobile technologies for diabetes self-management support provider perspectives from an academic medical center. J. Diabetes Sci. Technol. 2014; 8(1):74–82. [PubMed: 24876541]
5. Kropff J, Bruttomesso D, Doll W, Farret A, Galasso S, Luijf Y, Mader J, Place J, Boscari F, Pieber T, et al. Accuracy of two continuous glucose monitoring systems: a head-to-head comparison under clinical research centre and daily life conditions. Diabetes Obes. Metab. 2015; 17(4):343–349. [PubMed: 25132320]
6. Thabit H, Leelarathna L, Wilinska ME, Elleri D, Allen JM, Lubina-Solomon A, Walkinshaw E, Stadler M, Choudhary P, Mader JK, et al. Accuracy of continuous glucose monitoring during three closed-loop home studies under free-living conditions. Diabetes Technol. Therap. 2015; 17(11): 801–807. [PubMed: 26241693]
7. Facchinetti A, Del Favero S, Sparacino G, Castle JR, Ward WK, Cobelli C. Modeling the glucose sensor error. IEEE Trans. Biomed. Eng. 2014; 61(3):620–629. [PubMed: 24108706]
8. Facchinetti A, Del Favero S, Sparacino G, Cobelli C. Model of glucose sensor error components: identification and assessment for new Dexcom G4 generation devices. Med. Biol. Eng. Comput. 2015; 53(12):1259–1269. [PubMed: 25416850]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

9. C. for Disease Control, Prevention. National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States. US Department of Health and Human Services; 2014.
10. Boyle JP, Thompson TJ, Gregg EW, Barker LE, Williamson DF. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. Popul. Health Metrics. 2010; 8(1):29. http://dx.doi.org/10.1186/1478-7954-8-29.
11. Herman WH. The economic costs of diabetes: is it time for a new treatment paradigm? Diabetes Care. 2013; 36(4):775–776. [PubMed: 23520368]
12. Karter AJ, Stevens MR, Herman WH, Ettner S, Marrero DG, Safford MM, Engelgau MM, Curb JD, Brown AF. Out-of-pocket costs and diabetes preventive services the translating research into action for diabetes (triad) study. Diabetes Care. 2003; 26(8):2294–2299. [PubMed: 12882851]
13. Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, Hogg RJ, Perrone RD, Lau J, Eknoyan G. National kidney foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Ann. Intern. Med. 2003; 139(2):137–147. [PubMed: 12859163]
14. Beulens JW, Grobbee DE, Nealb B. The global burden of diabetes and its complications: an emerging pandemic. Euro. J. Cardiov. Prevent. Rehabil. 2010; 17(1 suppl):s3–s8.
15. Lloyd-Jones D, Adams RJ, Brown TM, Carnethon M, Dai S, De Simone G, Ferguson TB, Ford E, Furie K, Gillespie C. Heart disease and stroke statistics-- 2010 update. Circulation. 2010; 121(7):e46–e215. [PubMed: 20019324]
16. Nathan DM, Zinman B, Cleary PA, Backlund J-YC, Genuth S, Miller R, Orchard TJ. Modern-day clinical course of type 1 diabetes mellitus after 30 years duration: the diabetes control and complications trial/epidemiology of diabetes interventions and complications and pittsburgh epidemiology of diabetes complications experience (1983–2005). Arch. Intern. Med. 2009; 169(14):1307. [PubMed: 19636033]
17. Schwartz, S.; Scheiner, G. The role of continuous glucose monitoring in the management of type-1 and type-2 diabetes. Tfm Publishing; 2012.
18. Wilinska, ME.; Allen, JM.; Elleri, D.; Kumareswaran, K.; Nodale, M.; Evans, ML.; Dunger, DB.; Hovorka, R. 5th European Conference of the International Federation for Medical and Biological Engineering. Springer; 2012. Overnight closed-loop insulin delivery in patients with type 1 diabetes; p. 961-963.
19. Kovatchev BP, Renard E, Cobelli C, Zisser HC, Keith-Hynes P, Anderson SM, Brown SA, Chernavvsky DR, Breton MD, Mize LB. Safety of outpatient closed-loop control: first randomized crossover trials of a wearable artificial pancreas. Diabetes Care. 2014; 37(7):1789–1796. [PubMed: 24929429]
20. Basu A, Dube S, Slama M, Errazuriz I, Amezcua JC, Kudva YC, Peyser T, Carter RE, Cobelli C, Basu R. Time lag of glucose from intravascular to interstitial compartment in humans. Diabetes. 2013; 62(12):4083–4087. [PubMed: 24009261]
21. Guerra S, Facchinetti A, Sparacino G, De Nicolao G, Cobelli C. Enhancing the accuracy of subcutaneous glucose sensors: a real-time deconvolution-based approach. IEEE Trans. Biomed. Eng. 2012; 59(6):1658–1669. [PubMed: 22481799]
22. Salem MA, AboElAsrar MA, Elbarbary NS, ElHilaly RA, Refaat YM. Is exercise a therapeutic tool for improvement of cardiovascular risk factors in adolescents with type 1 diabetes mellitus? A randomised controlled trial. Diabetol. Metab. Synd. 2010; 2(1):47.
23. Goodyear LJ, Kahn BB. Exercise, glucose transport, and insulin sensitivity. Ann. Rev. Med. 1998; 49(1):235–261. [PubMed: 9509261]
24. Colberg-Ochs, S.; Riddell, M. Physical activity: regulation of glucose metabolism, clinical management strategies and weight control. American Diabetes Association; 2013.
25. Bouchard, C.; Shephard, RJ.; Stephens, T. Physical Activity, Fitness, and Health: International Proceedings and Consensus Statement. Human Kinetics Pub.; 1994.
26. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA. 2015; 313(6):625–626. [PubMed: 25668268]
27. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int. J. Behav. Nutr. Phys. Activ. 2015; 12(1):42.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

28. Rahman SA, Huang Y, Claassen J, Heintzman N, Kleinberg S. Combining fourier and lagged k-nearest neighbor imputation for biomedical time series data. J. Biomed. Inform. 2015; 58:198–207. http://dx.doi.org/10.1016/j.jbi.2015.10.004. [PubMed: 26477633]
29. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Bmj. 2009; 338:b2393. [PubMed: 19564179]
30. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. Stat. Med. 1991; 10(11):1665–1677. [PubMed: 1792461]
31. Weichselberger K. The theory of interval-probability as a unifying concept for uncertainty. Int. J. Approx. Reason. 2000; 24(2):149–170.
32. Dubois D. Possibility theory and statistical reasoning. Comput. Stat. Data Anal. 2006; 51(1):47–69.
33. Fagin R, Halpern JY. Uncertainty, belief, and probability. Comput. Intell. 1991; 7(3):160–173.
34. Spirtes, P.; Glymour, C.; Scheines, R. Causation, Prediction, and Search. MIT Press; 2000.
35. Pearl, J. Causality: Models, Reasoning, and Inference. Cambridge University Press; 2000.
36. Murphy, K.; Mian, S. Tech. Rep. Berkeley, CA: University of California; 1999. Modelling Gene Expression Data Using Dynamic Bayesian Networks.
37. Claassen T, Heskes T. A Bayesian approach to constraint based causal inference. UAI. 2012
38. Claassen, T.; Heskes, T. Bayesian probabilities for constraint-based causal discovery. IJCAI; 2013.
39. Flores MJ, Nicholson AE, Brunskill A, Korb KB, Mascaro S. Incorporating expert knowledge when learning bayesian network structure: a medical case study. Artif. Intell. Med. 2011; 53(3): 181–204. [PubMed: 21958683]
40. Granger CW. Testing for causality: a personal viewpoint. J. Econ. Dyn. Control. 1980; 2:329–352.
41. Kleinberg, S. Causality, Probability, and Time. Cambridge University Press; 2012.
42. Kleinberg, S. A logic for causal inference in time series with discrete and continuous variables; Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI); 2011. p. 943-950.
43. Kleinberg S, Mishra B. The temporal logic of causal structures. UAI. 2009
44. Rubin DB. Multiple imputation after 18+ years. J. Am. Stat. Assoc. 1996; 91(434):473–489.
45. O'malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. Health Serv. Res. 2005; 40(5p2):1620–1639. [PubMed: 16178999]
46. Kleinberg, S.; Elhadad, N. AMIA Annual Symposium Proceedings. Vol. 2013. American Medical Informatics Association; 2013. Lessons learned in replicating data-driven experiments in multiple medical systems and patient populations; p. 786
47. s Calabr MA, Lee J-M, Saint-Maurice PF, Yoo H, Welk GJ. Validity of physical activity monitors for assessing lower intensity activity in adults. Int. J. Behav. Nutr. Phys. Activ. 2014; 11(1):1.
48. Plasqui G, Bonomi A, Westerterp K. Daily physical activity assessment with accelerometers: new insights and validation studies. Obesity Rev. 2013; 14(6):451–462.
49. Freckmann G, Schmid C, Baumstark A, Pleus S, Manuela Link ME, Haug C. System accuracy evaluation of 43 blood glucose monitoring systems for self-monitoring of blood glucose according to din en iso 15,197. J. Diabetes Sci. Technol. 2012; 6(5):1060–1075. [PubMed: 23063032]
50. Hartemink AJ. 2008 <http://www.cs.duke.edu/amink/software/banjo/>.
51. Riddell MC, Perkins BA. Type 1 diabetes and vigorous exercise: applications of exercise physiology to patient management. Can. J. Diabetes. 2006; 30(1):63–71.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

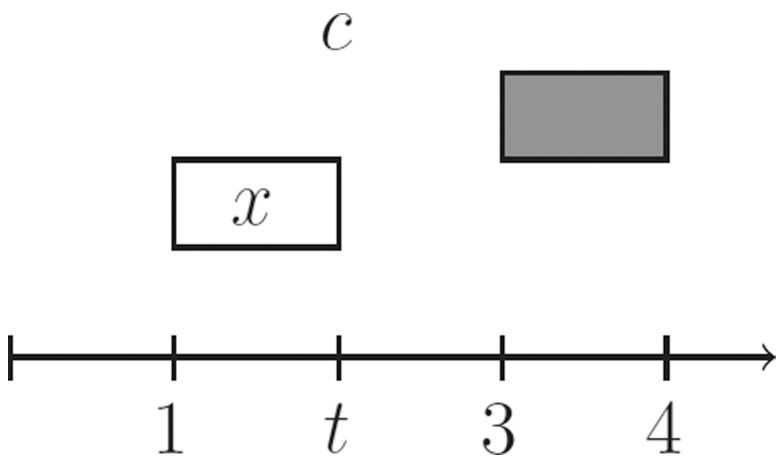

Fig. 1.
Example illustrating timing constraints for two causal relationships.

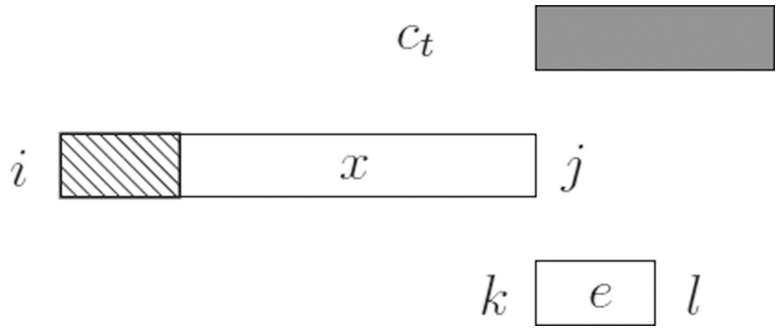

Fig. 2.
Illustration of time windows.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

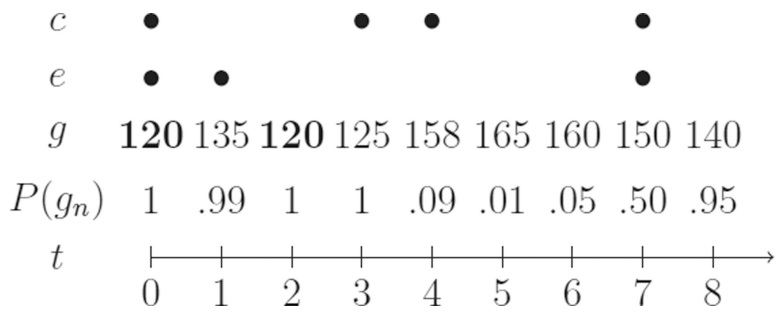

Fig. 3.

Example time series of two discrete variables (*c*, *e*) and glucose (*g*), which is mapped to probability of euglycemia ($P(g_n)$).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

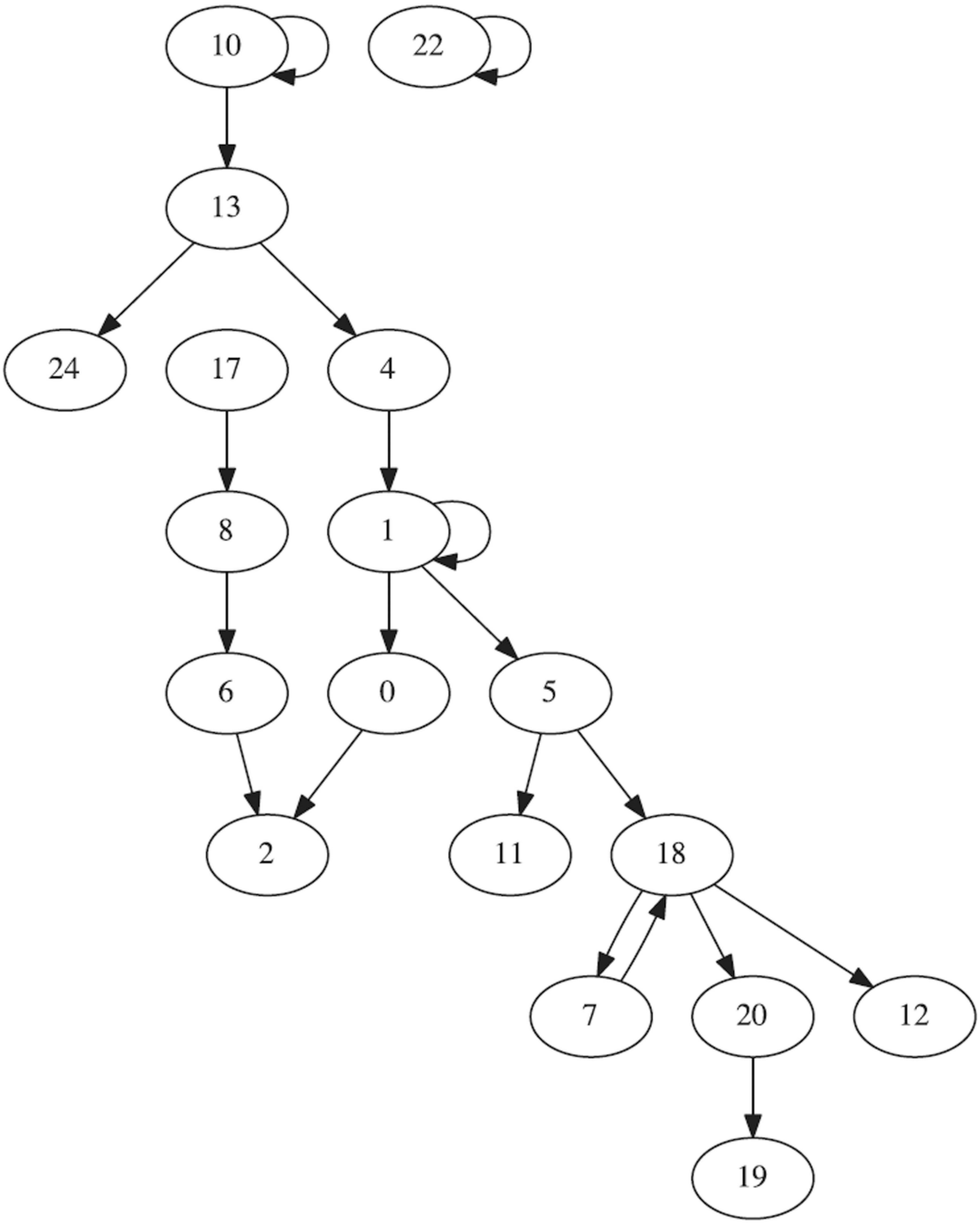

Fig. 4.
One of five randomly generated causal structures.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

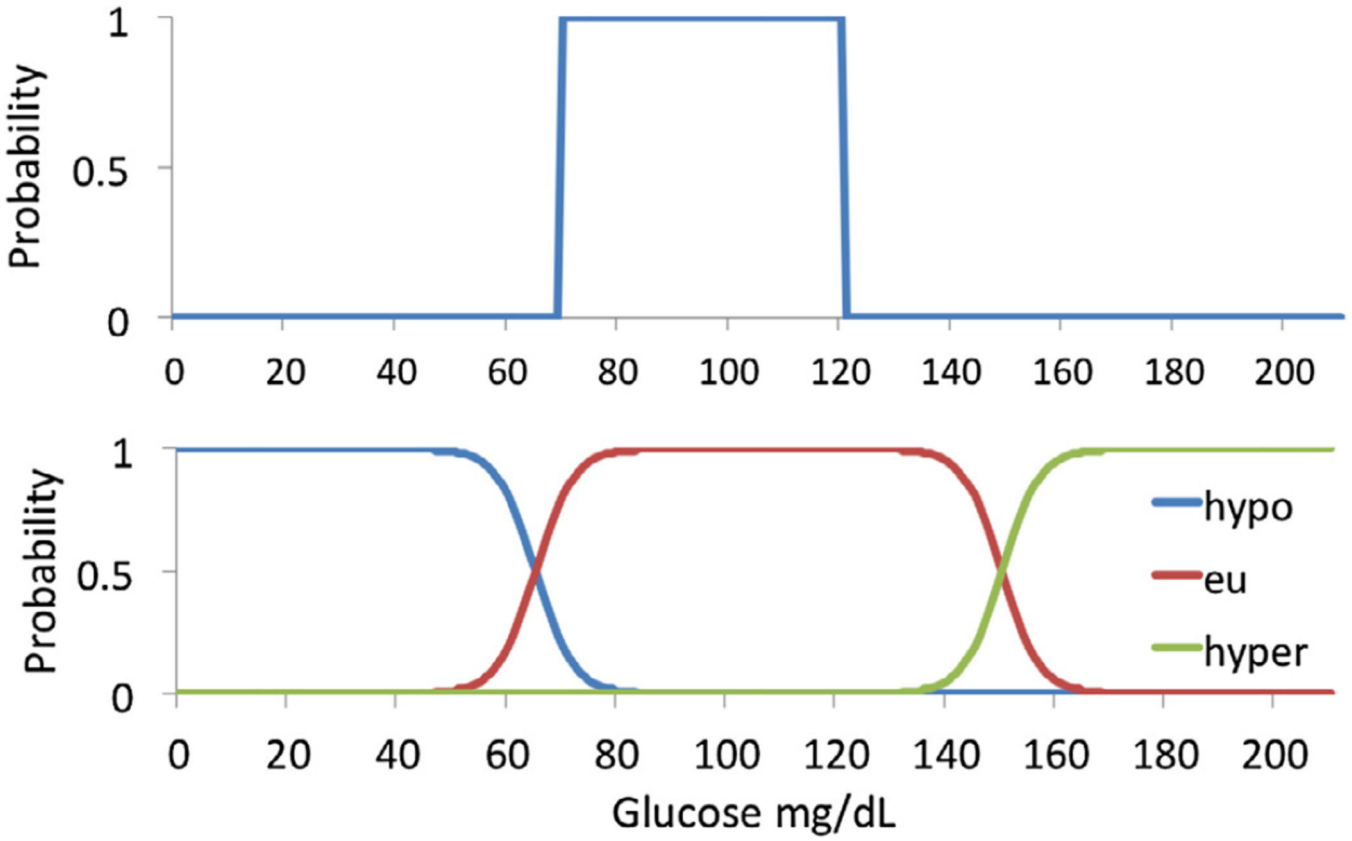

Fig. 5.

Probabilistic discretization of blood glucose (bottom) and traditional step-function (top), where values in [70, 120] are considered euglycemia. Based on the value reported by the CGM, probabilities of eu-, hyper- and hypoglycemia are assigned.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Comparison of causal inference with uncertain observation to DBNs on simulated time series for varied levels of uncertainty, with static uncertainty for each variable.

Method	1		0.95		0.9		0.85		0.75		0.55	
	FDR	***FNR***	***FDR***	***FNR***	***FDR***	***FNR***	***FDR***	***FNR***	***FDR***	***FNR***	***FDR***	***FNR***
Uncertain	.058	.089	.073	.131	.098	.138	.126	.263	.123	.644	.123	.644
DBN	.064	0	.204	0	.231	0	.270	0	.286	0	.272	.181

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Comparison of causal inference with uncertain observation to DBNs on simulated time series for varied levels of uncertainty, with varied uncertainty for each variable.

Method	[0.9, 1]		[0.55, 1]	
	FDR	*FNR*	*FDR*	*FNR*
Uncertain	0	0	0	0
DBN	0.032	0	0.048	0