The state of variability: a vision for descriptors of glycaemia

Tony Zhou¹, Jennifer Knopp¹ and J. Geoffrey Chase¹

¹: University of Canterbury, Dept of Mechanical Engineering, Centre for Bio-Engineering, Christchurch, New Zealand; tony.zhou@pg.canterbury.ac.nz

Abstract:

Glycaemic control (GC) in the intensive care unit is contentious. Hyperglycaemia, hypoglycaemia, and glycaemic variability are all associated with increased morbidity and mortality. While some studies and physiological evidence suggests GC should benefit hyperglycaemic patients, others show no or negative effects and increased incidence of hypoglycaemia. Interpretation of results is made more difficult by differences in the measurement and reporting of glycaemic control, blood glucose levels and variability in patients. In addition, target ranges for glycaemic control are not universally accepted, and higher targets are often used out of fear of hypoglycaemia, rather than their relationship to a clinical outcome. Importantly, current metrics are mostly mathematically derived, and then related to a potential clinical outcome, yielding highly variable results, while very few are clinically defined first. Thus, the goal metrics for control are not directly clinically defined.

This paper reviews differences in the reporting of BG level and its variability in literature. It then proposes a vision for improved description of glycaemia and presents a continuous glucose monitoring (CGM) sensor-based method to better quantify glycaemic level and variability, based on clinically defined metrics. A case study of this new method is presented using CGM sensor data from a study of 614 infants at risk of neonatal hypoglycaemia. Results show the new clinically defined method is able to describe changes in glycaemic level and variability in these patients and presents a flexible way forward for accurately describing state and variability from a clinically defined perspective. This method may provide better insight to patient glycaemia over time, and thus provide scope for improved control of glycaemia.

Keywords: Glycaemic Control; Glycaemia; Glycaemic State; Glycaemic Variability; Neonatal Intensive Care; Adult Intensive Care; Intensive Care Unit; ICU
1.0 Introduction

Glycaemic control (GC) in the critically ill is a contentious issue. There is a strong association between hyperglycaemia and intensive care unit (ICU) mortality and morbidity (Capes, Hunt, Malmberg, & Gerstein, 2000; C. Christiansen, Toft, Jorgensen, Andersen, & Tonnesen, 2004; J. S. Krinsley, 2003; Umpierrez et al., 2002). Early studies showed insulin therapy improved outcomes (Chase, Shaw, et al., 2008; Finney, Zekveld, Elia, & Evans, 2003; J. S. Krinsley, 2004; van den Berghe et al., 2001). However, several studies failed to replicate early beneficial results (Brunkhorst et al., 2008; Finfer et al., 2009; Kalfon et al., 2014; Preiser et al., 2009), while others showed beneficial outcomes (Dubois et al., 2017; Hersh et al., 2018) and several showed no benefit or harm (Chase et al., 2018; Kavanagh & McCowen, 2010; Wiener, Wiener, & Larson, 2008). Overall, the large randomised trial in 2009 reporting adverse outcomes has become a point of reference for many rejecting GC (Chase & Dickson, 2017; Finfer et al., 2009).

The weight of physiological studies examining mechanisms behind the effect of hyperglycaemia and insulin therapy on outcomes suggests insulin therapy should provide benefit, through reduced blood glucose (BG) and the insulin hormone itself (McCowen, Malhotra, & Bistrian, 2001; Van den Berghe, 2004; Vanhorebeek & Langouche, 2009; Wissing et al., 2018). Hyperglycaemia attenuates inflammation and impairs immune system function, resulting in increased incidence of sepsis and infection (Ali et al., 2008; Brunkhorst et al., 2008; Donati et al., 2014; Grey & Perdrizet, 2004). It also has been observed to induce oxidative stress and endothelial and microcirculation dysfunction (Ellahham, 2010), both of which contribute to organ failure (Aird, 2003; Motoyama et al., 2003). Insulin therapy directly counteracts each of these mechanisms of injury, reducing inflammation and infection, protecting endothelial cells and reducing oxidative stress. Thus, the physiological evidence suggests safe, effective glycaemic control should benefit patients in the ICU.

It has also been suggested poor glycaemic control, resulting in glycaemic variability and
hypoglycaemia, may counteract any potential benefits of reduced hyperglycaemia. In particular, glycaemic variability (Bagshaw et al., 2009; Egi, Bellomo, Stachowski, French, & Hart, 2006; J. S. Krinsley, 2008; S. Penning et al., 2015; C. G. Pretty et al., 2012; Signal, Le Compte, Shaw, & Chase, 2012) and hypoglycaemia (Bagshaw et al., 2009; Egi et al., 2010; Hermanides et al., 2010; Kalfon et al., 2015; S. Penning et al., 2015) are both independently associated with adverse outcomes. Poorly designed GC protocols may result in increased incidence of high BG variability and hypoglycaemia (Kalfon et al., 2015; Wiener et al., 2008), confounding study outcomes and comparisons across studies.

Hypoglycaemia appears more directly detrimental to patients than hyperglycaemia (S. Penning et al., 2015) and only one study, of all reported, reduced hypoglycaemia with insulin therapy for GC (Chase, Shaw, et al., 2008), indicating that poor control, rather than GC itself may be the cause of the variability in study outcomes.

More specifically, a recent study of metabolic variability and mortality outcomes showed no difference in underlying metabolic variability between survivors and non-survivors (V. Uyttendaele, J. L. Dickson, G. M. Shaw, T. Desaive, & J. G. Chase, 2017), suggesting patient outcomes are a function of the quality of control delivered and not patient-specific or cohort-specific characteristics. Poorly delivered control might thus affect study outcomes when looking for benefit or harm, as studies show good GC may need to be achieved for essentially all patients (Chase et al., 2010). As such, there is a clear need for good control algorithms able to directly manage the significant intra- and inter- patient variability that makes GC difficult (Chase et al., 2011) to provide safe, effective control for all patients. This outcome points to a clear need for greater input from the fields of control systems and automation.

Model based methods have been developed to achieve this goal (Dubois et al., 2017; Evans et al., 2012; Hovorka et al., 2004; Plank et al., 2006; Stewart et al., 2016). In particular, many monitor and respond to changes in patient-specific metabolic condition using an insulin sensitivity parameter. Insulin sensitivity is a key determinant of the glucose uptake response to an insulin dose, and this
sensitivity is most variable early in the ICU stay where most hypoglycaemia occurs (Bagshaw et al., 2009), both in response to patient condition and clinical interventions (C. Pretty et al., 2011; C. G. Pretty et al., 2012; C. G. Pretty et al., 2014; Thomas et al., 2014).

However, while the future looks optimistic for model-based GC and patient outcomes (Chase & Dickson, 2017), there is still much debate over what GC targets are appropriate and/or safe for an ICU context (Finfer et al., 2009). Many prefer higher targets out of fear of hypoglycaemia (Egi et al., 2010; Hermanides et al., 2010; Kalfon et al., 2015; S. Penning et al., 2015). Some studies have shown greater benefit from lower targets (Hersh et al., 2018; J. S. Krinsley & Preiser, 2015; Sophie Penning et al., 2014; Signal, Le Compte, Shaw, et al., 2012). Further, evidence suggests these targets may be patient specific and/or differ between ICU cohorts (James S Krinsley et al., 2013). Confounding this issue is a lack of consensus on how to measure and/or report GC outcomes and variability at a cohort and patient level (Chase et al., 2018; Eslami, Taherzadeh, Schultz, & Abu-Hanna, 2011).

Thus, there is a clear need for coherency and novelty in regard to assessing glycaemia, both in level and variability. Consensus in assessing state or level and variability with metrics well-correlated with clinically relevant outcomes would provide the means to assess all protocols and methods, where such metrics are lacking (Chase et al., 2018). Hence, while automation and control are emerging in this field, the key control parameter, the measure of desirable outcome parameters, glycaemia and its variability in this case, are missing.

This review summarises the state of the art with regards to assessing glycaemic level and variability in hospital ICU cohorts. It explores the common definitions, and their advantages or disadvantages. It then provides a vision for the future through a new, novel, state-based description of these quantities, in an attempt to better capture patient and cohort glycaemic behaviour. The methods, review, outcomes and analyses are generalisable to a range of clinical metrics in critical care.
2.0 State and Variability: Measures of Metabolic Quantities

2.1 State and Level of Blood Glucose

Blood glucose (BG) concentration, or its average across a population or time period or both, is the most common metric reported. Assumptions around appropriate concentrations and how they should be achieved underpin much of the glycaemic control research and clinical practice. Common thresholds for blood glucose and dysglycaemia differ in derivation and definition between different intensive care cohorts, and between adults and neonates in particular.

2.1.1 Adults

Adult hyperglycaemia is well researched and BG greater than 8.0 mmol/L is the typical threshold defining hyperglycaemia (Chase, Shaw, et al., 2008; Evans et al., 2012; J. Krinsley, 2003; J. S. Krinsley, 2004; Singer et al., 2009; Umpierrez et al., 2012; van den Berghe et al., 2001). This definition and the upper limit on GC target ranges varies within the literature from 6.1 – 9.0 mmol/L, where many individual clinicians, centres, and protocols define their own safety thresholds for normoglycaemia. One meta-analysis showed no particular advantage for GC ranges, although they found lower ranges were associated with higher incidence of hypoglycaemia (Yatabe, Inoue, Sakaguchi, & Egi, 2017). However, several other studies have shown lower ranges, while avoiding hypoglycaemia, offer reduced risk of death (J. S. Krinsley & Preiser, 2015; S. Penning et al., 2015; Signal, Le Compte, Shaw, et al., 2012). These differing outcomes, using different ranges, are a confounding factor in the pursuit of both consensus and reduced clinical hyperglycaemia.

Thresholds for mild and severe hypoglycaemia in adults are also well defined (Bagshaw et al., 2009; Cameron, Niemeyer, Gundy-Burlet, & Buckingham, 2008; Egi et al., 2010; Evans et al., 2012; Finfer et al., 2012; Hermanides et al., 2010; C. G. Pretty, Chase, Le Compte, Shaw, & Signal, 2010). The consensus is mild hypoglycaemia occurs at 4.0 – 4.4 mmol/L, while severe hypoglycaemia is most
commonly defined as BG < 2.2 mmol/L (Boom et al., 2014; van den Berghe et al., 2001). These thresholds result from clinical studies of decreased BG and the onset of physiological responses as a result of hypoglycaemia (Cryer, 1999), and they may be patient-specific (Boyle, Schwartz, Shah, Clutter, & Cryer, 1988). In particular, patients with type 2 diabetes and persistent hyperglycaemia can exhibit hypoglycaemic like symptoms at normal BG levels (Boyle et al., 1988).

2.1.2 Neonates

In contrast to the well reported clear limits of normoglycaemia for adults, normoglycaemia for infants is less intensively examined and defined. As a result, clinical practice and research definitions differ. Alsweiler et al. (Alsweiler, Kuschel, & Bloomfield, 2007) surveyed 27 tertiary neonatal units in Australasia for their definition and management of hyperglycaemia in very low birthweight infants. Variance in the clinical definitions of hyperglycaemia from these specific neonatal units ranged from 7.0 – 15.0 mmol/L, though most used 10.0 mmol/L.

Alsweiler et al. also observed large variance in target ranges for subsequent insulin therapy, though many defaulted to the normal ranges for adult glycaemia of 4.0 – 8.0 mmol/L. In some cases, the upper target ranged between 8.0 – 10.0 mmol/L. Some of the lower ends of these target ranges were much lower at 2.5-3.0 mmol/L, reflecting evidence of a possible tolerance for lower BG by infants (C. J. McKinlay et al., 2015; C. J. D. McKinlay et al., 2017). Thus there is overall little consensus for onset criteria and targets in neonatal GC, reducing effectiveness of protocols and ability to compare study outcomes.

In contrast to adult definitions, and also to the clinical thresholds observed by Alsweiler et al., the thresholds for neonatal hyperglycaemia used in literature by researchers is significantly lower, from 6.9-8.3 mmol/L (Cowett, Oh, Pollak, Schwartz, & Stonestreet, 1979; Hall, Peters, Eaton, & Pierro, 2004; Hays, Smith, & Sunehag, 2006; Louik, Mitchell, Epstein, & Shapiro, 1985; Vaucher & Walson, 1982). They are based on observed distributions of BG in term infants (Hey, 2005), where less common
extremes are used to define abnormal BG. However, these statistical definitions may simply describe measured BG in this cohort, rather than what may be clinically desirable, beneficial, or obtainable via GC (Hey, 2005).

Neonatal hypoglycaemia is controversial, where agreement has not been reached on either thresholds or treatment (Harris, Weston, & Harding, 2012; Harris, Weston, Signal, Chase, & Harding, 2013; C. J. McKinlay et al., 2015; Rozance & Hay, 2010), in part because it is often asymptomatic. Some studies show neonatal BG concentrations reach natural lows at approximately 2 hours after birth (Srinivasan, Pildes, Cattamanchi, Voora, & Lilien, 1986), while others fail to observe this nadir (Diwakar & Sasidhar, 2002; Hoseth, Joergensen, Ebbesen, & Moeller, 2000). Infants experiencing hypoglycaemia may also be asymptomatic (Cornblath et al., 2000; Griffiths, 1968; Griffiths & Bryant, 1971; Lucas, Morley, & Cole, 1988), unlike adults, and it is still unknown if asymptomatic hypoglycaemia is the same as symptomatic hypoglycaemia, or if such thresholds are patient-specific. However, neonatal hypoglycaemia (defined as <2.6 mmol/L) has been linked with poor neurologic outcomes in later life (C. J. D. McKinlay et al., 2017), and thus better definition in this area would lead to improved treatments and outcomes.

2.2 Measures of Glycaemic Variability

Though perhaps intuitive qualitatively, glycaemic variability is difficult to effectively quantify. Many reviews have undertaken to summarise variously measures of glycaemic variability, particularly in the context of Diabetes management, and more fully describe their advantages and limitations (DeVries, 2013; Rodbard, 2018; Service, 2013; Suh & Kim, 2015). Many metrics, and adaptations of metrics, exist, but in general most fall into the following broad categories: 1) descriptors of middle and range, 2) descriptors of total or summed variability or excursion length, and 3) descriptors of time in range.

The most standard methods for describing and quantifying glycaemic variability are statistical
descriptors of middle and range in data. The most common descriptor of variability is to report the
standard deviation (SD) of BG measurements alongside the mean BG value (Brunner, Adelsmayr,
Herkner, Madl, & Holzinger, 2012). Similarly, the IQR (inter-quartile range) is a non-parametric
alternative, reported alongside median BG. These metrics are popular because of their ease of use
(DeVries, 2013). However, the SD is a measure of dispersion rather than variability, and is limited in
its ability to reflect the time course of BG measurements (De Block et al., 2016).

The limitations of the SD for describing glycaemic variability are shown in Figure 1, where the mean
and SD are the same for two time courses of intermittent BG measurements. In this case, the same
set of BG measurements is arranged in two different patterns, one a decrease at a constant rate, and
the other moves between extremes. Because both have exactly the same measurement set, both have
the same mean and SD. If glycaemic variability is more truly a function of the change in BG, with more
rapid changes exerting different physiological effects than slower changes, than patient outcomes
could differ in these two cases. Thus, the SD is extremely limited in its ability to describe the time
course of glycaemic variability.

In addition, the mean and SD assume a normal distribution, which is inaccurate as BG is usually highly
lognormal and skewed. Thus, this measure for variability may be inappropriate to use, particularly if
there are a large number of measurements near the hypoglycaemic range of BG. Equally, all such
measures requiring a normal distribution assumption have this issue, and recent reviews recommend
non-parametric statistics (Chase et al., 2018).

The Coefficient of Variability (CoV), which is the mean divided by the standard deviation, is for some
a preferred measure of glycaemic variability (DeVries, 2013). CoV can be a good measure of overall
glycaemic variability, as research has shown that aiming for a CoV lower than a certain threshold (36%)
allows for the distinction between stable and unstable glycaemia (Hirsch, 2005; Monnier et al., 2017).
However, this metric suffers the same limitations as discussed regarding the mean and SD above. For the data in Figure 2, the CoV would by definition be the same between these two patients.

![Figure 1](image.png)

**Figure 1.** Two sensor traces with accompanying intermittent BG measures. The intermittent BG measures are the same between the two data sets, and so have the same mean and standard deviations, but display differing behaviours.

Metrics that look at total glucose excursion include Area under the BG curve (AUC) and Glucose Miles. AUC or area around a line is common in diabetes and non-critically ill cohort studies, often with more frequent continuous glucose monitoring (CGM) (Guido Freckmann et al., 2007; Pickup, Freeman, & Sutton, 2011; Salardi et al., 2002). Two zig-zagging sensor traces could theoretically have the same area, even though one trace is rising in level, while the other is falling. This behaviour would suggest the two traces have the same level of variability, but with differing behaviours, indicating how AUC is an overall measure of variance, but not specific to time course. This issue is also illustrated in Figure 2.

Glucose Miles is another way to measure variability, measuring the total ‘distance’ travelled by
movements in BG through intermittent BG or CGM traces. It is embedded as part of some other measures (McCall, Cox, Crean, Gloster, & Kovatchev, 2006; McDonnell, Donath, Vidmar, Werther, & Cameron, 2005). However, two traces can have the same Glucose Miles with very different mean BG, as shown in Figure 3 for a bias, and Figure 2 where the median is the same. Thus, this metric is similar to AUC in giving a total, but is not specific or reflective of the time-varying behaviour of blood glucose. Furthermore, while Glucose Miles and AUC can be useful descriptors of cohort variability when paired with measurements of mean or median BG, both have an inability to describe variation away from some longer term average or glycaemic state, which may be important to recognise clinically. Similar metrics, more suited to intermittent BG measurements, include the mean absolute difference (MAD), or mean of daily differences (MODD). Such metrics describe the average change in BG, or the difference in BG between days, but are limited in their ability to comprehensively describe the time course of variability and its relationship to the level, as per Figure 3.

Figure 2. Two sensor traces with very similar AUC, but differing behaviours.
Figure 3. Two sensor traces with similar Glucose Miles, but with very different mean BG.

Time in band is another typical measure of variance that is good for combining and capturing level and variability, though it tends to be more of an implicit measurement rather than explicit. It is a good aggregate measure for larger cohorts where central tendency makes time in band informative of overall cohort behaviour. Research has shown that time in band, or time in range, can be associated with clinical outcomes, in particular risk of microvascular complications (R. Beck et al., 2018). However, whether the measurement is inside or outside of the band itself is discrete or binary. Thus, all variability or measures in the band are assumed to be clinically acceptable, and those outside are not. It may be that a measure just inside the band is thus treated very different in analysis to one just outside, when both could be within measurement error. In addition, there is no agreement on appropriate bands, leading to difficulty in comparing the variability across studies (Chase et al., 2018). By example, Figure 4 shows two sensor traces for a 4.4 mmol/L to 8.0 mmol/L range, which have the same time in band but different, potentially clinically important behaviours both within and outside the band.
Other metrics for describing glycaemic variability exist, such as the mean amplitude of glucose excursions (MAGE), M-value, J-index, Low Blood Glucose Index (LBGI), High Blood Glucose Index (HBGI) and Average Daily Risk Range (ADRR) (DeVries, 2013; B. P. Kovatchev, Otto, Cox, Gonder-Frederick, & Clarke, 2006; Rodbard, 2018; Service, 2013; Suh & Kim, 2015). Each of these metrics attempts to compensate for some aspect of the limitations of the metrics previously discussed, and specifically attempt to take into account aspects of desirable glucose outcomes. Most of these metrics involves one or more aspects of the middle-range, time in band, or excursion size categories. All require some degree of correlation with outcomes to be clinically applicable, and have shown some degree of predictive capability (B. P. Kovatchev et al., 2006). While their disadvantages are presented elsewhere (DeVries, 2013; Rodbard, 2018; Service, 2013; Suh & Kim, 2015), the overall disadvantage of such metrics is that calculation and interpretation of these metrics is not transparent or intuitive to those unfamiliar with the metric, and this may have impeded widespread clinical uptake.

Furthermore, most of these glycaemic variability metrics have been derived using point glucose measurements, and most have only been validated using these intermittent BG measurements to
check the correlation with adverse outcomes. One of the metrics have been validated using CGM sensors, and was found to be correlated with increased hypoglycaemia given the measure of glycaemic variability over some threshold (Monnier et al., 2017), but more studies are needed for CGM sensor derived metrics and correlation with adverse outcomes.

In addition to the various metrics for variability that exist, variability in itself has been hard to define depending on what is clinically relevant. For example, observing the two sensor traces in Figure 1, on the timescale presented of 10 hours, the two sensor traces clearly have two differing levels of variability, i.e. CGM sensor trace 1 is far less variable than CGM sensor trace 2. However, stretching the timescale to a 24 hour period, the two sensor traces could be interpreted as having similar variability. Thus, there are currently no metrics that can adapt their timescales to take into account the full time course information that CGM sensors can provide.

In summary, the large range of glucose metric definitions are confusing, contentious, and complicated. No metric adequately describes the time-course of glycaemia. Variability in particular is poorly captured by commonly used metrics, most of which are mathematically rather than clinically centred and defined. These limitations are particularly important in the context of emerging CGM technologies, which provide greater time resolution with increasingly enhanced point accuracy (Crane et al., 2015). CGMs thus offer the potential to significantly improve GC. However, to obtain this benefit, the best clinically defined metrics to capture patient glycaemic level and state must be established, which may require a new approach to defining variability that is clinically focused rather than mathematically focused.
3.0 CGM Technology

CGM devices show significant as yet untapped potential for improving glycaemic monitoring and control, particularly due to their high measurement frequency. CGM sensors measure BG near continuously, with new generation devices able to measure at a rate of 1-6 times per minute (Crane et al., 2015; Schierenbeck, Franco-Cereceda, & Liska, 2017; Sechterberger, van der Voort, Strasma, & DeVries, 2015; Wollersheim et al., 2016), and standard devices providing measurements every 5 minutes (Akintola et al., 2015; R. W. Beck, Calhoun, & Kollman, 2012; Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study et al., 2008; B. Kovatchev, Anderson, Heinemann, & Clarke, 2008). Compared to 1-6 hourly point-of-care (POC) measurements in a well-staffed ICU (U Holzinger et al., 2006), these devices offer huge potential to improve care and reduce workload.

The increased measurement frequency has many benefits for care, including the ability to monitor patient condition and, importantly, the trajectory of their condition in real time. They also provide warning for hypoglycaemic events (McCall et al., 2006; C. G. Pretty et al., 2010), allowing early correction. Both benefits cannot be achieved with intermittent BG measures, where the minimum feasible regular measurement interval is ~1 hour, but clinical non-compliance can be high even at lesser rates (Campion, Waitman, Lorenzi, May, & Gadd, 2011; Carayon & Gurses, 2005; Chase, Andreassen, Jensen, & Shaw, 2008), as seen in a recent study where protocolised measurement interval was 1 hour and the clinical measurement was closer to 3 hours (Finfer et al., 2009; V. Uyttendaele, J. L. Dickson, G. Shaw, T. Desaive, & J. G. Chase, 2017). Even then, a patient’s condition may change significantly between hourly POC BG measures, resulting in hypoglycaemia remaining untreated for up to 50 minutes. A CGM sensor, on the other hand, can alarm both at occurrence and predictively before it occurs.

Despite these advantages, CGM sensors are not widely used in the ICU as CGM sensor technology still suffers limitations, including larger point error inaccuracies and sensor drift (Facchinetti et al., 2014;
Facchinetti, Del Favero, Sparacino, & Cobelli, 2015, 2016; Facchinetti, Sparacino, & Cobelli, 2010; Laguna, Rossetti, Ampudia-Blasco, Vehi, & Bondia, 2014; Lunn, Wei, & Hovorka, 2011; Rodbard, 2014; T. Zhou, Dickson, & Geoffrey Chase, 2018). They are also expensive, so it may be hard to justify the cost versus potential benefits. The larger point accuracy errors over traditional intermittent BG measurement techniques have been well documented, with new CGM devices usually reporting gradually improving MARD values up to 8-12% (Bochicchio et al., 2015; Bochicchio et al., 2017; Crane et al., 2015; Nohra et al., 2016) as the technology has developed (Bochicchio et al., 2015; Bochicchio et al., 2017; M. P. Christiansen et al., 2018; Crane et al., 2015; Damiano et al., 2014; Dungan, Han, Miele, Zeidan, & Weiland, 2012; Facchinetti et al., 2015; G. Freckmann et al., 2013; Kosiborod et al., 2014; B. Kovatchev et al., 2008; Lee et al., 2012; Luijf et al., 2013; Nohra et al., 2016; Saur et al., 2014; Wollersheim et al., 2016), compared to the 5% or lower common in intermittent POC measures (Critchell et al., 2007). Sensor drift is also still not widely recognised, even though it has shown to be a key driver of larger MARD and potential hypoglycaemia when used in GC (T. Zhou et al., 2018; Tony Zhou, Dickson, Shaw, & Chase, 2017).

CGM technology also has the potential to reduce GC related workload in the ICU, providing more bedside data with lower blood sampling requirements (Boom et al., 2014; U. Holzinger et al., 2010; Signal, Pretty, Chase, Le Compte, & Shaw, 2010; Tony Zhou et al., 2017). Clinical practices have cited high workload as a reason for reduced intermittent BG measures, as not every ICU has a low staff to patient ratio to justify the increased workload some GC protocols may require (Aragon, 2006; Bland et al., 2005; Gartemann et al., 2012). CGM sensors are able to give a continuous readout of the current BG measurement, and clinicians can adjust nutrition and insulin accordingly. However, this benefit has not yet reached regular care due to the limitations.

Overall, CGM technology has not reached regular clinical use in ICU care, despite common acceptance in lower-risk outpatient type 1 diabetes care (Barnard et al., 2017; Hovorka, 2011; B. Kovatchev et al.,
2017; B. Kovatchev, Tamborlane, Cefalu, & Cobelli, 2016; Lewis, Leibrand, & Community, 2016). As CGM technology improves it will be able to replace intermittent measures used in the ICU today. However, with increased temporal measures and benefits to control, the same issues of how to quantify level and state remain. In fact, with the increased measurement rate of CGM technology, these issues are exacerbated as measures that work for intermittent measures may not be representative or accurate for use with CGMs. There thus remains a significant need for better, more clinically defined metrics, and in particular, those able to maximise the measurements delivered by CGMs.
4.0 Vision for Glycaemic Control Metrics in the Future

Given the development of increasingly accurate CGM devices, which offer higher resolution glycaemic monitoring, and the limitations of current measures of glycaemic level and variability, improved methods must be developed to describe glycaemia. Such methods will be important for effective research into cohort and patient specific behaviours and the benefits and harms of dysglycaemia and insulin therapy. These metrics should be clinically defined and thus associated with clinical outcomes, avoid reliance on statistical assumptions, easy to implement computationally in real-time, and simple to understand to aid uptake.

Our vision for the future is a method for describing the time-course of glycaemic level and variability. Such a method would both improve reporting of glycaemic outcomes and better aid bedside glycaemic monitoring and treatment in patients. The measure would be able to fully utilise the increased resolution of modern CGM sensors allowing for analyses to relate the glycaemic states, and critically, any changes to these states, to clinical outcomes. The measure would be readily generalisable to other aspects of measurement and care, and would be easily compared across centres and cohorts under glycaemic control.
5.0 Glycaemic State Analysis

This section presents a novel method with the potential to fulfil the vision for the future defined. It develops a characterisation method for identifying Glycaemic States utilising CGM sensor data to monitor patient condition, both retrospectively, and in real-time. The algorithm builds on the idea of patient specificity, and also provides scope to adapt to changes in patient metabolism or clinical goals. The result is demonstrated on a cohort of 614 infants at risk for hypoglycaemia, and provides a first CGM-centric approach to simply quantifying Glycaemic State and variability in real-time.

5.1 Calculating Patient States

For each CGM trace, the mean of the whole CGM trace establishes a baseline average interstitial glucose (IG) level for that patient. The CGM data is then filtered using a centred 6-hour rolling average. Thus, there are 12x6-hour-average data points created using every hour of CGM data available (CGM sensor measurement frequency = 5 minutes), commencing 3 hours into the trace as the rolling average is calculated from the centre of the rolling window. The 6-hour rolling average is then bootstrapped by taking the 3 hour forward average of the CGM trace at the beginning to begin the rolling average. The rolling average is then rolled forward by one measurement, where the average of the next 3 hours of CGM data plus the first CGM data point is taken. The rolling average is then rolled forward again, and continues on until the bootstrapping meets the original 6-hour rolling average line. This is done in reverse for the end of the 6-hour rolling average line. Comparing the 6-hour rolling average to the arithmetic mean yields a baseline variation around the mean. Six hours is chosen because it filters out higher frequency glucose fluctuations, such as spikes just after feeding, and allows for any true, long term changes in average IG, characterised as a State Change in patient metabolic behaviour, to be shown.

Each time the 6-hour rolling average crosses the arithmetic average line, it is considered a possible
State Change if specific clinically defined conditions are met:

- More than 5 hours have passed since the last State Change, which assures “States” are 5 hours (or longer) periods of relatively constant average IG.
- The State Glycaemic Average, defined as the average IG for a given State, was more than 0.3 mmol/L higher or lower than the previous State Glycaemic Average, to reduce potential impacts of measurement error.

If either condition is not met, the CGM data for the current arithmetic average crossing is combined with the CGM data of the previous State, and that State’s Glycaemic Average recalculated using the longer CGM data. This State characterisation process is shown in Figure 5.

**Figure 5.** Steps of the Glycaemic State characterisation process, showing how the rolling average and States are calculated.

Importantly, while the 0.3 mmol/L threshold used here is based on clinical experience with neonates, it can be changed for any cohort or to find only larger changes. Similarly, a shorter or larger rolling average than 6-hours can be used, or they may be “nested” to find shorter State periods. Finally, State Changes per day (24 hours) are calculated to normalise results for comparison, but, equally, they could
be calculated in a roll-forward manner for use in real-time.

The median [IQR] hours of CGM per patient are calculated, alongside the average change in sensor measured interstitial glucose (IG) after a State Change, the maximum State Change, the minimum State Change, number of State Changes that resulted in a higher IG (also as a percentage), and the number of State Changes that resulted in a lower IG (also as a percentage). The median IG average is also calculated, alongside the IQR and 90% range, to allow for comparison of this cohort to others.

5.2 Case Study from a Neonatal Cohort

5.2.1 Subjects and Continuous Glucose Monitoring

The CHYLD Study recruited 614 infants born from 32 weeks gestation with one or more risk factors for neonatal hypoglycaemia, including the following: diabetic mother; preterm (<37 weeks); small (<10th centile or <2500 g); large (>90th centile or >4500 g); or acute illness (C. J. McKinlay et al., 2015). The aim was to examine the relationship between the incidence and severity of neonatal hypoglycaemia in at-risk infants and neurodevelopmental outcome in childhood (Harris et al., 2012; C. J. McKinlay et al., 2015). A total of 481 infants had an interstitial CGMS System Gold CGM sensor (Medtronic Inc., Northridge, CA) inserted soon after birth in the lateral thigh, as previously described (Harris, Battin, Weston, & Harding, 2010; Harris et al., 2012). Of these infants, 366 had more than 24 hours of CGM sensor data in the first 48 hours after birth, leading to 12356 total hours of CGM data (median [IQR]: 35.7 [30.5 38.4] hours/patient). Twelve further infants had more than 24 hours of CGM sensor data in the first 48 hours after birth, but had gaps in the data of more than 5 minutes (mean [IQR] = 8.1 [3.0 8.9] hours) and were thus excluded from the analysis. The CGM sensor recorded a measurement every 5 minutes, but results were masked and did not influence clinical care. CGM data were downloaded and recalibrated to all blood glucose concentrations, measured on a blood gas analyser (Signal, Le Compte, Harris, Weston, Harding, Chase, & Chyld Study, 2012). The study was approved by the New Zealand Northern Y Ethics Committee.
5.2.2 Results from a State and Variability Analysis

Table 1 presents the overall results of the State analysis algorithm. The number of State Changes experienced by each infant was calculated, along with the number of State Changes/day, average absolute change in IG over a State Change, minimum and maximum overall State Changes over the entire cohort, and the number of State Changes with higher and lower IG. The results reported here give greater depth to glycaemia, highlighting the frequency of glycaemically stable or variable patients.

Table 1. State Change analysis results.

<table>
<thead>
<tr>
<th>Patients</th>
<th>366</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total hours</td>
<td>12356</td>
</tr>
<tr>
<td>Hours/Patient (median) [IQR]</td>
<td>35.7 [30.5 38.4]</td>
</tr>
<tr>
<td>Median IG Average [IQR], (90% range) (mmol/L)</td>
<td>3.7 [3.3 4.1], (3.0-5.0)</td>
</tr>
<tr>
<td>Number patients with no State Changes</td>
<td>56 (15.3%)</td>
</tr>
<tr>
<td>1 State Change</td>
<td>177 (48.4%)</td>
</tr>
<tr>
<td>2 State Changes</td>
<td>81 (22.1%)</td>
</tr>
<tr>
<td>3 State Changes</td>
<td>46 (12.6%)</td>
</tr>
<tr>
<td>4 State Changes</td>
<td>5 (1.4%)</td>
</tr>
<tr>
<td>5 State Changes</td>
<td>1 (0.3%)</td>
</tr>
<tr>
<td>State Changes/day (median [IQR], (90% range))</td>
<td>0.65 [0.54 1.16], (0-1.73)</td>
</tr>
<tr>
<td>Median [IQR] absolute ΔIG State Change (mmol/L)</td>
<td>0.65 [0.45 0.92]</td>
</tr>
<tr>
<td>Max State Change (mmol/L)</td>
<td>3.7</td>
</tr>
<tr>
<td>Minimum State Change (mmol/L)</td>
<td>0.3</td>
</tr>
<tr>
<td>Number of State Changes from lower to higher average IG</td>
<td>311</td>
</tr>
<tr>
<td>Number of State Changes from higher to lower average IG</td>
<td>191</td>
</tr>
</tbody>
</table>

Figure 6 presents case studies from the State characterisation process. Figure 6(a) is a patient with 3 separate States identified. The initial patient State is low IG and rising, shown by the State Average of ~3.5 mmol/L. At around 25 hours after birth, the patient enters a State of higher average IG, but also higher variability, as shown by the large fluctuations in the original CGM signal about the State Average.
of ~4.4 mmol/L. At 38 hours after birth, the patient enters a steep decline in IG. However, this decline is not reflected in the rolling 6-hour average as a State Change until 40 hours after birth, where the rolling average intersects the IG average line. This State is less variable than the previous State and also has an average closer to the total IG average, implying the patient becomes more stable.

In Figure 6(b) only 2 States are identified, despite the CGM trace intersecting the IG average line frequently. This is because the resulting States are less than the 0.3 mmol/L threshold required to reach a new Glycaemic State, and so these are merged into one State. At approximately 33 hours, both the CGM trace and 6-hour rolling average decline enough, and for longer than the minimum 5 hour time frame, to describe a new, lower Glycaemic State. This patient has thus displayed a consistent and prolonged drop in glycaemia, which could be clinically significant.

Figure 6(c) shows a patient with similar States and State Changes to Patient 2 in Figure 6(a). The patient starts at a lower State, experiences a State Change to a higher level, and then drops again to a State Average closer to the overall IG average. In this case, high variability results in the rolling average crossing the IG average more frequently, although subsequent changes would not be significant (<0.3 mmol/L threshold). Clinically, it is of particular note that the IG spike at 37.4 hours after birth from 3.4 mmol/L to 7.7 mmol/L occurs in 20 minutes of measurements in the absence of parenteral dextrose boluses or buccal dextrose gel, and is likely to be sensor error, rather than a rise in true IG (Signal, Le Compte, Harris, Weston, Harding, Chase, & Grp, 2012), which the clinically defined approach discounts because the clinical definition of a State is far longer. In contrast, every other mathematically defined metric would be skewed by this error.

Figure 6(d) shows a relatively stable patient with regard to State, similar to Patient 3 in Figure 6(b), but with higher variability around that State. If a threshold >0.3 mmol/L was used, the algorithm would identify only one State. Thus, the choice of threshold depends on the granularity desired in defining
clinically significant States. It is possible to change the algorithm threshold to determine an optimum value for State differentiation based on clinical outcomes, interventions, and measurement error thresholds.

**Figure 6.** CGM data and Glycaemic State characterisation for Patients 2-7.
Finally, Figure 6(e) shows the State characterisation for an infant characterised by the algorithm as having 2 States, with moderate variability of the CGM trace. At the start of the CGM trace, IG starts low and appears to be hypoglycaemic for around 2.5 hours, appearing to be at a single State of low glucose concentrations. IG starts to rise back to normal levels at around 17 hours after birth. The 6-hour average crosses the IG average again at 43.5 hours, but does not exist long enough to define another State. Thus, only 2 States are characterised. However, different threshold choices or a change in beginning the moving average could capture such early potential States, if clinically relevant and desired, showing how the clinical definitions lead this algorithm and metric, thus better relating the metric and clinical goals.

Only 1 patient recorded 5 State Changes in the 48 hour period, the maximum number for this cohort with the State definitions used. This variability could be due to systematic sensor error or excessive sensor recalibration leading to an abnormal amount of quantified State Changes. Figure 6(f) shows this patient, where it is clearer the IG average is abnormally high for this cohort. In particular, in the last portion of the CGM trace the infant is hyperglycaemic.

5.2.3 Discussion of Case Study Results

The majority of infants experienced less than 2 State Changes in the first 48 hours of birth (233 of 366 patients, 64%), suggesting these patients at risk of hypoglycaemia remained in a relatively stable condition despite high rates of hypoglycaemia (Harris et al., 2010; Harris et al., 2012). There were more State Changes to a higher than lower IG. This bias may relate to fact that the CHYLD cohort recruited infants at risk of hypoglycaemia. Equally, BG tends to rise in days after birth as the infant’s metabolism stabilises and their ability to take up nutrition and absorb it develops.

Only ~25% of infants (96 out of 366) were parenterally fed. The majority were entirely enterally fed.
Thus, these State Changes are most likely not due to changes in parenteral feed rate, an outcome negated by the clinical choices of how a State is defined in size and length of time. The States, defined by the 0.3 mmol/L threshold chosen here, are by design larger than the expected rise due to a single enteral feed. The 5-hour minimum State length is greater than a feed interval. Thus, it is unlikely that feeds would cause false State Changes as defined and used here, which again relates the metric definition to the specific clinical situation and goals for the cohort and is not possible with other current metrics.

With respect to variability, it is captured first by State Changes per day, as discussed. It can also be further enhanced by examining area around the State level, a local AUC to compare or assess variability within a State. Similarly, a second level State analysis could be run with different clinical definitions to capture Sub-States. The method presented is thus clinically defined, as well as very flexible in this regard.

More adult CGM data is needed to compare to the infant results to see if there are any difference to level and variability between these two cohorts, or how much definitions of levels and States might need to vary between these cohorts. The method is generalisable enough to work with any CGM sensors and data, so using the method on other cohorts CGM data would yield valuable results to compare with this infant cohort or others. Equally, there is room to define the metric specific to clinical goals and cohorts, or even to compute multiple metrics simultaneously relevant to different clinical outcomes within a cohort.
6.0 Discussion

6.1 Comparison to Glycaemic Metrics

The method of characterising States and State Changes offers some advantages over the mean and standard deviation, median and IQR, CoV, AUC, Glucose Miles and time in band metrics. This method for level and variability characterisation allows for increased resolution and descriptiveness in the reporting of individual and cohort glycaemia. In particular, stability to a particular ‘State’, and the variability around that State, are easily reported and intuitively interpreted. A particular strength of the method is in its ability to capture any changes, as critically ill patients can experience rapid changes in metabolic dynamics (Dickson, Gunn, & Chase, 2014; Wong et al., 2006).

In comparison to mean and SD alone, or CoV, this method offers far greater insight to patient condition. As shown in Figure 2, two BG data sets can have the same mean and SD but very different glycaemic dynamics. A well-tuned state and variability analysis would distinguish between these States. Importantly, this method can distinguish between variability around a State, and variability between States, both of which have different time constants and thus may have different clinical implications. Similar outcomes hold regarding AUC and Glucose Miles.

The metric used in this study is considered clinically relevant because it is clinically defined, where most prior metrics are defined mathematically, applied to clinical data, and then associated with a clinical outcome of interest to show its clinical relevance. In this case, the methods presented allow clinicians to derive the variability that is relevant for the issues/situation they are examining, while still assessing it in a uniform mathematical fashion.

6.2 Potential Clinical Uses

As noted, potential clinical uses are widespread. Foremost, is real-time use to allow visual inspection
of changing glycaemia and variability around a State as it emerges and thus to track patient state. Such analysis could lead to better inputs to predictive models around patient outcomes (Del et al., 1995; Peres Bota, Melot, Lopes Ferreira, Nguyen Ba, & Vincent, 2002; Shaw & Chase, 2012; Tunnell, Millar, & Smith, 1998; Udekwu, Kromhout-Schiro, Vaslef, Baker, & Oller, 2004).

Research uses are the likely first choice, where they would enable better analysis between and across studies. The simple calculations mean any data set could be calibrated to local clinical norms for comparison or to a standard set of definitions. In both cases, better comparison would lead to better understanding of the success and failure of GC interventions, thus enabling optimised care.

### 6.3 Methodological Strengths and Limitations

An important strength of the method is that all aspects are clinically defined. Thus, the parameters can be changed in time or size to reflect the clinical behaviours to be captured. Hence, it is clinically defined and not dependent on any statistical distribution or assumptions. It can also be tuned to reflect clinical outcomes, where State thresholds may be adapted to reflect physiologically important changes in glycaemia.

The method does have some disadvantages. Currently, the method has only been used with retrospective data, and its use in real time and on other neonatal or adult cohorts has not been thoroughly analysed. Real time processing and updated States can be achieved with filtering and bootstrapping methods, and expansions to other cohorts simply requires the application of the method to available data. The part of the method where the whole sensor trace average is calculated and set as a baseline would also need to change, as the whole sensor trace would not be available in real time. However, the method can change slightly to be based off a moving whole IG average that continually updates as more data is received from the sensor. Although this may have room for some erratic behaviour at the start of data collection, the IG average should stabilise as more and more data
Another limitation of the method is that a new State can only be identified if the rolling average of the CGM sensor trace crosses the average glucose, IG. If, for example, a patient was highly variable in the short term above their IG, and then highly variable below their IG, only 2 States would be identified. This is partially affected by what the clinical parameters are set as for the method, for example setting shorter periods of interest can identify these shorter periods of variability, if that is what is of interest clinically. The variability within States can also be measured by combining with one of the metrics previously mentioned, e.g. Glucose Miles or AUC, if deemed appropriate for the analysis to be undertaken.

The method has not been applied to a data set with clinical outcomes to distinguish what State durations or thresholds are clinically significant. This future work will validate the overall usefulness of this method, and its contribution to the vision of improved descriptors of Glycaemic State and variability. Future work would also look to improve the method by making it robust to gaps in sensor readings using interpolation or other methods.

Reported outcomes are still a simplification of the time-course of states and variability. While the State trace can easily and intuitively show changes in states and variability, attempts to reduce its properties to reported numbers in results or tables will inevitably result in some loss of information and intuitive interpretation of results. However, this issue is always going to be the limitation of any method to describe variability as a summary metric over any time period, and this method has the potential to allow greater description of glycaemic results in tabular or numerical form. For these results in particular, CGM sensor traces with length between 24 hours and 48 hours long have been pooled together in Table 1, which may hinder easy interpretation of these results. Thus, the data for the median and IQR State Changes per day were included to mitigate this increase in difficulty in
interpretation.

7.0 Conclusions

The vision, analysis and review presented has addressed an emerging and critical aspect of glycaemic control, the need for consensus summary metrics of glycaemic level or State and variability. Current metrics have significant limitations as most are not clinically defined, and thus poorly represent many aspects needed for a control metric. It thus presents and defines the problem, current state of the art, and presents a vision for the future of what is required, including a first research effort to meet the goals defined. The overall work includes an ongoing focus to relate the ideas and current state back to traditional control systems methods and engineering approaches. There is still significant room for innovation and new approaches, particularly as more and more data becomes available to engineers, and clinicians become increasingly willing to take on more novel approaches driven by rapid changes in technology and the ability to measure and control patients. These changes will only occur more rapidly, driven in major part by the increasingly poor economics of healthcare mixed with the need to keep care more affordable via increasing automation.

8.0 Acknowledgements

The authors acknowledge the support of the EUFP7 and RSNZ Marie Curie IRSES program, the Health Research Council (HRC) of New Zealand, the MedTech CoRE and TEC, NZ National Science Challenge 7, Science for Technology and Innovation, Auckland Medical Foundation, the Eunice Kennedy Shriver National Institute of Child Health and Human Development, and the CHYLD Study investigators for sharing their CGM data.
9.0 References


10.1186/s13054-015-0908-7


