Clinical decision support: distance-based, and subgroup-discovery methods in intensive care
Nannings, B.

Citation for published version (APA):
Chapter 3. A SUBGROUP DISCOVERY APPROACH FOR SCRUTINIZING BLOOD GLUCOSE MANAGEMENT GUIDELINES BY THE IDENTIFICATION OF HYPERGLYCEMIA DETERMINANTS IN ICU PATIENTS


Barry Nannings, Robert-Jan Bosman, Ameen Abu-Hanna
3.1. Abstract

3.1.1 Objective
Despite the wide use of blood glucose management guidelines in Intensive Care (IC), hyperglycemia is still common. The aim of this study was the discovery of possible hyperglycemia determinants by applying the Patient Rule Induction Method (PRIM) to routinely collected data within the first 24 hours of admission, and to relate them to the literature.

3.1.2 Methods
PRIM was applied in two setups to data of 2,001 IC patients including 50,021 records of blood glucose levels and other variables. The independent predictors of blood glucose measurements were variables whose value is known before the time of the corresponding measurement, summarizing its “past”. These variables are candidates for inclusion in subgroup definitions and may constitute hyperglycemia determinants. Subgroups were validated using a random split design, and time-sensitivity of performance was analyzed. We compared our results to the literature.

3.1.3 Results
PRIM was able to identify relatively large subgroups having markedly high mean glucose values. Besides well-known determinants (e.g. the previous glucose value), PRIM also discovered possible determinants of which less is known about their relationship to hyperglycemia. Some possible determinants reported in the literature were not found by PRIM.

3.1.4 Conclusions
We demonstrated for the first time the utility of using subgroup discovery to uncover possible determinants for non-responsiveness to treatment. This implies the possible use of this technology to scrutinize the effects of various guidelines in clinical medicine on patient outcomes without requiring the development of a global predictive model. We hypothesize that by focusing on the identified subgroups, clinical guidelines may be improved. Further research is required to test this hypothesis.
3.2. Introduction

Glucose regulation is an increasingly important topic in Intensive Care (IC) where ways for improving guidelines to manage the blood glucose level are constantly sought. The landmark study by van den Berghe [1], which showed that normalization of the plasma glucose level of IC patients resulted in decreased morbidity and mortality, has been influential in setting up new guidelines for intensive-insulin therapy. Guidelines are however not always beneficial to all patients at all times, and providing tools to investigate the effects of guideline-based therapy on clinical outcomes is an important contribution of medical informatics research towards the improvement of guidelines.

The underlying biological mechanisms of glucose regulation are complex. The stress reaction of the body, in response to an injury, induces a release of hormones which increases hepatic glucose production [2]. The same hormones will inhibit insulin mediated glucose uptake to skeletal muscle [3]. Pre-existing diseases such as diabetes mellitus may contribute to hyperglycemia. Other factors to which hyperglycemia may be attributed, either reflect the severity of illness (e.g. acidosis, low potassium) or pertain to the treatment of the patient (e.g. the use of corticosteroids, diuretics, induced hypothermia) [2; 4-6]. An overview of important risk factors and determinants is given in [7].

To steer therapy, most recently suggested guidelines such as those described in [1; 8-13] rely primarily on the last measured glucose measurement, and sometimes the trend in previous glucose values and nutritional feed rates, but disregard other available clinical data. Although as a result of these guidelines the mean blood glucose level of the patient population as a whole might decrease, hyperglycemia is still often found in critically ill patients. A natural question to pose is which patients are at high risk of hyperglycemia despite having a blood glucose management guideline in place.

In this work we focused the search for such subgroups within glucose measurements from the first 24 hours because hyperglycemia is a prevalent problem in this period. This also means that a relatively large group of patients will be available.

The research presented in this paper is innovative for a number of reasons. First, we are not aware of efforts to apply the subgroup discovery algorithm PRIM to glucose data; our dataset is quite large, it contains time-oriented data, and is derived from guideline-based treatment. In addition, our research aims to help bridging the gap between a data-mining approach and the actual improvement of care whereas other work in the literature, focuses mainly only on one of these two.

3.3. Objectives

This paper is concerned with scrutinizing an intensive-insulin therapy guideline based on time-oriented data. The nature of time requires adequate representation of the data and
the validation of the acquired knowledge. The primary aim of this study was the identification of determinants of hyperglycemia, by means of the Patient Rule Induction Method (PRIM) [14], using commonly available clinical data residing in an Intensive Care Information System (ICIS), including laboratory results, vital signs and drug orders. Unlike current approaches for direct glucose level prediction based on modeling the underlying biological processes and the insulin resistance dynamics themselves [15] ours is aimed at focusing attention on observations that markedly deviate from the "rest" of the observations (in this case, ones with no hyperglycemia). The interpretation of these subgroups can provide insight into why some patients do not respond well to therapy and contribute to the improvement of treatment, e.g. by adjusting current guidelines to timely prevent the occurrence of these observations. A secondary aim was to investigate how our results relate to the literature on hyperglycemia risk factors and determinants.

3.4. Methods

3.4.1 Data

Between January 2005 and February 2006 data were prospectively collected in an 18-bed mixed general-surgical Intensive Care Unit (ICU) of a teaching hospital. All data were routinely collected for direct patient care in the ICIS (MetaVision®, iMD-soft, Tel Aviv Israel) and due to the design and the observational character of the study, obtaining informed consent was waived. Glucose regulation was performed through an algorithm incorporated in the ICIS as previously described [13]. The data included a total of 50,021 measurements of 2,001 patients collected during the patients' entire length of ICU stay.

![Flowchart showing the number of glucose measurements within the first 24 hours, and the developmental and validation sets.](image_url)
We used a split-sample design in which two thirds of the patients were randomly selected and all their measurements were assigned to the developmental set, and the measurements of the remaining patients were assigned to the validation set. See Fig. 1 for the respective set sizes. Characteristics of the patients in the developmental and validation sets are shown in Table 1.

The data included variables which were known at admission time to the ICU, among which were: admission type (medical, scheduled, unscheduled), acute renal failure, chronic renal insufficiency, chronic dialysis, cirrhosis, cardiovascular insufficiency, respiratory insufficiency, immunological insufficiency, burns, and whether the patient has pre-existing diabetes mellitus.

In addition to the data obtained at admittance we also included temporal variables that were repeatedly measured for each patient during the patient’s ICU stay. The outcome variable in our study is the measured plasma glucose, which is a temporal variable repeatedly sampled with an interval between two measurements ranging between 15 minutes and 4 hours. The independent variables include static variables and temporal variables whose value is known prior to the glucose measurement: For each glucose measurement taken at some time t we include values of other temporal variables obtained at times prior to t.

Based on expert opinion (of the second author of this paper) these temporal variables included:

- Glucose-related variables: the previous glucose value, the glucose trend based on the last two previous glucose measurements (mean change in concentration/min), and the average of the last three previous glucose measurements
- The most recent value during the last 6 hours of: bicarbonate, sodium and potassium
- The average value of: urine rate/hour, central temperature (both variables measured during the last 6 and also 2 hours), blood pressure, respiratory rate (both during the last 2 and 6 hours)
- The most recent value during the last 24 hours (that is, anytime prior to the measurement) of: albumin, white blood count, prothrombin time (PTT), thrombocyte count, and C-Reactive Protein (CRP)
- The binary variables corresponding to: whether renal replacement therapy was used during the last 12 hours, and whether corticosteroids were administered during the last 12 hours
- The insulin drip setting between previous and last glucose measurement.
Table 1. Patient characteristics in the total sample, the developmental set, and the validation set. Data are reported as mean ± SD, (median), interquartile range (25th to 75th percentiles). SAPS = Simplified Acute Physiology Score, APACHE = Acute Physiology and Chronic Health Evaluation, ICU = Intensive Care Unit

It should be noted that in addition to these temporal data preceding a glucose measurement, the static variables known at admission time are used as well. The static data and summaries of temporal data are dealt with in the same manner during the subgroup discovery procedure described below.

All data were collected in accordance to the Dutch National Intensive Care Evaluation (NICE) registry definitions [16]. To increase data quality, it was checked whether...
variables were within their value domains. A report on the quality of the data used in this study appeared in [17], although it should be noted that [17] reports on data from an earlier time-period.

For reasons to be described shortly, we also included a score variable reflecting the associated severity of illness for each of the continuous variables, with the exception of PTT, thrombocyte count and CRP, which could not be converted in the same way. Most of the scores were obtained by categorizing the continuous variables, using the Acute Physiology and Chronic Health Evaluation (APACHE) IV cut-off criteria [18]. Variables not included in the APACHE IV model were categorized according to criteria from APACHE II or Simplified Acute Physiology Score II (SAPS) [19], in this order. An example of categorization is converting a patient’s maximum body temperature value of 40 °C into a severity score of 4 units by using the APACHE IV categorization criteria. These categorizations, which result in ordered numeric scores, allow us to group very high and very low values together in a single condition. For uncovering the determinants of hyperglycemia, the algorithm has a choice between using a severity score and the raw data on which it is based, and although unlikely, it can also choose to use both.

We also had data needed to describe the patient sample and/or the subgroups such as mortality, length of stay, and scoring systems (APACHE II and SAPS II scores). These variables reflect outcome measures or, as in the case of scoring systems, their values can be calculated only after 24 hours of stay have elapsed.

3.4.2 Subgroup discovery

The Patient Rule Induction Method (PRIM) is a method proposed by Friedman and Fisher [14] that seeks subgroups in a high dimensional dataset having a markedly higher (or lower) value of an outcome than in the total sample. Initially PRIM includes all available observations (in our case the individual glucose measurements) in what is referred to as a box. It then attempts to shrink the box iteratively at either one side of the box, by peeling off a percentage ($q$) of the data of one of the variables, such that there is maximum increase in the mean at each successive sub-box (this is the procedure for continuous variables, it is slightly different for categorical variables). That is, at each step it considers the tails at the $q$ and $1-q$ quantiles of each variable’s distribution and removes the data under the tail rendering the highest mean sub-box. ‘Peeling’ continues until a user-specified minimum number of observations in the box is reached. At this point the PRIM algorithm performs a local inverse procedure to ‘peeling’ called ‘pasting’ aiming at recovering from possible sub-optimal choices made during the ‘peeling’ process. The term ‘patient’ in the algorithm refers to the fact that ‘peeling’ removes only a small proportion of the observations in each step.

The algorithm is formally described in [14] and the following example aims to illustrate its use in our application: Consider 100 patient records describing the weight and gender of 100 patients. Over a period of 24 hours body temperature is recorded at various times, say each hour. Similarly each patient’s BGL is recorded 10 times (for the sake of simplicity) over this period. Each glucose value will be associated with a glucose record,
resulting in 1000 glucose records in total. Each glucose record is described by the corresponding BGL measurement (the outcome) and by the weight, gender and a summary (e.g. mean in the last 6 hours) of all temperature measurements prior to the time of obtaining the BGL value. Let us configure PRIM to peel 1% of the data ($\alpha = 0.01$). The whole data comprises the initial box, and at the outset PRIM considers all the following candidate peeling operations: removing 1% of the records having the lowest weight; removing 1% of the records having the highest weight; removing 1% of the records with the lowest mean temperature; removing 1% of the records with the highest mean temperature; removing all records of male patients (operations on binary and categorical variables do not consider $\alpha$); and removing all records of the female patients. For each of the obtained subgroups PRIM calculates the mean BGL of the resulting box and it will retain the box with the maximum BGL value. The procedure is repeated recursively, with the peeling parameter still set to 1% (but now of the observations remaining after the peeling operation).

PRIM was used on the developmental data to find subgroups of high glucose measurements. Recall that only variables whose values are known prior to a glucose measurement, such as the previous glucose measurement, are considered. Subgroups that PRIM generates are described using conjunctive conditions. For example, a subgroup of measurements with a predicted glucose value $> 11$ mmol/l may be described by “temperature $< 36$ °C and the admission type is medical”. It cannot however, generate a rule using disjunctions on continuous variables such as “blood pressure $> 90$ or heart rate $> 110$”, nor “blood pressure $< 70$ or $> 90$”. However, the latter type of composite condition represents a variable-outcome relationship that is common in medicine in which a low and a high value of a variable is associated with adverse outcomes, while values in-between are associated with a normal value of the outcome. In order to generate conditions implicitly capturing this typical variable-outcome relationship, we included the categorizations of the continuous variables as described above. For example PRIM is able to generate a condition such as: “the severity/abnormality score of body temperature is greater than 4” which implicitly covers the respective high and low values of body temperature.

PRIM does not require the imputation of missing values. They are treated as illustrated in the following example: If the subgroup definition is: “bicarbonate $< 26$ mmol/l and temperature $> 30$ °C” it would include glucose measurements where bicarbonate and/or temperature are missing. The idea behind this is that if it really mattered for the subgroup to exclude missing values of a variable, PRIM would generate a rule explicitly excluding the missing value, e.g. “bicarbonate $< 26$ mmol/l and bicarbonate is not missing”. However, to avoid uncertainty, in our calculation of subgroup performance in both the developmental and validation sets, we excluded glucose measurements having missing values for variables defining a subgroup.

The implementation of PRIM that was used is called SuperGEM™ 1.0 [20]. PRIM was applied on the developmental dataset in two different setups. In Setup 1 PRIM was applied on the measurements in the developmental dataset using all input variables but excluding the glucose-variables: the previous glucose measurement, the mean of the
previous three glucose measurements and the glucose trend. Setup 1 is aimed at the
discovery of determinants other than glucose.

In Setup 2 PRIM was applied on the same dataset as in Setup 1 but with the inclusion of
the glucose-variables. Comparison of subgroups from Setup 1 and Setup 2 can provide
insight in the relative strength of the determinants. In both setups PRIM was run multiple
times, each time after the exclusion of measurements that were part of previously found
subgroups. In each run of PRIM we searched for a subgroup in the (remaining)
measurements in the development set covering at least 5 percent of the
measurements, and having a mean glucose value of at least 9 mmol/l, as chosen by the
clinical expert. Recall that in the development as well as the validation sets we exclude
missing values when reporting the number of measurements and their mean glucose
values in a subgroup. Hence, the percentage of the measurements considered for the
calculation of the mean may turn out to be slightly less than 5% of the development or
for the validation set.

For Setup 2, only the first two subgroups are reported in this paper, even though more
subgroups meeting the prerequisite minimum coverage of 5% and the minimum glucose
concentration of 9 mmol/l could be found. This is because the primary goal of Setup 2
was to understand the role of previous glucose, as a determinant, on the current BGL
when compared to Setup 1. It turned out that the third and later subgroups (that are not
shown in this paper) repeatedly rendered the previous glucose measurement as the
most important variable for the current glucose level, with a lower cut-off value in each
successive rule – no new variables of interest were hence discovered.

3.4.3 Validation

We validated the subgroup descriptions, in terms of size and mean, which were
generated from the development set on an independent validation set. In addition, we
performed an analysis to investigate the time-dependency of a subgroup’s mean
because we wanted to inspect whether observations in a subgroup have consistently
markedly higher glucose values over the whole period of a day, and not only in an
arbitrary interval thereof. We therefore applied the subgroup description to the validation
set within a sliding window of 4 hours width: all the measurements falling in the subgroup
in these four hours are obtained and their mean is calculated. The window was then slid
forward one hour and the procedure reapplied to the measurements falling between the
2nd and 5th hour after admission. This procedure was repeated 20 times covering the
first 24 hours of stay. Note that sliding the window may alter the composition of patients
corresponding to the measurements in the subgroup. The proportion of patients whose
measurements are part of the subgroup, with respect to the total number of patients
which had at least one measurement during the chosen window was also calculated.
Consider the following example. There might be a total of 100 patients staying at the ICU
between 6 and 10 hours after admission. Of these 100 patients, 80 had measurements
in this timeframe and only 40 of them had measurements falling within the subgroup in
this timeframe. This will result in a “subgroup to total” proportion in that timeframe of
50%.
3.5. Results

This section describes the subgroups discovered using PRIM, their validation, analysis of variable strength, and time-sensitivity investigation of the mean blood glucose level. When scores were part of the definition, the scores have been translated to the matching real attribute values. The only time a score was used to define a subgroup was in subgroup 2 of Setup 1.

In Setup 1 the first identified subgroup of glucose measurements had a mean of 12.5 mmol/l and was defined as follows:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Variable description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature &lt; 35.5 °C</td>
<td>Mean body temperature during the last 6 hours</td>
</tr>
<tr>
<td>Bicarbonate &lt; 14.9 mmol/l</td>
<td>Most recent bicarbonate measurement during the last 6 hours</td>
</tr>
</tbody>
</table>

The second identified subgroup in Setup 1, after removing the glucose measurements that were part of subgroup 1, had a mean of 9.1 mmol/l and was defined as:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Variable description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicarbonate &lt; 20.5 mmol/l</td>
<td>The most recent bicarbonate measurement during the last 6 hours</td>
</tr>
<tr>
<td>Admission type = medical</td>
<td>Medical reason for admission</td>
</tr>
<tr>
<td>Urine &lt; 2 l OR Urine &gt; 4 l</td>
<td>The amount of urine after 24 hours after extrapolation from average urine rate from the last 12 hours</td>
</tr>
<tr>
<td>21.5 &lt; Albumin &lt; 38.5 g/l</td>
<td>The most recent albumin during the last 24 hours</td>
</tr>
<tr>
<td>Temperature &lt; 36.85 °C</td>
<td>The mean body temperature during the last 6 hours</td>
</tr>
</tbody>
</table>

This is the only subgroup with a definition containing a score attribute (Urine severity score > 0), this can be deduced from the ‘OR’ in the definition (a definition using the actual score would be something like ‘score > x’).

The first identified subgroup in Setup 2 had a mean of 15.1 mmol/l and was defined as follows:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Variable description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous glucose &gt; 13.2 mmol</td>
<td>The previous glucose measurement</td>
</tr>
<tr>
<td>Bicarbonate &lt; 26 mmol/l</td>
<td>The most recent bicarbonate measurement during the last 6 hours</td>
</tr>
</tbody>
</table>

The second identified subgroup in Setup 2, after removing the glucose measurements that were part of subgroup 1, had a mean of 10.5 mmol/l and was defined as:
Table 2 characterizes the subgroups discovered. Table 3 displays the relative strength of each of the variables in the subgroup definitions by showing what the subgroup glucose mean would become if they were to be removed from the subgroup definition. In addition, Table 3 shows the percentages of missing values in the total dataset for the variables used to form the subgroups.

Table 2. Outcomes (average Glucose mmol/l) of the subgroups discovered. Subgroup 1 refers to the first subgroup discovered by PRIM while Subgroup 2 refers to the second group discovered after removing measurements belonging to Subgroup 1. Setup 1 refers to subgroups generated by excluding variables directly related to previous glucose measurements, while Setup 2 refers to subgroups based on all variables including variables directly related to the previous glucose measurements.

Fig. 2 displays the results of the time-sensitivity investigation of the mean glucose value based on the sliding window approach. In both setups, as time progresses, fewer measurements are available, and a negative trend can be seen. It can also be seen that the second subgroup of Setup 1 does not differ much from the mean of the remaining measurements (after excluding the measurements in subgroup 1).
3.6. Discussion

One of our clinical findings, based on the experiments in Setup 1, is that low bicarbonate and low body temperature form important physiological candidate determinants for hyperglycemia during insulin therapy. The performance of glucose management guidelines, may perhaps improve as a result of considering the values of these additional variables. Further research is necessary to investigate this. It is however unclear to us, given the available data, which of the patients have received hypothermic therapy, and as such we cannot make strong statements regarding the influence of this therapy on our results.

Another clinical finding, based on the results of Setup 2 is that a very high last value of the glucose level (> 13 mmol/l) is a main predictor for having a very high value of glucose in the next measurement. This is evidence for the utility of the common current use of the last glucose value as indicator to steer glucose regulation.

<table>
<thead>
<tr>
<th>Setup 1 (Glucose variables excluded)</th>
<th>Setup 2 (Glucose variables included)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable removed (%) missing in total data</td>
<td>Resulting Glucose mean</td>
</tr>
<tr>
<td>Subgroup 1</td>
<td>none</td>
</tr>
<tr>
<td>Body temperature (7%)</td>
<td>11.0</td>
</tr>
<tr>
<td>Bicarbonate (12%)</td>
<td>9.4</td>
</tr>
<tr>
<td>Subgroup 2</td>
<td>none</td>
</tr>
<tr>
<td>Bicarbonate (12%)</td>
<td>8.5</td>
</tr>
<tr>
<td>Admission type (0%)</td>
<td>8.8</td>
</tr>
<tr>
<td>Urine (1%)</td>
<td>9.1</td>
</tr>
<tr>
<td>Albumin (50%)</td>
<td>8.7</td>
</tr>
<tr>
<td>Body temperature (7%)</td>
<td>8.8</td>
</tr>
</tbody>
</table>

Table 3. The relative importance of variables defining a subgroup. The outcome values (glucose mmol/l) adjacent to a variable are obtained when the variable is removed from the definition of a subgroup. As an example, if the second variable of the second subgroup of Setup 1 would be removed (admission type), the mean glucose of the subgroup in the validation set would be 8.8 mmol/l. The statistics concerning the second subgroups in both setups are based on the measurements remaining after removing measurements belonging to the first subgroups.
The investigation of time-dependency of the subgroups showed that subgroups remained interesting during the first 24 hours of admission when related to the total sample mean. However, a negative trend could be discerned. This can be partly explained because the frequency of performing glucose measurements is generally higher near admission time. Because the mean glucose value of subgroup 2 in Setup 1 is only slightly higher than the mean of all glucose measurements, the added value of variables defining it as determinants should be further scrutinized.

How do our results relate to the medical literature? The use of the previous glucose in most of the glucose management guidelines is advocated in the literature [1, 8-13]. As indicated, this is also supported by our results. Our results are also concordant with the literature on temperature and bicarbonate [5, 21]. Surprisingly the glucose trend and the average of a number of previous glucose measurements, as we chose to represent them, did not provide much added value. Strangely, the often reported relation of hyperglycemia with the use of corticosteroids was not confirmed in our results [22]. This may be explained because, in accordance with the guideline used, most of the patients in this ICU received corticosteroids before or in the first hours of ICU admission. Also, the levels of steroids in the sample were already quite high, perhaps explaining why the use of steroids was not found to be a possible determinant.

Figure 2. The results of sliding a 4-hour window across time. The aim is to investigate whether subgroups are interesting during the entire 24 hours period from which the measurements were obtained. The centre and size of a circle represent, respectively, the mean glucose values and the number of measurements in a
subgroup in the corresponding timeframe. The height of a bar represents the number of patients corresponding to these measurements. The number appearing above a bar denotes the proportion of these patients among all patients having any measurement within the same time window. The dotted line, stretching from left to right, shows the mean of all glucose measurements within a timeframe; including those not part of a subgroup. The statistics on the second subgroups, in both setups, are based on the measurements remaining after excluding those belonging to the first subgroups. To help understand the figures, consider the timeframe labeled A at the top-left figure. The timeframe corresponding to A consists of all measurements between 9 and 13 hours after admission, the mean glucose within the subgroup is 12.9 mmol/l. There are 37 measurements in the subgroup of a total of 13 patients, amounting to 2 percent of the patients that had glucose measurements during this timeframe.

We found no relation between renal replacement therapy and glucose levels. Pre-existing diabetes mellitus was also not found to be an important determinant of hyperglycemia at the ICU. Factors which did have (a small) influence in our results and have not been reported before are albumin and the admission type to the ICU. Lower albumin serum levels related to (blood) loss and dilution due to fluid resuscitation are routinely found in intensive care patients.

Unlike many other statistical methods, PRIM is non-parametric, that is, it does not assume a pre-specified form of the association between predictors and outcomes, nor does it make distributional assumptions about the variables. Another non-parametric method, which in theory could be used to find subgroups of high glucose, is Classification and Regression Trees (CART) [23]. PRIM focuses, in our application, on discovery of only groups having a markedly higher glucose value. CART would fit a model for the whole sample (of observations). It might show all interesting subgroups at once but at the risk of sacrificing the quality of high-risk subgroups in favor of the quality of the whole model. A concrete example of this is when CART would not further split a set of observations which would have resulted in one small but interesting group when the quality of the other large group is not sufficiently improved.

Two other relevant subgroup discovery algorithms are typified by the work described in [24] and [25]. In [24], Lavrac and colleagues present the CN2-SD algorithm. It is an adaptation of the CN2 classification rule learner [26] to search for statistically deviant groups. That idea has also been applied to adapting association rule learning to subgroup discovery in the APRIORI-SD algorithm after the categorization of the input variables [27]. CN2-SD, which works on a binary outcome, performs a beam search in which Boolean conditions are combined with the AND operator to arrive at a subgroup description. For continuous variables such a condition is of the form: attribute < cut-off-value or attribute > cut-off-value. The cut-off values in these conditions are calculated from the data by first sorting the attribute values and then finding those values in which the associated class alters its value (i.e. switches from 0 to 1 or from 1 to 0). Aside from the focus in CN2-SD on binary outcomes, the main difference between PRIM and CN2-SD is in the search procedure. While PRIM makes only small steps toward the final subgroup definition only allowing small adjustments to its condition in each step, CN2-SD attempts to find the "best" cut-off point for an attribute. This makes CN2-SD
“greedier” in its search. However, to compensate, CN2-SD uses a beam-search in which multiple preliminary rules are stored for further evaluation although it is unclear which beam width is reasonable and perhaps it should be found by experimentation. Furthermore, CN2-SD does not apply the “pasting procedure” of PRIM that attempts to locally optimize the box to alleviate earlier sub-optimal choices made in the vicinity of the box. While PRIM removes subgroups before searching for new ones, CN2-SD provides a weighting mechanism to discourage the inclusion of old observations found in new subgroups. The Data Surveyor algorithm described in [25] is similar to CN2-SD in terms of the search algorithm, however, it seeks conditions of the form: lower-value < attribute < upper-value making the algorithm even “greedier” than CN2-SD. Data Surveyor and CN2-SD will in general be much faster than PRIM to arrive at a result but run a higher risk of missing an interesting subgroup. Further work consists of comparing CN2-SD, Data Surveyor and PRIM in various circumstances.

Formalisms, such as CART, that are able to express a split in “the middle of a box” would result in representational economy (one split would correspond to finding more than one subgroup in PRIM). However this representational economy comes at the expense of data fragmentation. Interestingly, extensions of the basic PRIM algorithm have also been described in [14] where regions of observations, other than the side of the current box, are also allowed to be removed. We have however not attempted this strategy.

Using PRIM is not new in itself although it is surprising that there are only very few applications described in the literature, most of them in Bioinformatics such as that described in [28]. In earlier work [29] we applied PRIM to identify patients having a high risk of mortality from an elderly IC population from a large dataset originating from various intensive care units in The Netherlands. The current work is different in at least two main aspects to [29]. First, we use PRIM to scrutinize clinical guidelines searching for non-responsive groups, indicating how medical informatics methods might be applied toward improvement of clinical guidelines. Second, the use of time-oriented data necessitates data abstractions and time-sensitivity analysis of subgroups as shown in Fig. 2. Our current abstractions of time-variant monitoring variables have focused on simple statistical summaries (like the mean of body temperature in the last 6 hours or the most recent albumin value) obtained from each variable separately. Such summaries may overlook relevant temporal characteristics of the signals such as trends, and the inter-relationships between them. Further work consists hence of investigating multivariate temporal patterns and the use of more expressive temporal abstractions. A good starting point for conducting such research is the framework described in [30].

This study has a number of limitations. First, the non exhaustive search for subgroups is not guaranteed to find the best subgroups, and adjustment of the parameter settings of the algorithm may further improve results. The results should therefore be considered as a set of validated subgroups associated with very high glucose value but does not necessarily include the set of the best possible subgroups.
Second, blood glucose measurements of the same patient are treated as independent observations without adjusting for their inter-correlations. This means that a patient may have more than one measurement (adjacent or not) in a subgroup and in this sense biases the results. However there is an important mitigating circumstance in this application: a measurement implies action (insulin provision) and hence if a problem persists (high BGL) even after the provision of more insulin then the seemingly “over-representation” of patients might be beneficial, depending on the goal of the analysis.

Third, the results are obtained from data generated during the glucose management of all consecutive patients of only a single ICU. The glucose management guideline used in this ICU, which is described in [13], shows a strong resemblance to the guideline suggested by van den Berghe [1] and is adopted in many ICUs. Though clinicians are expected to follow the guideline; it is unlikely that it was always followed. Adherence to the guideline may be a confounder in the analysis. If data are available on adherence one could first stratify the sample into a group where adherence was high and another in which it was not, and then perform PRIM analysis on each of them separately to try to isolate the effect of adherence on the results. It should also be noted that we only used data originating from the first 24 hours of stay; our results may not apply to periods beyond the first 24 hours.

3.7. Conclusions
As far as we know this is the first time the idea of subgroup discovery is linked to the identification of determinants of inadequate response to therapy. This is a powerful link as an increasing number of therapies are governed by guidelines, and this link allows one to investigate the effectiveness and/or efficiency, e.g. over time, of guidelines in terms of clinical outcomes. We demonstrated this idea in the identification of determinants which may be of use to further understand the glucose metabolism and possibly improve the current glucose management guidelines. PRIM proved to be useful in discovering subgroups whose interpretability agrees with clinical intuition and its application deserves much more attention than it is currently given in the literature. Our application should, however, be seen as an exploratory effort to understand the determinants of hyperglycemia, and further research is needed to investigate how guidelines can be improved in light of the discovered subgroups and what the benefits are in terms of patient outcomes.

3.8. Acknowledgements
This work was performed within the ICT Breakthrough Project “KSYOS Health Management Research”, which is funded by the grants scheme for technological co-operation of the Dutch Ministry of Economic Affairs and is also supported by the Netherlands Organization for Scientific Research (NWO) under the I-Catcher project, number 634.000.020.
3.9. References


