AALBORG UNIVERSITY

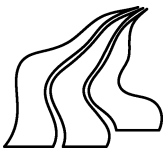
Prediction of the Insulin Sensitivity Index using Bayesian Network

by

Susanne G. Bøttcher and Claus Dethlefsen

May 2004 R-2004-14

Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7 G ▪ DK - 9220 Aalborg Øst ▪ Denmark
Phone: +45 96 35 80 80 ▪ Telefax: +45 98 15 81 29
URL: www.math.aau.dk/research/reports/reports.htm



ISSN 1399–2503 ▪ On-line version ISSN 1601–7811

Prediction of the Insulin Sensitivity Index using Bayesian Networks

Susanne G. Bøttcher and Claus Dethlefsen
Department of Mathematical Sciences
Aalborg University
Fr. Bajers Vej 7G
9220 Aalborg East
Denmark

Abstract

The insulin sensitivity index (S_I) can be used in assessing the risk of developing type 2 diabetes. An intravenous study is used to determine S_I using Bergmans minimal model. However, an intravenous study is time consuming and expensive and therefore not suitable for large scale epidemiological studies. In this paper we learn the parameters and structure of several Bayesian networks relating measurements from an oral glucose tolerance test to the insulin sensitivity index determined from an intravenous study on the same individuals. The networks can then be used in prediction of S_I from an oral glucose tolerance test instead of an intravenous study. The methodology is applied to a dataset with 187 patients. We find that the S_I values from this study are highly correlated to the S_I values determined from the intravenous study.

1 Introduction

Type 2 diabetes is a clinical syndrome that can result from several disorders that interfere with insulin secretion and/or the ability of the target tissues to respond to insulin. Martin, Warram, Krolewski, Bergman, Soeldner and Kahn (1992) found evidence in a 25 year follow-up study that the insulin sensitivity index (S_I) can be used to predict the development of type 2 diabetes up to a decade before diagnosis. Assessment of S_I is by Bergmans

minimal model, see Bergman, Ider, Bowden and Cobelli (1979), which is based on data from an intravenous glucose tolerance test (IVGTT). In the minimal model, the glucose and insulin kinetics are separately described by two sets of differential equations. The parameters in the model are traditionally estimated by a non-linear weighted least squares estimation technique, see for example Pacini and Bergman (1986). From these parameters, S_I can be determined.

However, an IVGTT is time consuming and expensive and therefore not suitable for large scale epidemiological studies. Interest is therefore in developing a method to assess the insulin sensitivity index from an oral glucose tolerance test (OGTT).

In Drivsholm, Hansen, Urhammer, Palacios, Vølund, Borch-Johnsen and Pedersen (2003), multiple linear regression is used to derive predictive values of S_I from measurements from an OGTT. These are compared with the values of S_I obtained from an IVGTT and calculated using Bergmans minimal model. The results show that it is possible to predict estimates of S_I, which are highly correlated to IVGTT-derived S_I for subjects with normal glucose tolerance.

In this paper, we express the relation between the observed variables in a Bayesian network. We try different approaches of establishing a Bayesian network, which can be used to predict S_I from measurements from an OGTT. We learn the parameters and structure of a Bayesian network from a training data set, where all patients underwent both an IVGTT and an OGTT. Bergmans minimal model were used to determine S_I from the IVGTT. We then calculate the predictive value of S_I from the Bayesian network and compare it with the value of S_I obtained from the IVGTT.

Like the multiple linear regression approach, the Bayesian network approach gives predictions of S_I that are highly correlated to IVGTT-derived S_I for subjects with normal glucose tolerance. In addition, the complex dependency structure between the variables is modeled adequately. Further, using Bayesian networks makes it possible to incorporate any prior information available, *e.g.* the physiological understanding of the problem or results from previous studies.

2 Data

In this paper we consider 187 non-diabetic glucose tolerant subjects, with one parent having diabetes. All the subjects underwent a 75 gram frequently

sampled OGTT. In such a test, the subject drinks 75 gram fluent glucose, after a 12 hour overnight fast. Venous blood samples are then drawn at 10, 5 and 0 minutes before the OGTT and after the start of the OGTT, at 10, 20, 30, 40, 50, 60, 75, 90, 105, 120, 140, 160, 180, 210 and 240 minutes. From these blood samples, the glucose and insulin concentrations are determined.

Within one week after the OGTT examination, all subjects underwent a tolbutamide modified frequently sampled IVGTT. In an IVGTT, glucose is injected directly into the venous. Blood samples are drawn at 10, 5 and 0 minutes before the injection and frequently up until 180 minutes after the injection. At 20 minutes, a bolus of tolbutamide is injected to elicit secondary pancreatic beta cell response. In the time between the two examinations, the subjects were asked not to change their lifestyle. The insulin sensitivity index (S_I) was for each subject calculated from the observations in the IVGTT using Bergmans minimal model and estimated by a non-linear weighted least squares estimation technique, as described in Pacini and Bergman (1986).

Other variables in the study are age, sex, weight, height, waist circumference, hip circumference, fat mass and information on physical activity. From the weight and height, the body mass index (BMI) can be calculated.

3 Bayesian Networks

We perform the analysis using Bayesian networks for discrete and continuous variables in which the joint distribution of all the variables are conditional Gaussian (CG), see Lauritzen (1992).

3.1 Bayesian Networks with Mixed Variables

Let $D = (V, E)$ be a Directed Acyclic Graph (DAG), where V is a finite set of nodes and E is a finite set of directed edges (arrows) between the nodes. The DAG defines the structure of the Bayesian network. To each node $v \in V$ in the graph corresponds a random variable X_v. The set of variables associated with the graph D is then $X = (X_v)_{v \in V}$. Often, we do not distinguish between a variable X_v and the corresponding node v. To each node v with parents $\mathrm{pa}(v)$, a local probability distribution, $p(x_v|x_{\mathrm{pa}(v)})$ is attached. The set of local probability distributions for all variables in the network is $\mathcal{P}$. A Bayesian network for a set of random variables X is then the pair $(D, \mathcal{P})$.

The possible lack of directed edges in D encodes conditional independencies between the random variables X through the factorization of the joint probability distribution,

$$p(x) = \prod_{v \in V} p\big(x_v | x_{\mathrm{pa}(v)}\big). \tag{1}$$

Here, we allow Bayesian networks with both discrete and continuous variables, as treated in Lauritzen (1992), so the set of nodes V is given by $V = \Delta \cup \Gamma$, where Δ and Γ are the sets of discrete and continuous nodes, respectively. The set of variables X can then be denoted $X = (X_v)_{v \in V} = (I, Y) = ((I_\delta)_{\delta \in \Delta}, (Y_\gamma)_{\gamma \in \Gamma})$, where I and Y are the sets of discrete and continuous variables, respectively. For a discrete variable, δ, we let $\mathcal{I}_\delta$ denote the set of levels.

To ensure availability of exact local computation methods, we do not allow discrete variables to have continuous parents. The joint probability distribution then factorizes into a discrete part and a mixed part, so

$$p(x) = p(i, y) = \prod_{\delta \in \Delta} p\big(i_\delta | i_{\mathrm{pa}(\delta)}\big) \prod_{\gamma \in \Gamma} p\big(y_\gamma | y_{\mathrm{pa}(\gamma)}, i_{\mathrm{pa}(\gamma)}\big).$$

A method for estimating the parameters and learning the dependency structure of a conditional Gaussian networks with mixed variables is presented in Bøttcher (2001) and implemented in the software package `deal`, see Bøttcher and Dethlefsen (2003).

3.2 Parameter and Structure Learning

To estimate the parameters in the network and to find the structure of the network, we use a Bayesian approach. So, considering the parameters, we encode our uncertainty about θ in a prior distribution $p(\theta)$, use data d to update this distribution, *i.e.* learn the parameters, and hereby obtain the posterior distribution $p(\theta|d)$ by using Bayes' theorem,

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}, \qquad \theta \in \Theta. \tag{2}$$

Here, Θ is the parameter space, d is a random sample of size n from the probability distribution $p(x|\theta)$ and $p(d|\theta)$ is the joint probability distribution of d, also called the likelihood of θ. As prior parameter distributions we use

the Dirichlet distribution for the discrete variables and the Gaussian inverse-Gamma distribution for the continuous variables. These distributions are conjugate to observations from the respective distributions and this ensures simple calculations of the posterior distributions.

Now, to learn the structure of the network, we calculate the posterior probability of the DAG, $p(D|d)$, which from Bayes' theorem is given by

$$p(D|d) = \frac{p(d|D)p(D)}{p(d)},$$

where $p(d|D)$ is the likelihood of D and $p(D)$ is the prior probability of D. As the normalizing constant $p(d)$ does not depend upon structure, another measure, which gives the relative probability, is

$$p(D, d) = p(d|D)p(D).$$

We use the above measure and refer to it as the *network score*. For simplicity, we choose to let $p(D)$ be the same for all DAGs, so we are only interested in calculating the likelihood $p(d|D)$. It is given as

$$p(d|D) = \int_{\theta \in \Theta} p(d|\theta, D)p(\theta|D)d\theta,$$

and we see that it, besides the likelihood of the parameters, also involves the prior distribution over the parameters, $p(\theta|D)$. This means that we for each possible DAG have to specify a prior distribution for the parameters. In the papers Heckerman, Geiger and Chickering (1995) and Geiger and Heckerman (1994) an automated method for doing this in respectively the purely discrete and the purely Gaussian case is developed. In Bøttcher (2001) this method is extended to the mixed case. With this method, the parameter priors for all possible networks can be deduced from one joint parameter prior, called a master prior. To specify this master prior, we only have to specify a prior Bayesian network, *i.e.* a prior DAG and a prior probability distribution, together with a measure of how confident we are in the prior network. With a few assumptions, the network score is obtainable in closed form.

If many DAGs are possible, it is computational infeasible to calculate the network score for all DAGs. In this situation it is necessary to use some kind of search strategy to find the DAG with the highest score, see e.g. Cooper and Herskovits (1992). In this paper we use a search strategy called greedy search. In greedy search we compare DAGs that differ only by a single arrow, either added, removed or reversed. The change that increases the network score the most is selected and the search is continued from this new DAG.

4 Inference

Having established a Bayesian network for a set of random variables, this represents the knowledge we, at this stage, have about these variables. When information on some or all of the variables becomes available, we can use this "knowledge base" to make inference about the unobserved variables in the network.

Inference in Bayesian networks is performed using Bayes' theorem. Consider a network for a set of random variables X and assume that some of the variables, B, are observed and the rest, A, are not. We can then, by using Bayes' theorem, calculate the conditional distribution of A given B as

$$p(A|B) \propto p(B|A)p(A).$$

Thus $p(A)$ is the prior distribution of A, *i.e.* the distribution of A before we observe B, $p(B|A)$ is the likelihood of A and $p(A|B)$ is the posterior distribution of A, *i.e.* the distribution of A, when we have observed B. Generally, finding these distributions is computationally demanding as it involves calculating huge joint distributions, especially if there are many variables in the network. Therefore efficient methods of implementing Bayes' theorem are being used. These implementations uses the fact that the the joint probability distribution of all the variables in a network factorizes according to (1). The marginal or conditional distributions of interest can then be found by a series of local computations, involving only some of the variables at a time, see *e.g.* Cowell, Dawid, Lauritzen and Spiegelhalter (1999) for a thorough treatment of these methods.

So having observed some of the variables in a network, we can use this new evidence to calculate the posterior distribution of any unobserved variable X_v, given the evidence. Notice that we do not need to observe all the other variables before calculating the posterior distribution, as we can update the prior distribution of X_v with any information available. Of course, the more information we have, the better the posterior distribution is determined. However, not all information will have an impact on the posterior distribution of a variable X_v. Consider the following result. A node v is conditional independent on the rest of the nodes in the network, given the *Markov blanket* of v, $\text{bl}(v)$, *i.e.*

$$v \perp\!\!\!\perp V \backslash v | \text{bl}(v).$$

The Markov blanket of v is the set of v's parents, children and children's parents, *i.e.*

$$\text{bl}(v) = \text{pa}(v) \cup \text{ch}(v) \cup \{w : \text{ch}(w) \cap \text{ch}(v) \neq \emptyset\},$$

where pa(v) is the parents of v and ch(v) is the children of v, see Cowell et al. (1999). So if all the variables in the Markov blanket are observed, we do not get further information about the distribution of X_v by observing the variables outside the Markov blanket. But if we have *not* observed all the variables in the Markov blanket, then observing some variable outside the Markov blanket, can influence the posterior distribution of X_v.

5 Results

We will now present the results obtained.

5.1 Preliminaries

In the present study, 187 subjects without known diabetes underwent both an OGTT and an IVGTT. In the OGTT, measurements were recorded of plasma glucose (G) and serum insulin levels (I) at time points 10, 5 and 0 minutes before intake of 75 gram glucose and at 10, 20, 30, 40, 50, 60, 75, 90, 105, 120, 140, 160, 180, 210 and 240 minutes after the intake.

In this analysis, the observations to time 10, 5 and 0 minutes before the glucose intake are, for both insulin and glucose, averaged and represented by the corresponding observation to time 0. Further, based on previous results, see Drivsholm et al. (2003), we use the logarithm of the insulin sensitivity index $\log S_I$ instead of S_I and we also include the sex of the patient and the body mass index (BMI) in the models. Sex is a binary variable, but we choose to treat it as a continuous variable. This has the effect that the variance is assumed equal for male and female observations, whereas the means can differ. If sex is treated as a discrete variable, the data is split into two groups with a parameter set for each group. and we have found that we do not have enough data to support this. Consider for example the simple case, where the only parent to $\log S_I$ is sex. If sex is treated as a continuous variable, the distribution of $\log S_I$ is given as

$$(\log S_I|\text{sex}) \sim \mathcal{N}(m + \beta\text{sex}, \sigma^2).$$

None of the parameters m, β and σ^2 depend on sex, but the mean is m if sex is 0 and $m + \beta$ if sex is 1. If sex is treated as a discrete variable, the distribution of $\log S_I$ is

$$(\log S_I|\text{sex}) \sim \mathcal{N}(m_{\text{sex}}, \sigma^2_{\text{sex}}),$$

i.e. both the mean and the variance depends on sex.

In the following we will try different ways of establishing a Bayesian network, which can be used to predict $\log S_I$ from measurements from an OGTT and from BMI and sex. So the networks we will consider in the following, only contain continuous variables. Notice that, when using the theory presented for mixed networks on networks with only continuous variables, it coincides with theory developed for purely continuous networks, see Bøttcher (2001). To learn the parameters and structure of a Bayesian network, we use the software package `deal`, see Bøttcher and Dethlefsen (2003). The package is written for R, see R Development Core Team (2004).

To validate the models, we split the dataset into a subset with 140 subjects, used as training data, and a subset with 47 subjects, used as validation data. For each model, we use `deal` with the training data to learn the parameters and structure of the Bayesian network. The posterior parameter distribution of $\log S_I$ is used to derive point estimates of the parameters. For the Gaussian parameters, we use the mean of the posterior and for the gamma distributed parameter, we use the mode of the posterior. These point estimates are then transfered to Hugin (`www.hugin.com`). For each subject in both the training data and the validation data, the conditional distribution of $\log S_I$ is calculated given the observations from the OGTT using Hugin. In the following, we call this distribution the predictive distribution of $\log S_I$. Notice, however, that if a fully Bayesian approach had been used, the predictive distribution for one subject is

$$p(\log S_I|d) = \int_{\theta\in\Theta} p(\log S_I|d,\theta)p(\theta)d\theta,$$

where d denotes the subjects OGTT measurements and θ are the parameters. This distribution is a t distribution with degrees of freedom increasing with, among other numbers, the number of subjects in the training dataset. In this study we have 140 subjects and we find that the error using a Gaussian distribution instead, is very small.

The predictive distribution is then, for each subject, compared with the corresponding $\log S_I$ value determined from the IVGTT in the following way. For each subject we use the predictive distribution to calculate the 95%'s credibility intervals $\mu \pm 1.96 \cdot \sigma$, where 1.96 is the 97.5%'s quantile in the Gaussian distribution. So if a Bayesian network can predict the value of $\log S_I$, we expect that 95% of the corresponding $\log S_I$ values found in the IVGTT study, will lie within this interval. If this is the case, we say that the predictive distribution of $\log S_I$ is *well calibrated*, see Dawid (1982).

Further, we perform an ordinary linear regression of the IVGTT obtained S_I on the predicted S_I and calculate the residual standard deviation, SD, and the correlation coefficient, R^2, obtained from this regression. To show that there is no systematic bias in these regressions, we report the intercept and slope of these regressions lines.

5.2 The Different Models

In the following we will present different approaches for finding a Bayesian network, that can model the dependency relations between the variables in the problem. Further, we will present the results of a previous approach, where multiple linear regression is used and also the results of using the leaps and bounds algorithm for best subset selection.

Bayesian regression network

Previous results have shown that predictions of $\log S_I$ from a multiple regression on OGTT plasma glucose and serum insulin levels, BMI and sex, are highly correlated to the corresponding IVGTT-derived S_I estimates, see *e.g.* Drivsholm et al. (2003). We will therefore learn the parameters and the structure of a network, where $\log S_I$ can depend on these variables, and these variables are marginally independent, *i.e.* the only arrows that are allowed in the model, are arrows into $\log S_I$. This network represents a regression model, so we will refer to it as the Bayesian regression network.

To learn this network, we need to specify a prior network, *i.e.* a prior DAG and a prior probability distribution. As prior DAG we use, for simplicity, the empty DAG, *i.e.* the one without any arrows. This DAG represents that all the variables are independent, so the local probability distribution for each node only depends on the node itself. To specify the prior probability distribution, we use the sample mean and the sample variance as an initial estimate of the mean and variance. As a measure of our confidence in this network, we use $N = 100$ for the size of the imaginary database.

Figure 1 shows the result of the structural learning procedure. We see that $\log S_I$ depends on almost all of the insulin measurements, except for I10, and a few of the glucose measurements.

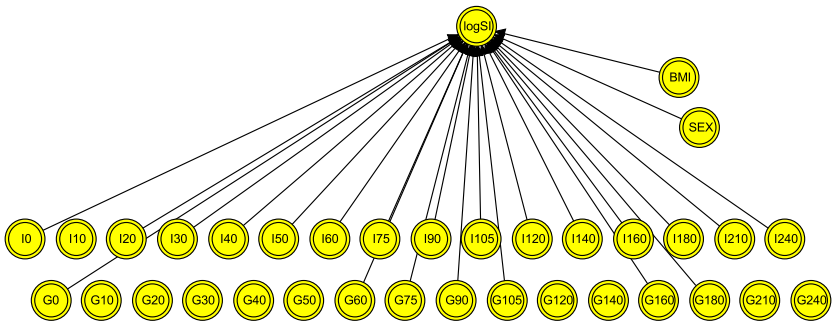


Figure 1: The result of the structural learning procedure for the Bayesian regression network.

Bayesian network with empty prior network

In situations where not all the variables are observed, information is gained by modeling the possible correlations between the explanatory variables. So we will now learn a network, where these correlations are allowed. We restrict us to networks, where arrows between the glucose and insulin measurements point forward in time, where BMI and sex can not have any parents and where $\log S_I$ can not have any children. Again we use the empty DAG as prior DAG, the sample mean and sample variance to specify the prior probability distribution and $N = 100$ as a measure of our confidence in this network. The result of the structural learning procedure reveals a complicated dependency structure between the variables, see Figure 2.

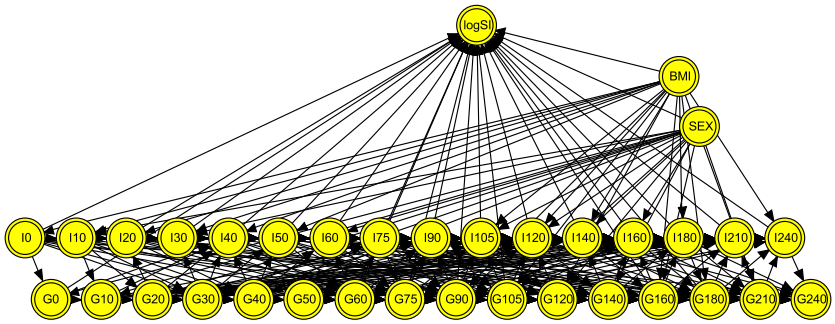


Figure 2: The Bayesian network with the empty network as prior.

The Markov blanket for $\log S_I$ in this network, is the same as the Bayesian regression network, see Figure 1. The reason for this is that $\log S_I$, in both networks, is not allowed to have any children and because we in both approaches have used the same prior network. So when all the variables in the Markov blanket is observed, as it is in our study, the prediction results are

exactly the same as for the Bayesian regression network.

Bayesian network with physiological prior network

In the previous two networks, we have for simplicity used the empty DAG as prior DAG. We will now use a prior DAG, called the physiological network, where the knowledge we have about the physiological relations between the variables is incorporated.

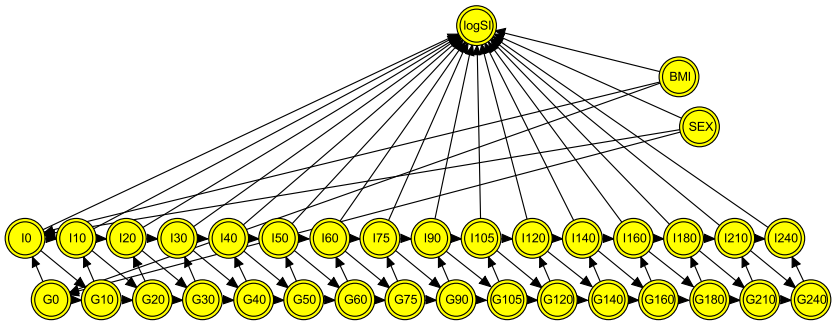


Figure 3: The physiological network.

In this network, insulin measurements and glucose measurements are assumed to be Markov processes. They are coupled so that the current glucose measurement depends on the previous insulin measurement and the current insulin measurement depends on the current glucose measurement, see Figure 3. This structure is consistent with the physiological model used in Bergmans minimal model to determine S_I from an IVGTT. In addition, we let the initial glucose and insulin measurements depend on BMI and sex.

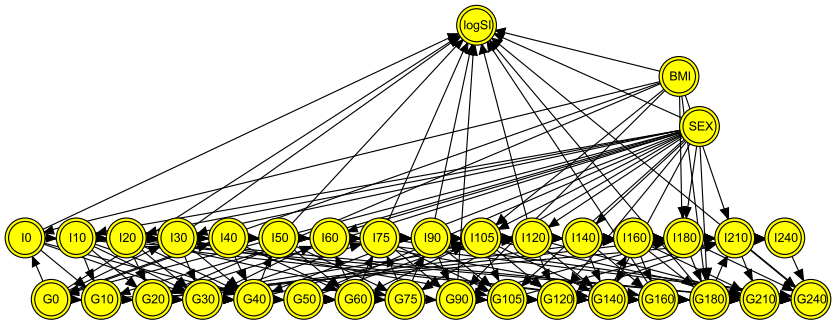


Figure 4: The Bayesian network with the physiological network as prior.

Like before, we estimate the prior probability distribution from data. However, contrary to the empty network, the variables in the physiological network depends on other variables, so we perform a linear regression on the

parents and use the sample mean and sample variance from these regressions as the mean and variance in the local prior probability distributions. Again we use $N = 100$ and we only consider networks where arrows between the glucose and insulin measurements point forward in time, where BMI and sex can not have any parents and where $\log S_I$ can not have any children.

The result of the structural learning procedure is shown in Figure 4. As before, we see a complicated dependency structure between the variables.

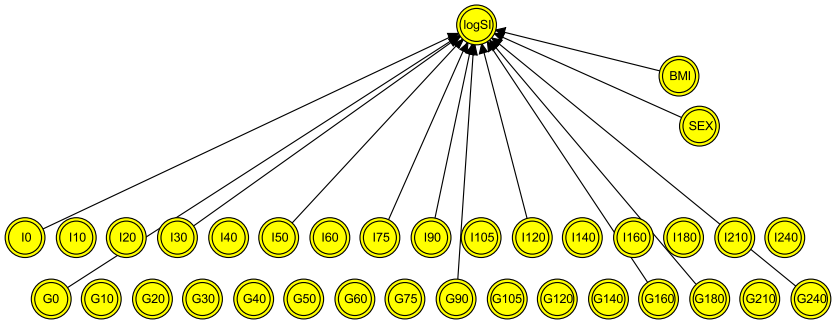


Figure 5: The Markov blanket for the Bayesian network with the physiological network as prior.

In Figure 5, the Markov blanket of $\log S_I$ is shown and we see that is quite different than with the empty prior, shown in Figure 1. Only 6 of the insulin measurements and 5 of the glucose measurements are included in the present blanket.

Results using multiple linear regression

In Drivsholm et al. (2003), multiple linear regression is used to derive predictive equations of $\log S_I$ using OGTT plasma glucose and serum insulin levels, BMI and gender. To limit the amount of blood samples drawn from the patients, they constrain the models to include glucose and insulin observations to the same time point. By a combination of backwards elimination and forward selection, they find the optimal model to be with sample time points 0, 30, 60, 105, 180, and 240. Notice, though, that they have found their model on the basis of a different training dataset than ours, as the partition of the dataset into training data and validation data is done randomly in both cases.

Results using the leaps and bound algorithm

Further, we have tried the leaps and bound algorithm by Furnival and Wilson (1974), using the Bayesian information criteria to find the best subset of the explanatory variables. With this approach, the optimal model is with I50, I90, G160, BMI and sex as explanatory variables. In theory, when the size of the database approaches infinity, using the Bayesian information criteria will result in the same subset of explanatory variables as when using the network score as selection criteria, see Haughton (1988).

5.3 Evaluation

To compare the different models, we first consider the network score. Notice that we can only compare network scores for networks that are learned using the same prior network. To be able to evaluate all models using the network score, we have also calculated the log scores for the results found in the multiple linear regression approach and the leaps and bounds approach. This is done by formulating these results as Bayesian networks and calculating the log scores using respectively the empty network and the physiological network as prior network. Likewise, for the Bayesian regression network found by using the empty network as prior network, we have calculated the log score using the physiological network as prior network.

Model	Empty prior	Physiological prior
BR	−17878.30	−17848.33
BN	−16528.39	−14851.44
MLR	−17886.17	−17849.06
L&B	−17894.95	−17846.12

Table 1: Network scores for the different models.

The results are reported in Table 1. The Bayesian network model (BN) has the lowest log score, *i.e.* the highest network score, both when the empty network and the physiological network are used as prior network. This is obvious as the BN is selected using the network score as selection criteria and because the Bayesian regression (BR), the multiple linear regression (MLR) and the leaps and bounds (L&B) networks are included in the search space, when searching for the BN with the highest score. So unless we have only found a local maximum, instead of a global maximum, the score for the BN must be higher than the score for the other networks.

When comparing the scores found using the empty prior, we see that the network scores for the BR network, the MLR network and the L&B network are almost all the same. The network score for the BN is over a thousand times higher than for any of the other networks, indicating that the BN provides a much better fit to data. Recall, however, that the Markov blanket for the BR network and the BN are the same, so when all the variables in the Markov blanket are observed, the BR network and the BN will predict the same $\log S_I$ values. So the higher network score is not important when data are complete, but can have an impact when data are incomplete.

When using the physiological network as prior network, we see almost the same result. The network score for the BR, MLR and L&B networks are almost all the same, whereas the network score for the BN is over 3000 times higher than for any of the other networks.

Model	Tr. data $R^2(SD)$	Val. data $R^2(SD)$	Tr. outside	Val. outside
BR with empty prior	0.76(0.31)	0.73(0.35)	1(1%)	1(2%)
BN with empty prior	0.76(0.31)	0.73(0.35)	1(1%)	1(2%)
BN with phys. prior	0.77(0.30)	0.73(0.36)	7(5%)	3(6%)
MLR	0.76(0.31)	0.66(0.40)	3(2%)	3(6%)
L&B	0.75(0.31)	0.73(0.36)	6(4%)	4(9%)

Table 2: The table lists the R^2 and SD values from the linear regressions of the IVGTT obtained $\log S_I$ on the predicted $\log S_I$ for both the training dataset and the validation dataset. Also listed are how many $\log S_I$ values that fall outside the credibility interval $\mu \pm 1.96 \cdot \sigma$.

In Table 2 the R^2 and SD values from the linear regression of the IVGTT obtained $\log S_I$ on the predicted $\log S_I$ are reported. The R^2 and SD values are for all five models acceptable and they are almost the same for all models, except for the multiple regression model, which on the validation dataset does not perform as well as the others.

Table 3 shows the intercept and slope of the estimated regression lines and there are no evidence of any systematic bias. We therefore conclude that an OGTT can be used to determine the insulin sensitivity index.

In Table 2 we have also listed how many $\log S_I$ values that fall outside the credibility interval $\mu \pm 1.96 \cdot \sigma$. Approximatively 5% of these predictions should lie outside and 95% inside the interval for the predictive distributions to be well calibrated. This is clearly fulfilled for the BN with the physiological network as prior network, so the predictive distribution for $\log S_I$ is, when using this network, well calibrated. With the MLR approach and the

Model	Tr. data (intercept, slope)	Val. data (intercept, slope)
BR empty prior	$(-0.19, 1.09)$	$(-0.05, 1.01)$
BN empty prior	$(-0.03, 1.01)$	$(0.25, 0.87)$
BN phsy. prior	$(-0.06, 1.03)$	$(0.14, 0.92)$
MLR	$(0, 1)$	$(0.06, 0.96)$
L&B	$(0, 1)$	$(0.11, 0.93)$

Table 3: The intercept and slope of the regressions lines from the regressions of the IVGTT obtained S_I on the predicted S_I. Reported to show that there is no evidence of systematic bias.

L&B approach it is almost fulfilled that 5% of the predictions lie outside the intervals. We will therefore conclude that the predictive distributions are also well calibrated in these cases. For the BR and the BN with the empty network as prior network, very few values lie outside the intervals, indicating that the variance is probably estimated to large.

Figure 6 shows the predicted $\log S_I$ values and the intervals for the BN with the empty prior and for the BN with the physiological prior. We see that for the two models, the predicted $\log S_I$ values are almost the same, but the intervals are much wider for the BN with the empty prior, meaning that the variance in this model is larger.

So to summarize, all the models give adequate predictions of the $\log S_I$ values. Evaluating the models using the different validation approaches all together, the BN with the physiological prior model gives a more precise predictive distribution of $\log S_I$ compared to the other models. We therefore suggest that this model should be used to derive the predictive values of $\log S_I$.

6 Discussion

We have established a promising way of determining the insulin sensitivity index from an oral glucose tolerance test rather than from an intravenous glucose tolerance test. All approaches give adequate predictions of S_I. The Bayesian network with the physiological prior estimates the most precise predictive distribution of S_I, so we claim that this is the best model. There are also other advantages by using a Bayesian network instead of an ordinary regression model. In a Bayesian network, we can use any prior knowledge available from *e.g.* previous studies or from the physiological understanding of the problem. Further, we can calculate the predictive distribution of $\log S_I$

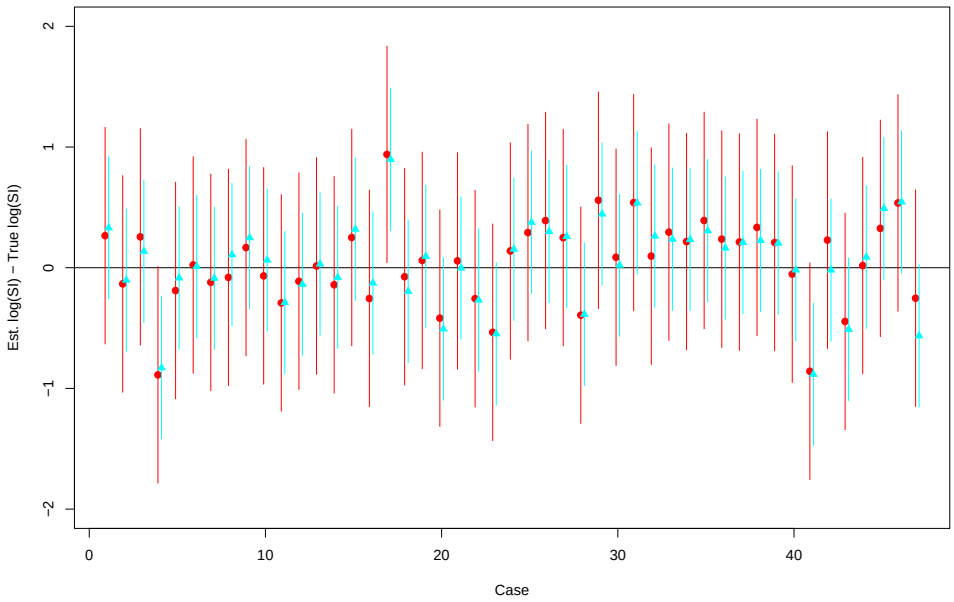


Figure 6: The predicted $\log S_I$ values and the credibility intervals for the Bayesian network with empty prior (dark and disks) versus the Bayesian network with physiological prior (light and triangles).

in situations, where some of the observations are missing. This can be used when a single or a few observations are missing for a specific subject. It can also be used when certain time points are not observed at all, which could be the case if a dataset from another study, using fewer time points, is analyzed.

Acknowledgements

This work has been supported by Novo Nordisk A/S. The collection of data were done by Torben Hansen, Novo Nordisk A/S.

References

Bergman, R. N., Ider, Y. Z., Bowden, C. R. and Cobelli, C. (1979). Quantitative estimation of insulin sensitivity, *American Journal of Physiology* **236**: E667–E677.

Bøttcher, S. G. (2001). Learning Bayesian Networks with Mixed Variables, *Artificial Intelligence and Statistics 2001*, Morgan Kaufmann, San Francisco, CA, USA, pp. 149–156.

Bøttcher, S. G. and Dethlefsen, C. (2003). `deal`: A Package for Learning Bayesian Networks, *Journal of Statistical Software* **8**(20): 1–40.

Cooper, G. and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data, *Machine Learning* **9**: 309–347.

Cowell, R. G., Dawid, A. P., Lauritzen, S. L. and Spiegelhalter, D. J. (1999). *Probabilistic Networks and Expert Systems*, Springer-Verlag, Berlin-Heidelberg-New York.

Dawid, A. P. (1982). The Well-Calibrated Bayesian, *Journal of the American Statistical Association* **77**(379): 605–610.

Drivsholm, T., Hansen, T., Urhammer, S. A., Palacios, R. T., Vølund, A., Borch-Johnsen, K. and Pedersen, O. B. (2003). Assessment of insulin sensitivity index and acute insulin reponse from an oral glucose tolerance test in subjects with normal glucose tolerance, Novo Nordisk A/S.

Furnival, G. M. and Wilson, R. W. (1974). Regression by Leaps and Bounds, *Technometrics* **16**(4): 499–511.

Geiger, D. and Heckerman, D. (1994). Learning Gaussian Networks, *Proceedings of Tenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, CA, USA, pp. 235–243.

Haughton, D. M. A. (1988). On The Choice of a Model to fit Data From an Exponential Family, *The Annals of Statistics* **16**(1): 342–355.

Heckerman, D., Geiger, D. and Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data, *Machine Learning* **20**: 197–243.

Lauritzen, S. L. (1992). Propagation of probabilities, means and variances in mixed graphical association models, *Journal of the American Statistical Association* **87**(420): 1098–1108.

Martin, B. C., Warram, J. H., Krolewski, A. S., Bergman, R. N., Soeldner, J. S. and Kahn, C. R. (1992). Role of glucose and insulin resistance in development of type 2 diabetes mellitus: results of a 25-year follow-up study, *The Lancet* **340**: 925–929.

Pacini, G. and Bergman, R. N. (1986). MINMOD: a computer program to calculate insulin sensitivity and pancreatic responsivity from the frequently sampled intraveneous glucose tolerance test, *Computer Methods and Programs in Biomedicine* **23**: 113–122.

R Development Core Team (2004). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.