Stochastic latent residual approach for consistency model assessment

Stochastic Latent Residual Approach for Consistency Model Assessment

Hani Syahida Zulkafli[1,*], George Streftaris[2], Gavin J. Gibson[2]

[1]Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, Selangor, Malaysia
[2]School of Mathematics and Computer Sciences, Heriot-Watt University, United Kingdom

Abstract Hypoglycaemia is a condition when blood sugar levels in body are too low. This condition is usually a side effect of insulin treatment in diabetic patients. Symptoms of hypoglycaemia vary not only between individuals but also within individuals making it difficult for the patients to recognize their hypoglycaemia episodes. Given this condition, and because the symptoms are not exclusive to only hypoglycaemia, it is very important for patients to be able to identify that they are having a hypoglycaemia episode. Consistency models are statistical models that quantify the consistency of individual symptoms reported during hypoglycaemia. Because there are variations of consistency model, it is important to identify which model best fits the data. The aim of this paper is to asses and verify the models. We developed an assessment method based on stochastic latent residuals and performed posterior predictive checking as the model verification. It was found that a grouped symptom consistency model with multiplicative form of symptom propensity and episode intensity threshold fits the data better and has more reliable predictive ability as compared to other models. This model can be used in assisting patients and medical practitioners to quantify patients' reporting symptoms capability, hence promote awareness of their hypoglycaemia episodes so that corrective actions can be quickly taken.

Keywords Latent Residual, Posterior Predictive Checking, Model Verification, Model Assessment

1. Introduction

Hypoglycaemia is a condition of low glucose level in blood, i.e. below 4mmol/L. It is a common side effect of insulin treatment in diabetic patients. It is crucial to treat a hypoglycaemia episode promptly to avoid severe hypoglycaemia episode, where patient needs other people's help to recover. However, it is not easy for the patient to identify a hypoglycaemia episode because symptoms of hypoglycaemia vary within individuals. A given symptom is not equally covarying with blood glucose levels [1] implying a degree of between-subject variability. Individuals experiencing various symptoms of hypoglycaemia are not necessarily able to recognize a hypoglycaemic episode because the individuals' ability to recognize hypoglycaemia is significantly correlated with the number of symptoms reported per episode [2]. There are marked variability of the reported symptoms between episodes of hypoglycaemia [3] but the study is limited to children respondents. A consistency model was developed to quantify the consistency of reporting the symptoms of hypoglycaemia by adult patients [4]. Zulkafli *et al* [5] then introduced the grouped symptoms models as one of the consistency estimations models. This model adds another source of variation to symptoms' reportings by distributing the 26 symptoms to several groups according to the causes. Other functional form was briefly introduced as an alternative to be used in the consistency models [5].

With several consistency models developed, the challenge is to evaluate the performance of each model before making decisions on which model can give better consistency estimates.

Residual analysis is one prominent way in validating a statistical model. Cox and Snell, [3] introduced a general definition of residual for non-linear models. Deviance, Pearson and Anscombe residuals are examples of type of residuals commonly used in residual analysis. However, these residuals have unknown sampling distribution which will affect the interpretation of the analysis [7].

Among works that has been done in measuring the performance of statistical models related to diabetic data use coefficient of determination (R^2) goodness-of-fit measure [8] and robust method [9]. However, the work does not apply to the concept of latent residuals.

Latent residual analysis was used in analyzing binary response variable in regression framework [10,11]. χ^2 test for latent model testing is sensitive to distributional properties of the observed variables [12]. The test also will have high probability of Type 1 error with complex model [13]. Therefore, the intent of this paper is to present a method for assessing the adequacy of the stochastic model with latent variables utilissing the concept of stochastic latent residuals, z_{ijk}.

Also, one of the important aims of this work is to develop a model which can be used to make prediction of values of interest with quantified confidence. A good predictive model, enables us to predict how consistent a patient is in reporting hypoglycaemia when given some of his/her specific characteristics. This can be used in order to assist early detection of hypoglycaemia and give necessary advice to the patient. Therefore, the second objective of this paper is to examine the consistency model's predictive capability by employing a validation approach relying on the posterior predictive distribution.

2. Materials and Methods

The methods of model assessment discussed in this paper are applied to data collected from 66 diabetic patients where each subject is given a unique ID number [14]. Each patient recorded his/her symptoms in each hypoglycaemia episode experienced for a duration of 9 to 12 months.

2.1. The Consistency Model

A consistency model was developed under Bayesian approach [4]. Observed variable Y_{ijk} takes value 1 if patient $i = 1,\dots,$ I reports symptom $j = 1,\dots,$ J in episode $k = 1,\dots,$ K$_i$ by patient $i = 1,\dots,$ I. Otherwise, Y_{ijk} takes value 0. $Y_{ijk} \sim$ Bernoulli (p_{ijk}) where p_{ijk} is the probability of symptom j is reported in episode k by patient i. A threshold, τ_{ijk} is defined for patient i reporting symptoms j at episode k and symptom j is considered as reported when the threshold τ_{ijk} exceeded by a functional form $h(\alpha_{ij}, \beta_{ik}.)$, i.e $\tau_{ijk} \leq h(\alpha_{ij}, \beta_{ik}.)$. α_{ij} and β_{ik} are latent variables which correspond to the propensity of symptoms j for patient i and intensity of

episodes k in patient i respectively.

The threshold, τ_{ijk} is assumed to follow a log-normal distribution, $\tau_{ijk} \sim$ Log-Normal$(0, \sigma_i^2)$ where σ_i^2 is the parameter associated with the variability of symptoms reported by individual patient i. The consistency estimate is defined as $c_i = \frac{1}{100+\sigma_i^2}$.

Each of the parameter is assigned a prior distribution as follows;

$$\alpha_{ij} \sim \text{Gamma}(1,0.1), i = 1,\dots,66 \text{ and } j = 1,\dots,26$$

$$\beta_{ik} \sim \text{Gamma}(1,0.1), i = 1,\dots,66 \text{ and } k = 1,\dots,K_i$$

$$\sigma_i^2 \sim \text{Inv-Gamma}(1,0.1), i = 1,\dots,66.$$

This consistency model was later expanded by separating the symptoms into different groups according to their causes in order to have an additional source of variation [5]. Therefore, the prior corresponding to the symptoms propensity then become

$$\alpha_{ijl} \sim \text{Gamma}\left(\theta, \frac{\theta}{u_l}\right), l = 1,\dots,6$$

giving $\text{E}(\alpha_{ijl}) = u_l$ and $\text{Var}(\alpha_{ijl}) = u_l^2/\theta, l = 1,\dots,6$

Earlier work of the consistency model assumed a threshold form $h(\alpha_{ij}, \beta_{ik}.) = \alpha_{ij}\beta_{ik}$ [4]. Later, another option for the functional form was introduced, i.e. $h(\alpha_{ij}, \beta_{ik}.) = \alpha_{ij} + \beta_{ik}$ and their differeces were briefly discussed [5].

2.2. Stochastic Latent Residual

The stochastic latent residuals, z_{ijk}, would give rise to the observed data under the considered model. Following the concept of generalised residuals [6], the data can be regarded as generated through a functional model, $g(\cdot)$, [15] depending on the vector of all model parameters and latent variables, say θ, i.e.

$$y = g_\theta(z) \tag{1}$$

where $z \sim U(0,1)$ are generalised residuals. Then, in the general case, (1) can be inverted to give the stochastic latent residuals

$$z = g_\theta^{-1}(y) \tag{2}$$

For the assumed discrete model we have

$$y_{ijk} = I\{z_{ijk} \leq p_{ijk}\}$$

where $z_{ijk} \sim U(0,1)$ and $I\{\cdot\}$ is the indicator function. This implies that, under the assumed model,

$$z_{ijk} = y_{ijk}u_1 + (1 - y_{ijk})u_2$$

where $u_1 \sim U(0, p_{ijk})$ and $u_2 \sim U(p_{ijk}, 1)$. Therefore, if the model is adequate, $z_{ijk} \sim U(0,1)$ and a p-value for testing the hypothesis of this uniform distribution can be obtained. To implement this method, 10,000 MCMC iterations were run for this model and obtained the latent residual, z_{ijk}, for each subject such that

$$\text{If}\begin{cases} y_{ijk} = 1, & z_{ijk}^{(1)} \sim U(0, \hat{p}_{ijk}) \\ y_{ijk} = 0, & z_{ijk}^{(0)} \sim U(\hat{p}_{ijk}, 1) \end{cases}$$

where $\hat{p}_{ijk}$ is the estimated probability of patient i reporting symptom j at episode k at each iteration. Therefore, if the tested hypothesis is correct,

$$z_{ijk} = (z_{ijk}^{(1)}, z_{ijk}^{(0)}) \sim U(0,1).$$

A Kolmogorov-Smirnov goodness-of-fit test was conducted on each posterior sample of residuals obtained in each MCMC iteration, resulting in a corresponding p-value, π_γ, where $\gamma = 1, 2, 3, \ldots, 10,000$ iterations. This will give a posterior distribution $f(\pi|y_{ijk})$ where y_{ijk} denotes the observation data.

2.3. Posterior Predictive Checking

This approach is commonly used for checking the model's suitability, and is based on work that was elaborated in [16] and later expanded in [17]. The purpose of the analysis is to compare the observed data with values predicted from the model.

The observations, Y_{ijk} are binary data that take value 1 if patient i reported symptom j in episode k value zero otherwise. $Y_{ijk}^{(p)}$ is defined as the predicted data, such that these are the data that will be obtained if we use the same model to do prediction. 10% of the total number of observations are randomly selected, which are then used as the validation sample. Then, the examined model is fitted to the remaining data. The fitted model is subsequently used to do prediction on symptom reporting for the sampled patients episodes, $Y_{ijk}^{(p)}$ for i, j, k in the sample. Recall that, Y_{ijk} is Bernoulli distributed with probability, p_{ijk}. The posterior probability, p_{ijk}, is sampled from the fitted model and is used to obtain the reporting prediction, $Y_{ijk}^{(p)}$. Consequently, we compare the total number of predicted reportings, N_p to the total number of observed reportings, N_{obs}. Accordingly, the distributions of Y_{ijk} and $Y_{ijk}^{(p)}$ were compared.

Four other measures are used to assess and describe the usefulness of the model's predictions [18]. The measures are related to sensitivity, specificity and predictive values. Here, sensitivity is defined as the proportion of experienced symptoms that are correctly predicted as being reported by the models whereas specificity is the proportion of symptoms that have not been experienced which are correctly predicted as not reported by the model. Ideally, a good predictive model should have high sensitivity and specificity. However, these two measures are often inversely proportional, meaning as sensitivity increases specificity decreases and vice versa. The probability of the model giving correct prediction were evaluated by using

the positive predictive value (PPV) and negative predictive value (NPV). These four measures were calculated using Y_{ijk} and $Y_{ijk}^{(p)}$ in the validation sample and are defined as follows

a)   PPV=$\frac{\sum_{ijk} Y_{ijk} Y_{ijk}^{(p)}}{\sum_{ijk} Y_{ijk}^{(p)}}$ for i, j, k in the sample

PPV is the proportion of symptoms with positive prediction that was correctly classified as reported. PPV measures the probability of patient i truly experiencing symptom j at episode k given that the model predicts the symptom is likely to be experienced.

b)   NPV=$\frac{\sum_{ijk}(1-Y_{ijk})(1-Y_{ijk}^{(p)})}{\sum_{ijk}(1-Y_{ijk}^{(p)})}$ for i, j, k in the sample.

NPV is the proportion of symptoms with negative reporting prediction that was correctly classified as absent. NPV measures the chance of patient i having symptom j not present at episode k given that the model predicts that it is not likely to be reported.

c)   TPR=$\frac{\sum_{ijk} Y_{ijk} Y_{ijk}^{(p)}}{\sum_{ijk} Y_{ijk}}$ for i, j, k in the sample.

True Positive Rate (TPR), also known as the sensitivity of the predictive model, measures the ability of the model to correctly predict if symptom j occurs at episode k.

d)   TNR=$\frac{\sum_{ijk}(1-Y_{ijk})(1-Y_{ijk}^{(p)})}{\sum_{ijk}(1-Y_{ijk})}$ for i, j, k in the sample.

True Negative Rate (TNR), or also called specificity, represents the capacity of the model to predict that symptom j is not reported at episode k when the symptom is truly absent.

3. Results and Discussion

3.1. Model Assessment

As preliminary checking, we observe the histogram of the residuals $\mathbf{z}$ for each patient i for grouped symptom model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. Recall that patient i reports symptom j at episode k when $\tau_{ijk} \leq \alpha_{ij}\beta_{ik}$. Thus, the observed variable, Y_{ijk} is equal to 1 when symptom j is reported at episode k by patient i. Otherwise, Y_{ijk} takes value zero. Figure 1 presents the histogram for one patient, Subject 4028. The distribution pattern suggests that the residuals do follow a Uniform (0,1) distribution. To further confirm the distribution of $\mathbf{z}$ we also check on the histogram of p-values for this patient, $\pi_{(4028)}$ (Figure 2). From this histogram, we can say there is no evidence against the adequacy of fit of the model.
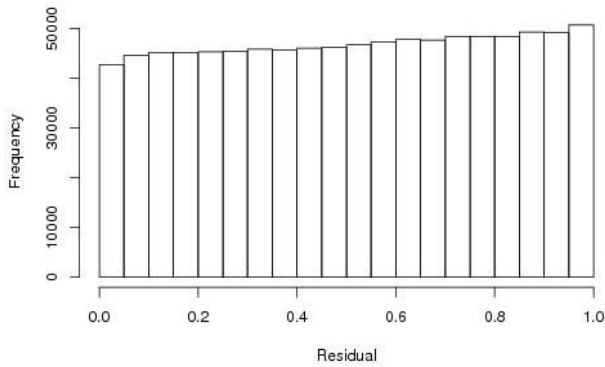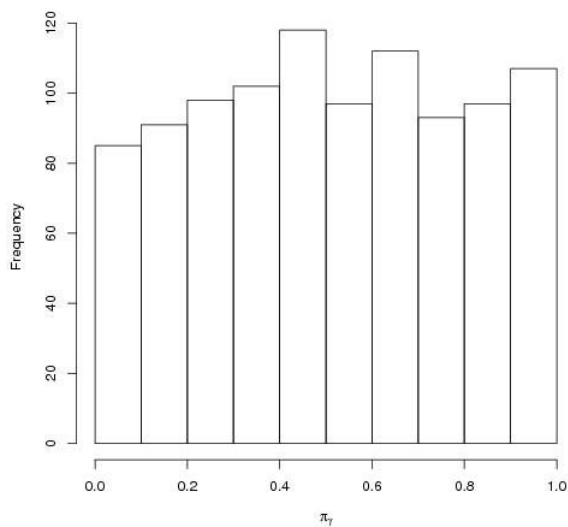
Figure 1. Histogram of stochastic latent residuals for model with grouped symptoms using threshold $(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ for patient 4028.
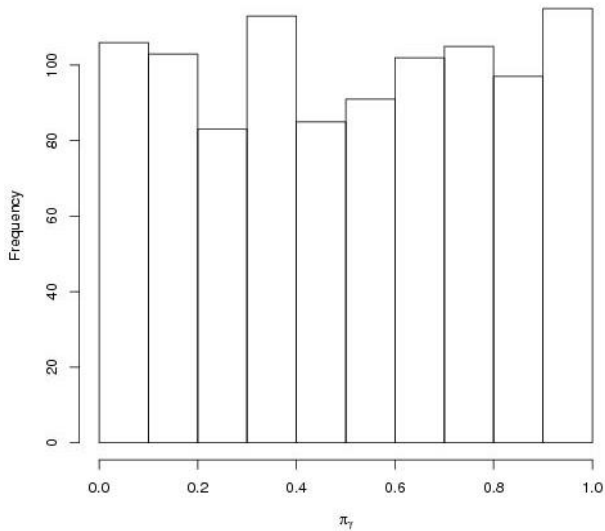
Figure 2. Posterior distribution of p-values, π, for fit of model with grouped symptoms using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ for patient 4028

For comparison purposes, Figure 3 presents histograms

of p-values, π_γ, representing another patient, Subject 5088 when using the two thresholds, $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ and $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik.}$ Observing the posterior distributions of π_γ for subject 5088 it can be seen that there is no strong evidence against the models tested, although it appears that there is more evidence against the model when the model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$ is fitted. This is evidenced from the higher concentration of p-values close to zero.

As implied earlier, to have strong evidence against a tested model, i.e. to reject the hypothesis that the model is adequate, the posterior p-values, π_γ, should be very small. Therefore as a measure of model goodness of fit, the proportion of π_γ less than 0.05, $\text{Pr}(\pi_\gamma<0.05)$ is calculated for each subject. For comparison purposes, cases with greater $\text{Pr}(\pi_\gamma<0.05)$ show stronger evidence against the model fit.

For 67% of the 66 subjects the proportions of $\pi_\gamma <0.05$ suggest that better fit of the model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. Bar plots in Figure 4 display the $\text{Pr}(\pi_\gamma<0.05)$ obtained from the grouped symptoms model when using different thresholds for Subjects 3022, 4028, 5088, 4045, 4023 and 2013. For these patients, their $\text{Pr}(\pi_\gamma <0.05)$ when using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$ (yellow bars) are higher than when using $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ which is indicated by the red bars.

The same procedure was repeated for comparing the models with and without grouped symptoms. For both models the $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ threshold is used, and the proportion $\text{Pr}(\pi_\gamma <0.05)$ is calculated. Only seven patients show higher $\text{Pr}(\pi_\gamma <0.05)$ when the grouped symptoms model is used compared to the model without grouped symptoms. This suggests that the model with grouped symptoms fits the data better.
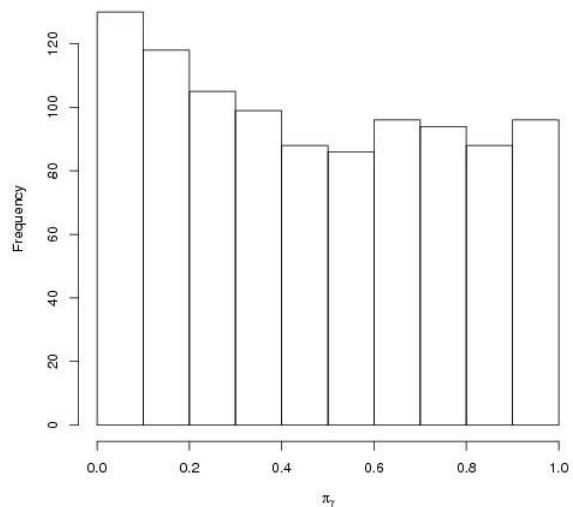
Figure 3. Posterior distribution of p-values, π, for fit of model with thresholds $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ (left) and $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$ (right) for patient 5088 in grouped symptoms model.
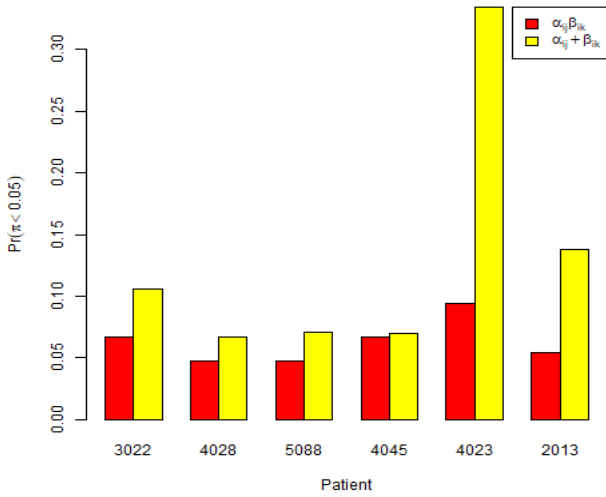
Figure 4. Bar plots comparing the proportion of p-values, $\pi < 0.05$ between different thresholds when using the grouped symptoms model

3.2. Model Verification

The posterior predictive checking approach was applied to study the predictive ability of the core model (without grouped symptoms) and the grouped symptoms model,

using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. Graphical plots in Figure 5 show the posterior distributions of the total predictednumber of reporting symptoms, with blue (dotted) lines marking the total number of predicted reportings, N_p, whereas the red (solid) lines refer to the total observed value, N_{obs} for subjects 4045.

The reporting symptoms for patient 4045 are very well predicted by the grouped symptoms model as indicated by the blue and red lines that almost overlap. The prediction made was $N_p = 15.26$ with 95% CI (9,22) and the symptom reportings i.e. the total observed value, N_{obs}, is 15. However, the nongrouped symptoms model also made a good prediction, although it is slightly over estimated ($N_p = 17.63$).

We also test the performance of different thresholds with the core model. Figure 6 gives the posterior distributions of the total predicted number of symptom reportings for subjects 5009. With each threshold, the predicted distributions comfortably contain the total number of observation, N_{obs} (represented by red solid lines). This indicates that for this patient, we cannot distinguish between the three threshold models in terms of their predictive ability. Note that graphs for all patients exhibit similar trend.
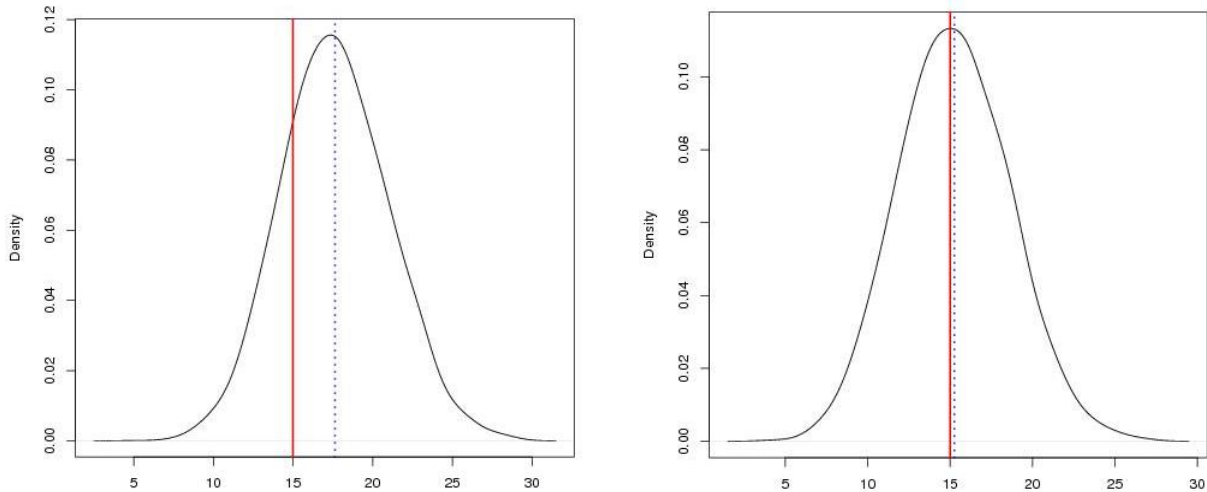
Figure 5. Posterior density plots of number of reportings, N_p, for patients 4045 under non-grouped symptoms model (left) and grouped symptoms model (right) using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. Blue dotted lines show the number of predicted symptoms reportings of each model, and red lines represent the true number of reported symptoms, N_{obs}.
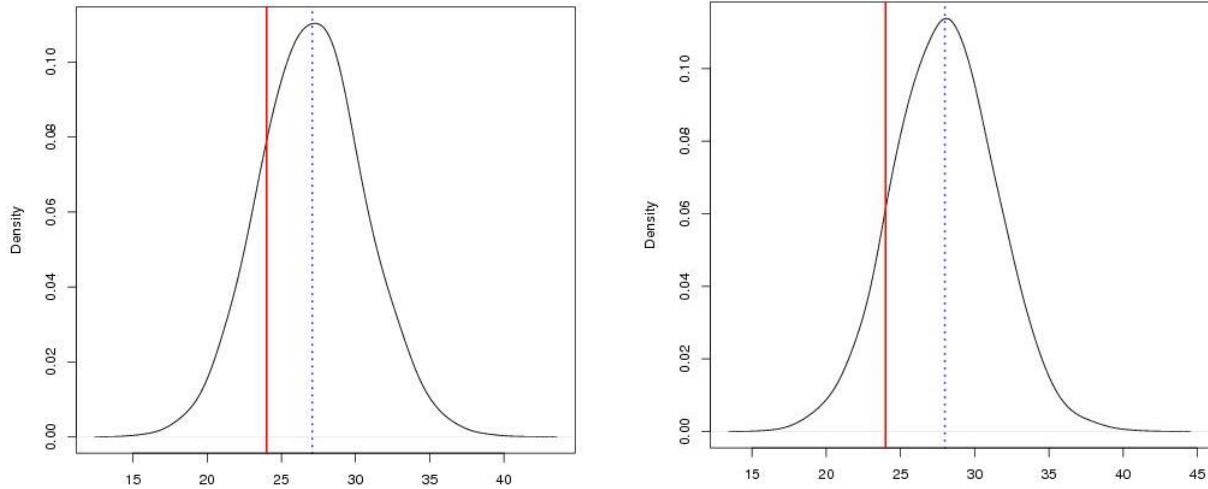
Figure 6. Posterior density plots of number of reportings, N_p, for patients 5009 under non-grouped symptoms model using thresholds $\alpha_{ij}\beta_{ik}$ (left), and $\alpha_{ij}+\beta_{ik}$ (right). Blue dotted lines show the number of predicted symptoms reportings of each model, and red lines represent the true number of reported symptoms, N_{obs}.

Finally, the performance of prediction for different models when using data from all patients in the analysis is compared, i.e. the models with and without grouped symptoms using thresholds $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ and $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$. The results are provided in Table 1. Among the four models, the model with grouped symptoms with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ gives the closest predicted value, N_p, to the observed number of symptoms reported. Figure 7 shows the posterior distributions of the predicted number of symptoms reported. The total number of symptoms predicted to be reported, is 754.3, with a 95% credible interval of (713,797), which contains the observed number of reported symptoms, 771. The other three models considered here do not perform well in terms of this prediction, with the corresponding posterior predictive distributions failing to contain the true value.

Table 1. Model predictions for validation sample for all subjects (The number of reported symptoms, N_{obs} = 771).

	$\alpha_{ij}\beta_{ik}$		$\alpha_{ij} + \beta_{ik}$	
	mean	95% CI	mean	95% CI
PPV	0.413	(0.390,0.436)	0.330	(0.310,0.348)
NPV	0.933	(0.930,0.936)	0.931	(0.928,0.935)
TPR	0.464	(0.435,0.489)	0.469	(0.437,0.498)
TNR	0.919	(0.913,0.925)	0.883	(0.876,0.890)
N_p	866.63	(820,908)	1095.22	(1044,1142)

(a)   Non-grouped symptoms model

	$\alpha_{ij}\beta_{ik}$		$\alpha_{ij} + \beta_{ik}$	
	mean	95% CI	mean	95% CI
PPV	0.408	(0.383,0.432)	0.380	(0.359,0.402)
NPV	0.926	(0.923,0.930)	0.931	(0.928,0.935)
TPR	0.399	(0.435,0.489)	0.457	(0.431,0.492)
TNR	0.929	(0.913,0.925)	0.908	(0.902,0.914)
N_p	754.3	(713,797)	928.084	(876,969)

(b)   Grouped symptoms model
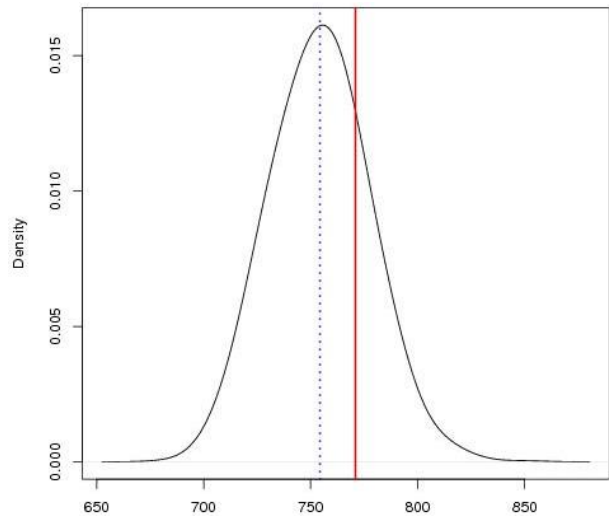
Figure 7. Posterior density plots of number of reportings, N_p, for all patients under grouped symptoms model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. Blue dotted lines show the number of predicted symptom reportings of the model, and red lines represent the true number of reported symptoms, N_{obs}.

Regarding the other four predictive measures presented in Table 1, the models explored here do not display substantial differences. It is also obvious that prediction referring to symptoms not being experienced (NPV, TNR) is much more successful, as compared to prediction for reported symptoms (PPV, TPR). The fact that the developed models perform better in predicting that symptoms will not be reported, may be explained by the nature of the data, where the frequency of reporting symptoms is relatively low (771/7033). The proportion of symptoms with positive prediction that was correctly classified as reported is highest when using the core model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$, i.e. PPV = 0.413, whereas the chance of a symptom not present in an episode given that the model predict it will not be reported is also

highest with this model (NPV = 0.933).

4. Conclusions

This paper discusses the assessment of models with different thresholds using the concept of stochastic latent residuals. It was found that the grouped symptoms model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ fits the data better. Performing the model verification and posterior predictive checking to verify which model is best in predicting symptom reporting, it is concluded that the grouped symptoms model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ has more reliable predictive ability as compared to other models.

Acknowledgements

REFERENCES

[1] Pennebaker, J. W, Cox, D. J., Gonder-Frederick, L., Wunsch, M. G., Evans, W. S., & Pohl, S. "Physical symptoms related to blood glucose in insulin-dependent diabetics," Psychosomatic Medicine, 43(6), 489–500, 1981.

[2] Cox DJ, Gonder-Frederick L, Antoun B, Cryer PE, Clarke WL. "Perceived symptoms in the recognition of hypoglycemia," Diabetes Care, 16(2):519-527, 1993.

[3] Macfarlane PI, Smith CS. "Perceptions of hypoglycaemia in childhood diabetes mellitus: a questionnaire study," Pract Diabetes, 5:56–58, 1988.

[4] Zammitt, N., Streftaris, G., Gibson, G., Deary, I., and Frier, B. Modeling the consistency of hypoglycemic symptoms: high variability in diabetes, Diabetes Technology & Therapeutics, 13(5):571–578, 2011.

[5] Zulkafli, H., Streftaris, G., Gibson, G., and Zammitt, N. Bayesian modelling of the consistency of symptoms reported during hypoglycaemia for individual patients, Malaysian Journal of Mathematical Sciences, 10(S):27-39, 2016.

[6] Cox, D. R. and Snell, E. J. A general definition of residuals. Journal of the Royal Statistical Society. Series B (Methodological), pages 248–275, 1968.

[7] McCullagh, Peter, and John A. Nelder. "Generalized Linear Models 2nd Edition Chapman and Hall." London, UK, 1989.

[8] McEwan, P., Foos, V., Palmer, J.L., Lamotte, M., Lloyd, A. and Grant, D. "Validation of the IMS CORE diabetes model," Value in Health, 17(6), pp.714-724, 2014

[9] Mohd Saifullah Rusiman, Siti Nasuha Md Nor, Suparman, Siti Noor Asyikin Mohd Razali, "Robust Method in Multiple Linear Regression Model on Diabetes Patients," Mathematics and Statistics, Vol. 8, No. 2A, pp. 36 - 39, 2020. DOI: 10.13189/ms.2020.081306.

[10] Albert, J. and Chib, S. "Bayesian residual analysis for binary response regression models." Biometrika, 82(4), pp.747-769, 1995.

[11] Farias, R.B. and Branco, M.D. "Efficient algorithms for Bayesian binary regression model with skew-probit link," In Recent Advances in Biostatistics: False Discovery Rates, Survival Analysis, and Related Topics (pp. 143-168), 2011.

[12] Fouladi, Rachel T. "Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality," Structural Equation Modeling 7, no. 3: 356-410, 2000.

[13] Moshagen, M. "The model size effect in SEM: Inflated goodness-of-fit statistics are due to the size of the covariance matrix," Structural Equation Modeling: A Multidisciplinary Journal, 19(1), pp.86-98, 2012.

[14] UK Hypoglycaemia Study Group: Risks of hypoglycaemia in types 1 and 2 diabetes: effects of treatment modalities and their duration, Diabetologia, 50:1140–1147, 2007.

[15] Dawid, A. P. and Stone, M. "The functional-model basis of fiducial inference," The Annals of Statistics, pages 1054–1067, 1982

[16] Rubin, D. B. "Bayesianly justifiable and relevant frequency calculations for the applies statisticial," The Annals of Statistics, pages 1151–1172, 1984.

[17] Gelman, A., Meng, X.-L., and Stern, H. "Posterior predictive assessment of model fitness via realized discrepancies," Statistica sinica, 6(4):733–760, 1996.

[18] Streftaris, G., Wallerstein, N., Gibson, G., and Arthur, S. "Modeling probability of blockage at culvert trash screens using bayesian approach," Journal of Hydraulic Engineering, 139(7):716–726, 2013.