From the: *Institut für Medizinische Informationsverarbeitung Biometrie und Epidemiologie*

Ludwig-Maximilians-Universität zu München

Dissertation

zum Erwerb des Doctor of Philosophy (Ph.D.) an der

Medizinischen Fakultät der

The usefulness of dynamic approaches in predicting risk of overweight in children within the PEACHES cohort

vorgelegt von:

Lien Dung Le

aus:

Hanoi, VIetnam

Jahr:

2023

Mit Genehmigung der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

First evaluator: *Univ. Prof. Dr. rer. nat. Ulrich Mansmann*

Second evaluator*: Univ. Prof. Dr. rer. nat. HDR Anne-Laure Boulesteix*

Third evaluator: *Prof. Dr. Rüdiger von Kries*

Fourth evaluator: *Prof. Dr. Christa Meisinger*

Dean: **Prof. Dr. Thomas Gudermann**

Datum der Verteidigung:

27.03.2023

Table of content

Table of content **3**

Abstract **6**

List of figures **8**

List of tables **11**

List of abbreviations **12**

1. Introduction **13**

2. Material and Methods **17**

2.1 General remarks 17
2.1.1 Wordings 17
2.1.2 Discrete vs. continuous time variable 17

2.2 PEACHES study 17
2.2.1 Description of original data set 17
2.2.2 Preparation of data set for analysis 18
2.2.3 PEACHES data used for analyses 19
2.2.4 Effect of time on growth curve 23

2.3 Main simulation study design 24
2.3.1 Model specification 24
2.3.2 Underlying distributions of additive time covariates for the simulation 25
2.3.3 Calculation of simulated outcome 25

2.4 Simulation ICC study design 29
2.4.1 Variance of random intercept $\tau 2$ 29
2.4.2 Model specification 30
2.4.3 Distributions of time variables 30
2.4.4 Calculation of simulated outcome 30
2.4.5 Sample size of the main simulation study 31

2.5 Prediction models 31
2.5.1 Overview of performed prediction models 31
2.5.2 Bayesian dynamic models 34
2.5.3 Bayesian static models 47
2.5.4 Generalized linear mixed regression model 1 (GLMER1) 54

2.6 Applied measures 56
2.6.1 Measures comparing prediction performance at the individual level 56
2.6.2 Measures comparing prediction performance at the aggregated level 57
2.6.3 Measures describing the relative usefulness of models 58

2.7 Graphical presentation of the applied measures 59
2.7.1 Logarithmized relative prediction errors (LRPE) on individual level 60
2.7.2 Brier Score over time 60
2.7.3 Average prediction error rate over time 60
2.7.4 Scaled Brier Score over time 61
2.7.5 Brier Skill Score vs. GLMM over time 61
2.7.6 Brier Skill Score vs. “U1-model” over time 61

2.7.7 Scatter plots of Brier Score for different ICC scenarios 61
2.7.8 Calibration plots 61
2.8 Quantification of the association between prediction performance and overall time effect 62
2.9 Results on the individual level 62
2.10 Examining factors that associate with prediction error 62
2.11 Supplementary analyses 63
2.11.1 Extended GLMM models 63
2.11.2 Comparing models using error rate 70
2.11.3 Analyses with smaller sample size 70
2.12 Literature search 71
2.13 Statistical program 71

3. Results 72

3.1 PEACHES study 72
3.1.1 Relative prediction error 72
3.1.2 Brier Score of Bayesian models vs. GLMER1 80
3.1.3 Calibration of Bayesian models vs. GLMER1 83
3.1.4 Scaled Brier Score of Bayesian models vs. GLMER1 86
3.1.5 Brier Skill Score 86
3.1.6 Results at individual level - PEACHES study 88
3.2 Simulation main study 93
3.2.1 Simulated data used for analyses 93
3.2.2 Relative prediction error 93
3.2.3 Brier Score of Bayesian models vs. GLMER1 101
3.2.4 Calibration of Bayesian models vs. GLMER1 104
3.2.5 Scaled Brier Score of Bayesian models vs. GLMER1 107
3.2.6 Brier Skill Score 107
3.2.7 Results at individual level - Simulation main study 108
3.3 Simulation ICC study 114
3.3.1 Brier Score among Bayesian models 114
3.3.2 Brier Score of Bayesian models vs. GLMER1 118
3.3.3 Calibration of Bayesian models vs. GLMER1 121
3.3.4 Scaled Brier Score of Bayesian models vs. GLMER1 127
3.3.5 Brier Skill Score 128
3.3.6 Results at individual level - Simulation ICC study 132
3.3.7 Influence of ICC and time on prediction performance 137
3.4 Supplementary analyses 139
3.4.1 Extended GLMM models 139
3.4.2 Comparing models using error rate 153
3.4.3 Analyses with smaller sample size 158
3.5 Literature search 158
3.6 Summary of the results 159

4. Discussion 160

References 166

Appendix A: Technical information **169**

Appendix B: Supplementary results **171**

Appendix C: Results of literature search **174**

Appendix D: Reproducibility of the results **176**

Acknowledgements **177**

Affidavit **178**

Confirmation of congruency **179**

List of publications **180**

Abstract

Background and objectives:

Researchers are often motivated to update established and validated prognostic scores, due to the well-known issue of degraded predictive performance of prediction models over time and across populations. Dynamic prediction approaches have been shown to have the potential to tackle this issue. However, there is little evidence about the usefulness of such dynamic prediction approaches. Especially, in clinical settings with repeated measurements, depending on data's correlation structure, such dynamic approaches bring greater complexity and more difficult implementation with them. This thesis aims to study the usefulness of dynamic approaches, compared with generalized linear mixed model (GLMER) in predicting risk of overweight in children within the Programming of Enhanced Adiposity Risk in Childhood - Early Screening (PEACHES) cohort, where arthrometric measurements were obtained over ten well-child visits during the first living years of the participating children. Results of this analysis are compared with those of a simulated study, where the attempt was made to imitate children's BMI development over their first five living years. This thesis also aims to explore factors that potentially influence the usefulness of dynamic approaches.

Methods: Analyses were performed with 1) PEACHES cohort study: this data set contains growth data of 1,707 children and pregnancy weight data of their mothers, who were recruited and observed in the prospective PEACHES cohort study between during 18th August 2010 and 16th July 2018; 2) simulation data that imitated PEACHES study settings but introduced more control over the data randomness; and 3) simulated data of the same design as in 2) but different random intercept variance τ^2 values were considered. Common metrics such as Brier Score, Scaled Brier Score, Brier Skill Score, and calibration plots were used to compare prediction performance of the models. The results of all analyses were then compared graphically. The difference in prediction performance among models was quantified applying linear mixed models on logarithmized relative prediction error. The impact of τ^2 on the usefulness of dynamic approaches was then quantified applying linear mixed models. The following models were compared with one another: 1) generalized linear mixed model trained (GLMER1), where prediction was made at U1 and no updates over time were made; 2) Bayesian static model, future outcome are updated using outcome of the past visits (BSM1); 3) Bayesian static model, where prediction was made at U1 and no updates over time were made (BSM2); 4) Bayesian dynamic model, where individual random intercepts and outcome are updated using outcome of the past visits and RIs from the last visits (BDM1); and 5) Bayesian dynamic model, where fixed effects, individual random intercepts and outcome are updated using outcome of the past visits and estimated random intercepts and fixed effects from the last visits (BDM2). In an extended analysis, three other GLMER models were considered, where child's and/or mother's covariates were incorporated.

Results: In PEACHES study, BSM1 and BSM2 show overall similar prediction errors as GLMER1, while it was shown that with increasing amount of information from past visits, model updating in BDM1 and BDM2 leads to improvement in prediction. However, this improvement was rather observed at later visits. Results of the analyses with error rates show similar but less pronounced results. Overall, results of different metrics show great agreement. Results of the simulation main study mostly agreed with those of PEACHES study. The overperformance of BDM1 and BDM2, which was observed with Brier Skill Score and Scaled Brier Score seems to be more pronounced in the simulation settings. Results of the simulation study with different τ^2 show that inter-individual variability strongly influences the overperformance of BDM1 and BDM2, relative

to GLMER1. With higher τ^2, the overperformance of BDM1 and BDM2 is more pronounced. Results of analyses at individual level in the simulation study with different τ^2 show that the more amount of information from past visits is available, the better BDM1 and BDM2 can capture the overall distribution of the simulated random intercepts.

Conclusion: Dynamic prediction approaches, despite the well-known challenges they bring, have the potential to offer advantages over traditional prediction methods. These challenges can be tackled with the development of information technologies and increased data quality. It is necessary to carefully evaluate the usefulness of dynamic approaches in the designing stage of clinical or epidemiological studies under consideration of study settings and assumed parameters.

List of figures

Figure 1. (left) BMI Z score of children in training dataset; (right) mean risk of overweight throughout the study. Average numbers of days from U1 at each visit are depicted (0d, 3d, 33d, etc…) 20
Figure 2. (left) BMI Z score of children in validation dataset; (right) mean risk of overweight throughout the study. Average numbers of days from U1 at each visit are depicted (0d, 3d, 33d, etc…) 22
Figure 3. LRPE of GLMER1 vs. BSM1 over time in PEACHES study 73
Figure 4. LRPE of GLMER1 vs. BSM2 over time in PEACHES study 75
Figure 5. LRPE of GLMER1 vs. BDM1 over time in PEACHES study 77
Figure 6. LRPE of GLMER1 vs. BDM2 over time in PEACHES study 79
Figure 7. Brier Score of Bayesian models vs. GLMER1 over time in PEACHES study 82
Figure 8. Calibration of Bayesian models vs. GLMER1 predicting U2 outcome using U1 data in PEACHES study 84
Figure 9. Calibration of Bayesian models vs. GLMER1 predicting U9 outcome using cumulated data from U1 until U8 in PEACHES study 85
Figure 10. Scaled Brier Score of Bayesian models vs. GLMER1 over time in PEACHES study 86
Figure 11. Brier Skill Score of Bayesian models vs. GLMER1 over time in PEACHES study 87
Figure 12. Brier Skill Score vs. “U1-model” over time in PEACHES study 87
Figure 13. Distributions of estimated random intercepts over time –Bayesian models vs. GLMER1 in PEACHES study 89
Figure 14. Random intercept estimated by Bayesian models vs. GLMER1 over time in LOOCV with validation ID = 2 (left: expected value, right: standard error) 90
Figure 15. Fixed effects estimated by Bayesian models vs. GLMER1 over time in LOOCV with validation ID = 2 91
Figure 16. Outcome predicted by Bayesian models vs. GLMER1 over time in LOOCV with validation ID = 2 92
Figure 17. Mean risk of overweight in large simulation data set 93
Figure 18. LRPE of GLMER1 vs. BSM1 over time in simulation main study 94
Figure 19. LRPE of GLMER1 vs. BSM2 over time in simulation main study 96
Figure 20. LRPE of GLMER1 vs. BDM1 over time in simulation main study 98
Figure 21. LRPE of BDM2 vs. GLMER1 over time in simulation study 100
Figure 22. Brier Score of Bayesian models vs. GLMER1 over time in simulation main study 103
Figure 23. Calibration of Bayesian models vs. GLMER1 predicting U2 outcome using U1 data 105
Figure 24. Calibration of Bayesian models vs. GLMER1 predicting U9 outcome using cumulated data from U1 until U8 106
Figure 25. Scaled Brier Score of Bayesian models vs. GLMER1 in simulation main study 107
Figure 26. Brier Skill Score of Bayesian models vs. GLMER1 over time in simulation main study 107
Figure 27. Brier Skill Score of Bayesian models vs. “U1-model” over time in simulation main study 108
Figure 28. Distributions of random intercepts over time – compared between Bayesian vs. GLMER1 models in simulation main study 109
Figure 29. Random intercept estimated by Bayesian models vs. GLMER1 over time in LOOCV with validation ID = 1 111
Figure 30. Fixed effects estimated by Bayesian models vs. GLMER1 over time in LOOCV with validation ID = 1 112

Figure 31. Outcome predicted by Bayesian models vs. GLMER1 over time in LOOCV with validation ID = 1 113
Figure 32. Impact of ICC on difference of performance between BSM1 and BSM2 115
Figure 33. Impact of ICC on difference of performance between BDM1 and BDM2 116
Figure 34. Impact of ICC on difference of performance between BSM1 and BDM1 117
Figure 35. Impact of ICC on difference of Brier Score between BSM1 and GLMER1 119
Figure 36. Impact of ICC on difference of Brier Score between BDM1 and GLMER1 120
Figure 37. Calibration of Bayesian models vs. GLMER1 predicting U2 outcome using U1 data - $ICC = 0$ 121
Figure 38. Calibration of Bayesian models vs. GLMER1 predicting U2 outcome using U1 data - $ICC = 0.549$ 122
Figure 39. Calibration of Bayesian models vs. GLMER1 predicting U2 outcome using U1 data - $ICC = 0.884$ 123
Figure 40. Calibration of Bayesian models vs. GLMER1 predicting U9 outcome using cumulated data from U1 until U8- $ICC = 0$ 124
Figure 41. Calibration of Bayesian models vs. GLMER1 predicting U9 outcome using cumulated data from U1 until U8- $ICC = 0.549$ 125
Figure 42. Calibration of Bayesian models vs. GLMER1 predicting U9 outcome using cumulated data from U1 until U8-$ICC = 0.884$ 126
Figure 43. Scaled Brier Score of Bayesian models vs. GLMER1 over time in simulation study - $ICC = 0$ 127
Figure 44. Scaled Brier Score of Bayesian models vs. GLMER1 over time in simulation study - $ICC = 0.549$ 127
Figure 45. Scaled Brier Score of Bayesian models vs. GLMER1 over time in simulation study - $ICC = 0.884$ 128
Figure 46. Brier Skill Score of Bayesian models over time in simulation study - $ICC = 0$ .128
Figure 47. Brier Skill Score of Bayesian models over time in simulation study – $ICC = 0.549$ 129
Figure 48. Brier Skill Score of Bayesian models over time in simulation study - $ICC = 0.884$ 129
Figure 49. Brier Skill Score of Bayesian models (reference = “U1-model”) over time in simulation study - $ICC = 0$ 130
Figure 50. Brier Skill Score of Bayesian models (reference = “U1-model”)” over time in simulation study – $ICC = 0.549$ 131
Figure 51. Brier Skill Score of Bayesian models (reference = “U1-model”) over time in simulation study - $ICC = 0.884$ 131
Figure 52. Distributions of random intercepts over time – compared between Bayesian vs. GLMER1 models in simulation study with $ICC = 0$ 132
Figure 53. Distributions of random intercepts over time – compared between Bayesian vs. GLMER1 models in simulation study with ICC$= 0.233$ 132
Figure 54. Distributions of random intercepts over time – compared between Bayesian vs. GLMER1 models in simulation study with $ICC = 0.549$ 133
Figure 55. Distributions of random intercepts over time – compared between Bayesian vs. GLMER1 models in simulation study with $ICC = 0.884$ 133
Figure 56. Distributions of random intercepts over time –compared between Bayesian vs. GLMER1 models in simulation study with $ICC = 0.968$ 134
Figure 57. Random intercept estimated by Bayesian vs. GLMER1 models over time in LOOCV with validation ID = 1 in simulation study with $ICC = 0$ 134
Figure 58. Random intercept estimated by Bayesian vs. GLMER1 models over time in LOOCV with validation ID = 1 in simulation study with $ICC = 0.549$ 135
Figure 59. Random intercept estimated by Bayesian vs. GLMER1 models over time in LOOCV with validation ID = 1 in simulation study with $ICC = 0.884$ 135
Figure 60. Outcome predicted by Bayesian vs. GLMER1 models over time in LOOCV with validation ID = 1 in simulation study with $ICC = 0$ 136

Figure 61. Outcome predicted by Bayesian vs. GLMER1 models over time in LOOCV with validation ID = 1 in simulation study with $ICC = 0.549$ 136
Figure 62. Outcome predicted by Bayesian vs. GLMER1 models over time in LOOCV with validation ID = 1 in simulation study with $ICC = 0.884$ 137
Figure 63. Prediction performance of GLMER4 vs. BSM1 over time in PEACHES study. 140
Figure 64. Prediction performance of GLMER4 vs. BSM2 over time in PEACHES study. 142
Figure 65. Prediction performance of GLMER4 vs. BDM1 over time in PEACHES study 144
Figure 66. Prediction performance of GLMER4 vs. BDM2 over time in PEACHES study 146
Figure 67. Brier Score of Bayesian vs. all GLMM models in PEACHES study..... 148
Figure 68. Brier Score of Bayesian vs. all GLMM models, from U4 in PEACHES study... 149
Figure 69. Scaled Brier Score over time in PEACHES study 152
Figure 70. Brier Skill Score vs. GLMER4 over time in PEACHES study 152
Figure 71. Error rate of Bayesian models vs. GLMER1 over time – PEACHES study..... 154
Figure 72. Error rate of Bayesian models vs. GLMM models over time – PEACHES study 155
Figure 73. Error rate of Bayesian models vs. GLMER1 over time - simulation main study 157

List of tables

Table 1. Inclusion and exclusion criteria for PEACHES analyses 18
Table 2. Number of available (non-missing) BMI Z-score at each visit in training dataset .. 20
Table 3. Mean absolute BMI Z-scores at all visits in training dataset 20
Table 4. Mean risk of overweight at all visits in training dataset 21
Table 5. Mean absolute BMI Z-scores at each visit in validation dataset 22
Table 6. Mean risk of overweight at each visit in validation dataset 22
Table 7. Example of time variables in PEACHES data set 23
Table 8. Parameters used for simulation design 24
Table 9. Assumptions for time window of visits, recommended by BzgA 25
Table 10. Simulation steps using package `simstudy` (Version 0.2.1) 26
Table 11. Hypothetical $\tau 2$values defined and their corresponding ICC 29
Table 12. Parameters used for simulation design 30
Table 13. Overview of the prediction models 33
Table 14. Model specification of model BDM1 34
Table 15. Model specification of model BDM2 40
Table 16. Model specification of model BSM1 47
Table 17. Model specification of model BSM2 51
Table 18. Model specification of model GLMER1 54
Table 19. Model specification of model GLMER2 63
Table 20. Model specification of model GLMER3 66
Table 21. Model specification of model GLMER4 68
Table 22. LRPE of GLMER1 vs. BSM1 in PEACHES study – output of LMM 72
Table 23. LRPE of GLMER1 vs. BSM2 in PEACHES study – output of LMM 74
Table 24. LRPE of GLMER1 vs. BDM1 in PEACHES study – output of LMM 76
Table 25. LRPE of GLMER1 vs. BDM2 in PEACHES study– output of LMM 78
Table 26. RPE of GLMER1 vs. Bayesian models in PEACHES study 78
Table 27. Brier Score of Bayesian models vs. GLMER1 over time in PEACHES study 81
Table 28. Calibration slopes and intercepts of Bayesian models vs. GLMER1 in PEACHES study 83
Table 29. LRPE of GLMER1 vs. BSM1 in simulation main study – output of LMM 93
Table 30. LRPE of GLMER1 vs. BSM2 in simulation study – output of LMM 95
Table 31. LRPE of GLMER1 vs. BDM1 in simulation main study – output of LMM 97
Table 32. LRPE of GLMER1 vs. BDM2 in simulation study – output of LMM 99
Table 33. RPE of GLMER1 vs. Bayesian models in simulation main study 99
Table 34. Brier Score of dynamic models vs. GLMER1 over time 101
Table 35. Calibration slopes and intercepts of Bayesian models vs. GLMER1 – simulation main study 104
Table 36. Influence of ICC and time on Brier Score – Output of LMM 137
Table 37. Influence of ICC and time on MLRPE – Output of LMM 138
Table 38. Prediction performance of BSM1 vs. GLMER4 – quantified with LMM 139
Table 39. Prediction performance of BSM2 vs. GLMER4 - quantified with LMM 141
Table 40. Prediction performance of BDM1 vs. GLMER4 – quantified with LMM 143
Table 41. Prediction performance of BDM2 vs. GLMER4 – quantified with LMM 145
Table 42. RPE of GLMER4 vs. Bayesian models in PEACHES study 147
Table 43. Calibration of Bayesian models vs. GLMER4 151
Table 44. Error rate of Bayesian models vs. GLMER1 over time in PEACHES study 153
Table 45. Error rate of Bayesian models vs. GLMER1 over time in simulation main study 156

List of abbreviations

APER	Average Prediction Error Rate
BMI	Body-mass index
BS	Brier score
BSS	Brier Skill Score
BzgA	Bundeszentrale für gesundheitliche Aufklärung
GLMM	generalized linear mixed model
ICC	Intraclass correlation coefficient
LGA	Large-for-gestational-age
LMM	linear mixed model
PEACHES	Programming of Enhanced Adiposity Risk in Childhood - Early Screening
RPE	Relative Prediction Error (based on Brier's score) on the natural logarithmic scale
RPER	Relative Prediction Error Rate
sBS	Scaled Brier Score
SGA	small-for-gestational-age
zBMI	BMI Z-score

1. Introduction

Objectives

This thesis aims to study the usefulness of dynamic approaches in predicting risk of overweight in children within the PEACHES cohort. Results of this analysis are compared with those of a simulated study, where randomness presumably due to measurement errors is introduced into the data in a controlled way. This thesis also aims to explore factors that influence the usefulness of dynamic approaches, if any exists.

Dynamic prediction models – definition and usage in the literature

The followings present the definition and usage of dynamic prediction models in the literature. The findings presented below result from a focused literature review that was conducted within this thesis. The motivation to update prediction models or established prognostic scores was encountered in the identified articles. The issue of degraded predictive performance of models and scores over time is behind that motivation. From the focused literature review, five identified studies raised the necessity for updating and attempted to update established risk scores or prognostic models. The studies by Hafkamp-De Groen, Hickey et al. and Hippisley-Cox et al. dealt with the PIAMA score (for predicting asthma in pre-school children with asthma like-symptoms) (Hafkamp-De Groen, 2013), the EuroSCORE (for cardiac surgery) (Hickey, 2013) and QRISK (for cardiovascular disease) (Hippisley-Cox et al., 2011), respectively. The study by Houwelingen and Thorogood aimed to update model predicting kidney graft survival (Houwelingen and Thorogood, 1995), while Genders et al. suggested an updated version of the Diamond–Forrester model for estimating the probability of obstructive coronary disease (Genders, 2011) and Hafkamp-De Groen provided a revised version of the risk model. Hoewelingen and Thorogood applied fine-tuning of the prognostic model by shrinkage with validation data set (Houwelingen and Thorogood, 1995). Hafkamp-De Groen, Hippisley-Cox et al. and Genders et al. chose model revision, where model coefficients were refitted (Hippisley-Cox et al., 2011, Genders, 2011). After reassessing model performance, Hickey et al. came to the conclusion that the EuroSCORE model showed systematic deterioration in the calibration over time for patient population in England and Wales (Hickey, 2013). Consequently, they performed analyses comparing different updating approaches for this model, which include 1) periodic update every one or two year(s), 2) monthly update (“rolling window”), and 3) dynamic logistic regression (Hickey et al., 2013).

Other identified studies compared different approaches using real data (Siregar et al., 2016, Davis et al., 2019a, Steyerberg et al., 2004, Su et al., 2018, Vergouwe et al., 2017, Janssen et al., 2008). Approaches that were addressed in these studies include recalibration (Siregar et al., 2016) (Davis et al., 2019a, Steyerberg et al., 2004, Su et al., 2018, Vergouwe et al., 2017, Janssen et al., 2008), structural model revision (Siregar et al., 2016, Steyerberg et al., 2004, Su et al., 2018, Vergouwe et al., 2017, Janssen et al., 2008), Bayesian dynamic model (Siregar et al., 2016) and Bayesian dynamic model incorporating a forgetting factor (Su et al., 2018).

In time-to-event data settings, joint models seem to be commonly used to incorporate repeated measurements into a time-to-event prediction model. In the literature, for instance, in the two identified studies (Andrinopoulou et al., 2017, Posch et al., 2020), this approach was often referred to as a dynamic approach. As an alternative, landmark analysis or “landmarking” have been used to allow time-dependence and certain dynamics into time-to-event prediction models. A few examples in the literature can be named: (Heyard et al., 2019, Keogh et al., 2019, Li et al., 2017, Rizopoulos et al., 2017, Suresh et al., 2017). These approaches are not in line with the

data settings and the aim of this thesis, no further details regarding these two approaches will be provided in the followings.

More formally, the approach of gradual model updating was formulated in the text book “Clinical Prediction Models” by Steyerberg (Steyerberg, 2009). There, it was recommended to start with simple updating methods, such as re-calibration, then continue with model revision and model extension (Steyerberg, 2009). However, no recommendations were made for settings with correlated data such as repeated measurements. Considering the identified articles, I believe that the systematic review by Jenkins et al. (Jenkins et al., 2018) could capture pretty well the research landscape in the field of dynamic prediction and provided the current state of the art. They categorized their findings into three groups: 1) model updating in discrete steps with frequentist methods, 2) Bayesian updating methods, and 3) methods based on time-varying coefficients (Jenkins et al., 2018). Dynamic approaches that are demonstrated in this thesis fall into the second category. Here, the focus lies in the comparison of Bayesian methods of different dynamic level with a frequentist static model (“GLMER1”). Additionally, the comparison was performed with extended frequentist model (“GLMER4”), which incorporates updated data and introduces dynamics in the model in a discrete manner.

The followings give a deeper focus on the study by Finkelman et al. (Finkelman et al., 2016) and present the findings from the literature review regarding potential benefits of dynamic approaches. This study used simulated data to compare 1) a static linear static model, 2) a static linear mixed effects model with random intercept only, 3) a static linear mixed effects model with both random intercept and random slope 4) a dynamic linear model, 5) a Bayesian linear mixed effects (BLME) model with random intercept only, and 6) a BLME model with both random intercept and random slope with regard to prediction accuracy in clustered populations. Dependent variable of these models represented a hypothetical normally distributed, continuous clinical outcome. Dynamics in the dynamic linear model and the two BLME models was introduced in a way that after making predictions on a certain group of patients, the predicted outcome data on these individuals were then used as data for the predictions of the next group of patients. This was repeated until predictions for all patients in the test data set had been made. The model priors were not affected, only data was updated after each prediction (Finkelman et al., 2016). The authors of this study came to the conclusion that model updating provided great gain in accuracy, while the effect of overfitting is comparably small and even complex dynamic models are more useful than static models (Finkelman et al., 2016). Results of this study also suggested substantial improvement of prediction performance in a clustered data setting by using dynamic approaches (Finkelman et al., 2016). Siregar et al. considered Bayesian models as good alternative for model updating in both small and large data sets (Siregar et al., 2016).

Results of the studies mentioned above (Siregar et al., 2016, Davis et al., 2019a, Steyerberg et al., 2004, Su et al., 2018, Vergouwe et al., 2017, Janssen et al., 2008, Hickey et al., 2013) all agree on the improvement in prediction performance by model updating. Davis et al., Steyerberg et al., Su et al., Vergouwe et al., and Janssen et al. all agree on the statistical overperformance, thus preference for simpler updating methods over extensive model revision (Janssen et al., 2008, Steyerberg et al., 2004, Su et al., 2018, Vergouwe et al., 2017, Davis et al., 2019a). Hickey et al. and Siregar et al. concluded the potential benefit of Bayesian dynamic modeling over static approaches (Hickey et al., 2013, Siregar et al., 2016).

The findings of Plate et al. (Plate et al., 2019) suggested an increasing number of studies with repeated measurements in healthcare, and at the same time, their reluctant usage. Incorporating

these measurements can potentially optimize prediction models. The reasons for failing to implement such measurements into the prediction models that was stated by the authors included the uncertainty of the usefulness of (probably more complex) methods accounting for repeated measurements and the lack of added clinical value of these models due to their complicated implementation (Plate et al., 2019). The authors also suggested the necessity of a framework of possible approaches to help researchers with optimizing prediction models (Plate et al., 2019).

This suggestion and other findings in the literature about potential benefit of dynamic approaches motivated this research work and support the relevance of its results.

Common prediction performance measures

Common measures that have been used in the literature to capture prediction performance were summarized by Steyerberg et al. (Steyerberg et al., 2010). This summary served as basis to choose criteria measuring prediction performance of the models in this thesis. As suggested by the authors, discrimination and calibration are two important aspects that need be addressed when assessing performance of prediction models.

In identified published works, the following discrimination and calibration measures were used: area under the receiver operating characteristic curve (AUC) (Siregar et al., 2016, Davis et al., 2019a, Su et al., 2018), Brier Score (Davis et al., 2019a), logarithmic score (Davis et al., 2019a), Observed to Expected ratio (Davis et al., 2019a), calibration plots, intercept, and/or slope (Steyerberg et al., 2004, Su et al., 2018, Vergouwe et al., 2017, Siregar et al., 2016, Janssen et al., 2008), and mean absolute error (Finkelman et al., 2016).

In this thesis, the performance measures applied are Brier Score, (logarithmized) relative prediction error, scaled Brier Score, Brier Skill Score, as well as calibration plots.

Comparison between static and dynamic models

While calibration and discrimination measures have been widely used to compare static and dynamic prediction models, no established measure quantifying or evaluating the overperformance of dynamic over static models seem to exist. From the focused literature review, three proposals have been identified. Finkelman et al. suggested using “relative improvement” for such purpose (Finkelman et al., 2016). In their article, “relative improvement” was described as a metric that ranges from 0 to 1 and demonstrates the improvement in mean absolute error of a model over the intercept-only model, relative to the improvement that would have been observed with the true model (Finkelman et al., 2016). The advantage of this metric lies in its interpretability.

Vergouwe et al. proposed a closed testing procedure to choose an appropriate update method, while controlling for overfitting (Vergouwe et al., 2017). In this procedure, a series of likelihood ratio tests against the original were performed. Alternative models included recalibration and model revision, where the number of coefficients to be updated increased (Vergouwe et al., 2017). The authors suggested that performing these likelihood ratio tests in a step-wise manner helped balance the amount of information and the risk for overfitting (Vergouwe et al., 2017).

Davis et al. proposed a data-driven annual updating strategy, where the simplest updating method with no difference in accuracy compared with the best model, while the best model was defined as model with best median accuracy over all bootstrap samples (Davis et al., 2019b). Interestingly, this strategy considered the updating sample size and aimed to minimize overfitting.

In this thesis, besides the common prediction performance measures mentioned above, relative prediction error on logarithmic scale and linear mixed models were applied to visualize and quantify the performance of dynamic approaches, relative to the static approach.

Data

Data used for the analyses in this thesis come from two sources: 1) the prospective mother-child cohort study PEACHES and 2) simulation.

In PEACHES cohort, pregnant women examined in maternity clinics in Germany (Munich, Düsseldorf, and some provinces in northern Germany) were recruited between 2010 and 2015. In essence, the arthrometric data of their children from birth until their fifth living year as well as mothers' pre-conceptional, prenatal and postpartum data were collected. One main objective of the study focuses on the association between mothers' pregnancy obesity status and risk of overweight and metabolic diseases in children's early life course (Gomes et al., 2018). The nature of arthrometric data contains randomness due to different sources of systematic and unsystematic measurement errors: instrument error, intra-rater error, inter-rater error for example due to different measurement procedures for newborns at different clinics and medical facilities. These sources of measurement errors are difficult to be captured in prediction models. For that reason, simulated data was also used in this thesis, where PEACHES cohort's data of children's BMI over time is imitated. The same analyses were performed with these two sources of data and the results were compared. Besides the so-called simulation main study, as described above, the simulation study was extended to examine the usefulness of dynamic approaches when modifying the structure of random effects. In other words, different intraclass correlation coefficients of the underlying model were analyzed in this approach. This extended study, the so-called simulation ICC study, used different data sets, which were simulated using the same model components as in the simulation main study but various random intercept variance.

Outline of the thesis

This thesis is outlined in the following chapters: Chapter 2 (page17) describes statistical methods and data used. Chapter 3 (page 72) and Chapter 4 (page 160) presents and discusses core results of the analyses, respectively. Additional information are presented in the appendices: Appendix A: Technical information (page 169), Appendix B: Supplementary results (page 171), Appendix C: Results of literature search (page 174), and Appendix D: Reproducibility of the results (page 176).

2. Material and Methods

2.1 General remarks

2.1.1 Wordings

The following wordings are applied throughout the thesis:

- Visit: if not specified otherwise, throughout the thesis, the term “visit” represents the frequent preventive medical check-up that is recommended for children in Germany, the so-called well-child visits. These visits are scheduled at birth (U1), 3 to 10 days after birth (U2), 4 to 5 weeks after birth (U3), 3 to 4 months after birth (U4), 6 to 7 months after birth (U5), 10 to 12 months after birth (U6), 21 to 24 months after birth (U7), 34 to 36 months after birth (U8), 46 to 48 months after birth (U8), and 60 to 64 months after birth (U9) (Bundeszentrale für gesundheitliche Aufklärung, 2020).
- Prediction visit: the visit, where data is collected, and prediction is made for future risk of overweight.
- Future visits: visits that lie in the future, risk of overweight at these timepoints will be predicted.
- ID: a child that participated in the PEACHES data set, or a simulated case in the simulation study.

2.1.2 Discrete vs. continuous time variable

Graphical presentation of most of the results in this thesis does not consider time as a continuous variable. The analyses in section 2.8 attempted to quantify the association between prediction performance and number of previous visits. These analyses were performed on the individual level. Here, child’s age in months was considered.

2.2 PEACHES study

2.2.1 Description of original data set

The data set contains growth data of 1,707 children and pregnancy weight data of their mothers, who were recruited in the Programming of Enhanced Adiposity Risk in Childhood - Early Screening (PEACHES) cohort study between during 18th August 2010 and 16th July 2018.

One of the objectives of PEACHES cohort study was to investigate factors that associate with child’s future risk of overweight. Risk of overweight was measured with BMI-Z score. The cut-off 1 was chosen, which indicates risk for overweight as recommended by the WHO (de Onis M, 2010). If BMI-Z score of a certain child is >1 at a certain visit, he/she is considered to be at risk of overweight.

More details about background as well as variables that were measured and collected in the PEACHES study can be found in the study protocol (Gomes et al., 2018).

2.2.2 Preparation of data set for analysis

The following calculations were done, in order to prepare the data sets PEACHES for analyses within this thesis.

2.2.2.1 Inclusion and exclusion criteria

The inclusion and exclusion criteria applied are shown in the following table.

Table 1. Inclusion and exclusion criteria for PEACHES analyses

Inclusion criteria
- Pre-conceptionally obese or non-obese
- Singleton pregnancy
- Absence of type 1 diabetes mellitus (T1DM) or type 2 diabetes mellitus (T2DM) in mothers (pre-conceptionally)
- Full-term (>=37 weeks 0 days of gestation) live birth
Exclusion criteria
- Underweight mothers
- Twin/multiple pregnancy
- Presence of T1DM or T2DM in mothers (pre-conceptionally)
- Preterm children (gestational age <36 weeks 6 days of gestation)

Please refer to the script "scripts/exclude.R" to see single calculations of this step.

2.2.2.2 Calculation of child's BMI z-scores

The scripts provided by WHO were used to calculate the BMI z-score of the children in the PEACHES and PEPO cohorts. More details about this calculation method can be found in in Appendix D: Reproducibility of the results, 'peaches/create_data_peaches/scripts/who2007_R/ReadMe.pdf' and 'peaches/create_data_peaches/scripts/igrowup_R/ReadMe.pdf'. Please also refer to the script "peaches/create_data_peaches/scripts/calc_bmiz.R" in Appendix D: Reproducibility of the results to see single calculations of this step.

Offspring BMI z-scores were categorized according to the World Health Organisation categories (World Health Organization, 2006).

2.2.2.3 Calculation of total gestational weight gain (GWG)

Total gestational weight gain (GWG) in kilograms was calculated as the difference between the last measured weight before delivery and pre-conception weight and was classified as inadequate, adequate, or excessive according to the BMI-specific recommendations of the Institute of Medicine (Rasmussen et al., 2009). Please refer to the script "peaches/create_data_peaches/scripts/calc_gwg.R" in Appendix D: Reproducibility of the results to see single calculations of this step.

Women with GWG values within the recommended range are considered as having adequate GWG. Women with GWG values below the lower cut-off were considered as having inadequate GWG while those having GWG values above the upper cut-off were considered as having excessive GWG. Please refer to the script "peaches/create_data_peaches/scripts/derive.R" in Appendix D: Reproducibility of the results to see single calculations of this step.

2.2.2.4 Calculation of birth weight categories for gestational age and sex

The file 'peaches/create_data_peaches/input/Voigt2014_ga_long.txt' in Appendix D: Reproducibility of the results with birth weight percentile cutoffs was used to determine the categories large-for-gestational-age (LGA, >90th percentile), average-for-gestational-age (AGA, 10th to 90th percentile), and small-for-gestational-age (SGA, <10th percentile) birth weight for gestational age and sex. These cut-offs were based on the German reference population (Voigt et al., 2014).

Please refer to the script "peaches/create_data_peaches/scripts/calc_ga.R" in Appendix D: Reproducibility of the results to see single calculations of this step.

2.2.2.5 Split data set for training and validation

1,000 children were randomly sampled from the original data set. These 1,000 children entered the training data set. The remaining children were set aside for the purpose of validation. The validation study is not part of this thesis. Therefore, from now on, the validation data set will only be shortly described in 2.2.3.3, but no further analyses were done with it.

2.2.3 PEACHES data used for analyses

2.2.3.1 Variables of interest

The following variables are considered in the analyses within this thesis.

Exposure variables:

- time (in months) after birth at a certain visit (U1, U2, U3, U4, U5, U6, U7, U7a, U8, or U9),
- mother’s pre-conceptional obesity status (obese or non-obese), and
- child’s large-for-gestational age at birth (yes or no)

Outcome variable:

- child’s risk of overweight (defined as BMI Z-score >1) at a certain visit

2.2.3.2 Training dataset

Number of IDs included in the validation dataset is 1000. The following figures describe the course of BMI-Z score and its population mean, as well as the average risk of overweight in all children included in the training dataset.

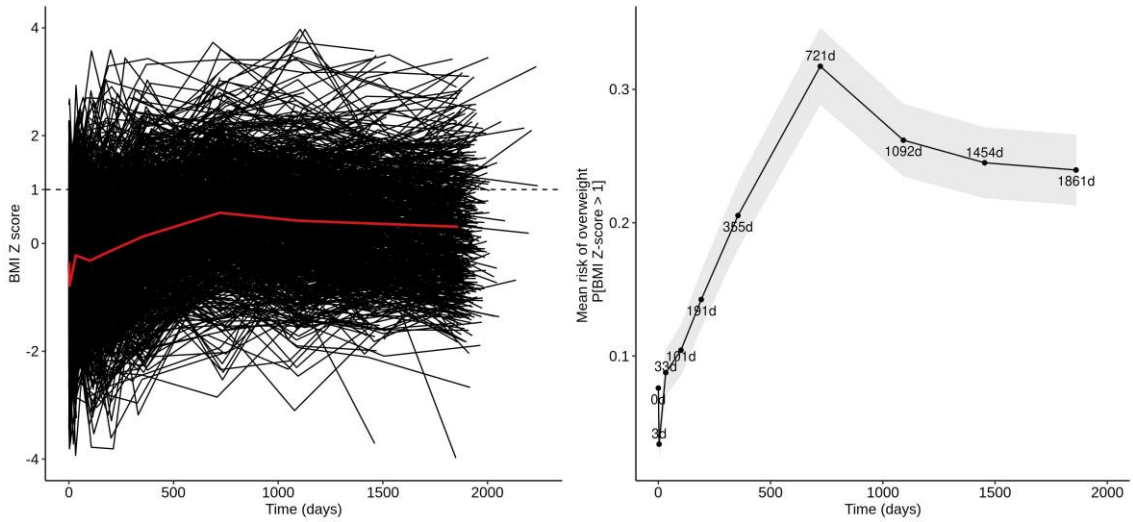


Figure 1. (left) BMI Z score of children in training dataset; (right) mean risk of overweight throughout the study. Average numbers of days from U1 at each visit are depicted (0d, 3d, 33d, etc…)

Table 2. Number of available (non-missing) BMI Z-score at each visit in training dataset

Visit	Number of available observed outcomes	Proportion of available observed outcome
U1	1000	1
U2	972	0.972
U3	982	0.982
U4	977	0.977
U5	962	0.962
U6	969	0.969
U7	936	0.936
U7a	882	0.882
U8	845	0.845
U9	597	0.597

Mean absolute BMI Z-scores of children included in the training dataset can be seen in Table 3. On average, BMI-Z score of children in PEACHES training dataset are lower than 0. From U5 to U7, it increases. From U7 it decreases continually over time

Table 3. Mean absolute BMI Z-scores at all visits in training dataset

Visit	Average number of days from U1	Mean BMI Zscore	[2.5% - 97.5%] of BMI Zscore
U1	0	-0.35	[-2.31;1.63]
U2	3	-0.78	[-2.68;1.24]
U3	33	-0.22	[-2.09;1.59]
U4	101	-0.32	[-2.43;1.81]
U5	191	-0.15	[-2.29;1.94]
U6	355	0.13	[-1.94;2.25]
U7	721	0.57	[-1.42;2.47]

U7a	1,092	0.42	[-1.45;2.57]
U8	1,454	0.37	[-1.46;2.48]
U9	1,861	0.31	[-1.50;2.80]

Table 4 presents the average risk of overweight in PEACHES training dataset at each visit. From U2 to U7, the average risk of overweight increases over time. At U7 (about 2 years after birth), children are at highest risk of overweight. From U7, it tends to decrease.

Table 4. Mean risk of overweight at all visits in training dataset

Timepoint	**Average number of days from U1**	**Mean risk for over-weight**	**[2.5% - 97.5%] of risk for over-weight**
U1	0	0.076	[0;1]
U2	3	0.034	[0;1]
U3	33	0.088	[0;1]
U4	101	0.104	[0;1]
U5	191	0.142	[0;1]
U6	355	0.205	[0;1]
U7	721	0.317	[0;1]
U7a	1,092	0.262	[0;1]
U8	1,454	0.245	[0;1]
U9	1,861	0.24	[0;1]

2.2.3.3 Validation dataset

Number of IDs included in the validation dataset is 557. The following figures describe the course of BMI-Z score throughout the study and the average risk of overweight in all children included in the validation dataset. At U7 (about 2 years after birth), children in the validation dataset are at highest risk of overweight. The shape of the BMI z-score curve observed in the validation dataset seems to be similar to the curve observed in the training dataset. Table 5 and Table 6 present detailed numbers for these figures.

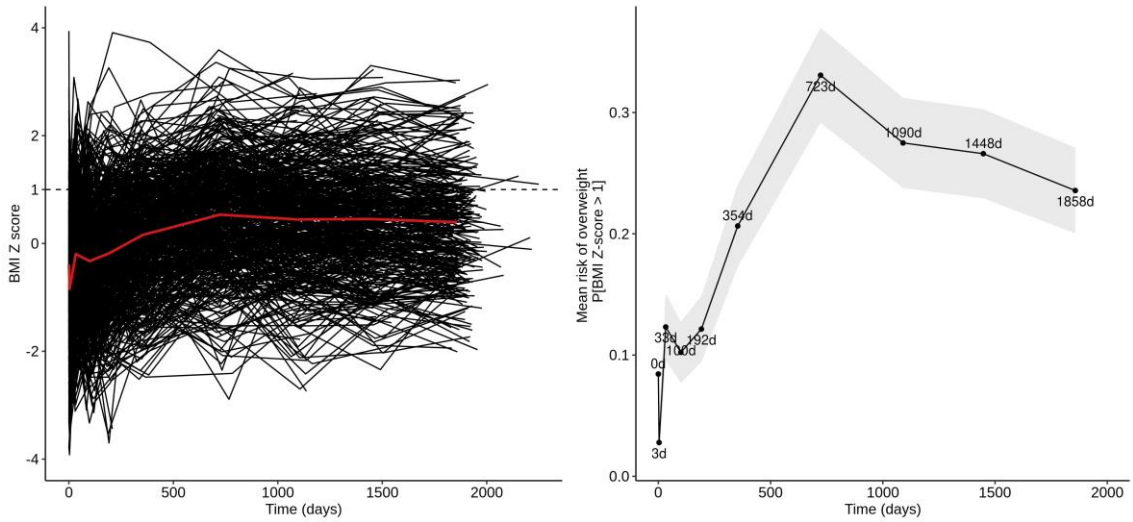


Figure 2. (left) BMI Z score of children in validation dataset; (right) mean risk of overweight throughout the study. Average numbers of days from U1 at each visit are depicted (0d, 3d, 33d, etc…)

Table 5. Mean absolute BMI Z-scores at each visit in validation dataset

Timepoint	Average number of days from U1	Mean BMI Zscore	[2.5% - 97.5%] of BMI Zscore
U1	0	-0.389	[-2.42;1.57]
U2	3	-0.843	[-2.75;1.02]
U3	33	-0.198	[-2.28;1.72]
U4	100	-0.327	[-2.42;1.87]
U5	192	-0.183	[-2.22;1.79]
U6	354	0.158	[-1.91;2.00]
U7	723	0.535	[-1.65;2.53]
U7a	1,090	0.446	[-1.55;2.52]
U8	1,448	0.453	[-1.42;2.71]
U9	1,858	0.395	[-1.54;2.59]

Table 6. Mean risk of overweight at each visit in validation dataset

Timepoint	Average number of days from U1	Mean risk for overweight	[2.5% - 97.5%] of risk for overweight
U1	0	0.076	[0;1]
U2	3	0.034	[0;1]
U3	33	0.088	[0;1]
U4	101	0.104	[0;1]
U5	191	0.142	[0;1]

U6	355	0.205	[0;1]
U7	721	0.317	[0;1]
U7a	1,092	0.262	[0;1]
U8	1,454	0.245	[0;1]
U9	1,861	0.24	[0;1]

2.2.4 Effect of time on growth curve

In order to look for relevant time points that will serve as basis for the simulation design, a linear mixed model with random intercept, the effect of linear and additive time effects at each visit on BMI-Z score over time was examined. Dependent variable was defined as BMI-Z score over time. Independent variables were the time effects. No random slopes were specified. It was assumed that the population homogenous is and possible inhomogeneity can be covered by the random effects. Models were backward selected in a stepwise manner according to their AIC by using the function `step()` in package `lmerTest` (Kuznetsova et al., 2017). The time effects of the selected model then served as input of the simulation design. The full model was specified as the followings:

$$BMIz_{i,j} = \alpha_i + \beta_0 + \beta_1 * t_{1_{i,j}} + \beta_2 * t_{2_{i,j}} + \beta_3 * t_{3_{i,j}} + \beta_4 * t_{4_{i,j}} + \beta_5 * t_{5_{i,j}} + \beta_6 * t_{6_{i,j}} + \beta_7 * t_{7_{i,j}} + \beta_{7a} * t_{7a_{i,j}} + \beta_8 * t_{8_{i,j}} + \varepsilon_{i,j}$$

$$\alpha_i \sim N(0, \tau^2)$$

$$\varepsilon_{i,j} \sim N(0, \sigma^2)$$

An example of time variables for one ID that were examined in this analysis was shown in Table 7. Note that the number of days from birth (U1) at each visit differ from those in Table 3 to Table 6. The numbers shown in Table 7 relate to data of an individual, while elsewhere they represent an aggregated measure overall children in the data set.

Table 7. Example of time variables in PEACHES data set

Visit	**Number of days from birth**	**Number of months from**								
		U1	**U2**	**U3**	**U4**	**U5**	**U6**	**U7**	**U7a**	**U8**
U1	0	0	0	0	0	0	0	0	0	0
U2	2	0.07	0	0	0	0	0	0	0	0
U3	30	0.99	0.92	0	0	0	0	0	0	0
U4	107	3.52	3.45	2.53	0	0	0	0	0	0
U5	213	7	6.93	6.01	3.48	0	0	0	0	0
U6	360	11.83	11.76	10.84	8.31	4.83	0	0	0	0
u7a	688	22.6	22.54	21.62	19.09	15.61	10.78	0	0	0
U7a	1,016	33.38	33.31	32.39	29.86	26.38	21.55	10.78	0	0
U8	1,456	47.84	47.77	46.85	44.32	40.84	36.01	25.23	14.46	0
U9	1,906	62.62	62.55	61.63	59.1	55.62	50.79	40.02	29.24	14.78

The primary purpose of this analysis focuses on time factor, thus no further mother's or child's covariables were incorporated in the model. More details about this analysis can be found in Appendix B: Supplementary results.

2.3 Main simulation study design

Results of the analysis described in 2.2.4[1] revealed that number of months after U1, U2, U3, U4, U6, and U7 were identified as variables that significantly influence the average growth curve of the children in PEACHES data set. These time variables served as input of the design of the simulation main and ICC studies. The subsections 2.3.1 and 2.3.2 will represent the model specification as well as the components that enter the main simulation design. The simulation steps will be shown in subsection 2.3.3.

2.3.1 Model specification

The generalized linear mixed model that underlies the simulation design is specified as follows:

$$LP_{i,j} = \alpha_i + \widehat{\beta_0} + \widehat{\beta_1} * t_{1_{i,j}} + \widehat{\beta_2} * t_{2_{i,j}} + \widehat{\beta_3} * t_{3_{i,j}} + \widehat{\beta_4} * t_{4_{i,j}} + \widehat{\beta_6} * t_{6_{i,j}} + \widehat{\beta_7} * t_{7_{i,j}}, \text{ where}$$

$$\alpha_i \sim N(0, \tau^2)$$

$LP_{i,j}$: $linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$ and $j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

The following logit link function was applied: $P[BMIz > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1+e^{LP_{i,j}}}$

Variance τ^2 as well as the coefficients of the fixed effects $\widehat{\beta_0}$, $\widehat{\beta_1}$, $\widehat{\beta_2}$, $\widehat{\beta_3}$, $\widehat{\beta_4}$, $\widehat{\beta_6}$, $\widehat{\beta_7}$ were estimated from the training dataset (see 2.2.3.2). The estimate of τ^2 is 3.551, which indicates an ICC of 3.551/(3.551+$\pi^2/3$) = 0.519. This estimate suggests that variances between individuals contributes to more than 50% of the overall variance and that it is necessary to use mixed models that incorporate random intercepts. More details about the output of the underlying model can be found in Appendix B2 (Appendix B: Supplementary results).

Table 8. Parameters used for simulation design

Parameter	Estimated value that entered the simulation model (SE)
τ^2	3.551
$\widehat{\beta_0}$	-4.65186 (0.51158)
$\widehat{\beta_1}$	0.10092 (2.94906)
$\widehat{\beta_2}$	0.56928 (3.06273)
$\widehat{\beta_3}$	-0.37908 (0.54987)
$\widehat{\beta_4}$	-0.19887 (0.21051)
$\widehat{\beta_6}$	-0.03126 (0.07224)
$\widehat{\beta_7}$	-0.07443 (0.03589)

[1] Detailed results of analysis about time effect on growth curve can be found in Appendix B: Supplementary results.

2.3.2 Underlying distributions of additive time covariates for the simulation

In order to imitate the variability of the number of days from birth (U1) at each visit, it is necessary to define an underlying distribution for each time covariate that enters the simulation. Uniform distributions are chosen to do so. The time window for each visit includes the earliest possible and the latest possible date suggested by the Bundeszentrale für gesundheitliche Aufklärung (BzgA) (Bundeszentrale für gesundheitliche Aufklärung, 2020) (see Table 9). These time windows served as parameters of the uniform distribution for each visit, respectively.

Table 9. Assumptions for time window of visits, recommended by BzgA

Visit	**Earliest recommended days after birth**	**Latest recommended days after birth**
U1	0	0
U2	3	10
U3	28	35
U4	90	120
U5	180	210
U6	300	360
U7	630	720
U7a	1,020	1,080
U8	1,380	1,440
U9	1,800	1,920

2.3.3 Calculation of simulated outcome

The presence of risk of overweight at a certain visit for a certain child was simulated by using the logit link function and the simulation model presented in 2.3.1, whereas the variance of estimated random intercept and estimates of fixed effects were obtained from fitting a GLMM with the training PEACHES data set (see Table 8). The simulated outcome is binary.

Simulation was done using the package `simstudy` (Version 0.2.1) (Wujciak-Jens, 2020). First of all, the estimated standard error τ for random intercepts extracted from the fitted GLMM (see Table 8) is defined (as `sigma.ri`, see Table 10). In the second step, the uniform distributions of the time distance between each visit and birth (U1) are defined (as `u1` to `u9`, see Table 10). Next, the normal distribution of the random intercept (as variable `RI`) is defined with mean = 0 and variance = $\mathrm{sigma.ri}^2$. With these components, the first part of the data is generated considering the correlation between data points within one individual. In this process, the random intercept and the number of days from birth (U1) of each individual are simulated simultaneously according to the defined distributions. In the next step, the additive time variables are calculated as shown

in an example in Table 7. The coefficients of the fixed effects $\widehat{\beta_0}$, $\widehat{\beta_1}$, $\widehat{\beta_2}$, $\widehat{\beta_3}$,$\widehat{\beta_4}$,$\widehat{\beta_6}$,$\widehat{\beta_7}$ are then extracted from the GLMM model (see Table 8) and serve as fixed components of the simulation design. In the last step, the outcome is simulated, which undergoes a binomial distribution with a logit link function. The linear predictor is calculated using the formula in 2.3.1. R codes for data generation that follow the steps described above are shown in Table 10. 1,000 IDs were simulated in the main simulation study.

Table 10. Simulation steps using package `simstudy` (Version 0.2.1)

R codes for data generation

```
simulate <- function(nid, seed, sigma.ri = NULL) {
  require(plyr)
  require(lme4)
  load("sim_bin/input/siminput.RData")

  # Extract from glmer model
  if (is.null(sigma.ri)) {
    sigma.ri <-
      as.data.frame(lme4::VarCorr(m0.bin, comp = "Std.Dev."))[1, "sdcor"]
  }



  def <-
    defData(varname = "u1",
            dist = "uniformInt",
            formula = "0;0")
  def <-
    defData(def,
            varname = "u2",
            dist = "uniformInt",
            formula = "3;10")
  def <-
    defData(def,
            varname = "u3",
            dist = "uniformInt",
            formula = "28;35")
  def <-
    defData(def,
            varname = "u4",
            dist = "uniformInt",
            formula = "90;120")
  def <-
    defData(def,
            varname = "u5",
            dist = "uniformInt",
            formula = "180;210")
  def <-
```

R codes for data generation

```
    defData(def,
            varname = "u6",
            dist = "uniformInt",
            formula = "300;360")
  def <-
    defData(def,
            varname = "u7",
            dist = "uniformInt",
            formula = "630;720")
  def <-
    defData(def,
            varname = "u7a",
            dist = "uniformInt",
            formula = "1020;1080")
  def <-
    defData(def,
            varname = "u8",
            dist = "uniformInt",
            formula = "1380;1440")
  def <-
    defData(def,
            varname = "u9",
            dist = "uniformInt",
            formula = "1800;1920")
  def <-
    defData(
      def,
      varname = "RI",
      formula = 0,
      dist = "normal",
      variance = sigma.ri ^ 2
    )


  #Reference:  https://www.kindergesundheit-info.de/themen/entwicklung/frueherkennung-
u1-u9-und-j1/untersuchungen-u1-bis-u9/
  set.seed(seed)
  utab <- genData(nid, def) %>%
    tidyr::gather(udata, udata_age_days, -id, -RI)


  udatatab <-
    plyr::join_all (
      list(
        ddply(utab, .(id), u1func) %>% gather(visit, months_from_u1, -id),
        ddply(utab, .(id), u2func) %>% gather(visit, months_from_u2, -id),
        ddply(utab, .(id), u3func) %>% gather(visit, months_from_u3, -id),
        ddply(utab, .(id), u4func) %>% gather(visit, months_from_u4, -id),
```

R codes for data generation

```
        ddply(utab, .(id), u5func) %>% gather(visit, months_from_u5, -id),
        ddply(utab, .(id), u6func) %>% gather(visit, months_from_u6, -id),
        ddply(utab, .(id), u7func) %>% gather(visit, months_from_u7, -id),
        ddply(utab, .(id), u7afunc) %>% gather(visit, months_from_u7a, -id),
        ddply(utab, .(id), u8func) %>% gather(visit, months_from_u8, -id),
        ddply(utab, .(id), u9func) %>% gather(visit, months_from_u9, -id)
      )
    )

  udatatab$udata <-
    factor(
      udatatab$visit,
      levels = c("V1", "V2", "V3", "V4", "V5", "V6", "V7", "V8", "V9", "V10"),
      labels = unique(utab$udata)
    )

  udatatab <- join(
    udatatab %>% select(-visit),
    utab %>% select(id, RI, udata, udata_age_days) %>% distinct()
  )

  udatatab$beta0 <- getME(m0.bin, "beta")[1]
  udatatab$beta1 <- getME(m0.bin, "beta")[2]
  udatatab$beta2 <- getME(m0.bin, "beta")[3]
  udatatab$beta3 <- getME(m0.bin, "beta")[4]
  udatatab$beta4 <- getME(m0.bin, "beta")[5]
  udatatab$beta6 <- getME(m0.bin, "beta")[6]
  udatatab$beta7 <- getME(m0.bin, "beta")[7]

  set.seed(seed)
  def2 <-
    defDataAdd(
      varname = "risk",
      dist = "binary",
      formula                                                                    =
"beta0+RI+beta1*months_from_u1+beta2*months_from_u2+beta3*months_from_u3+beta4*months_
from_u4+beta6*months_from_u6+beta7*months_from_u7",
      link = "logit"
    )
  return(addColumns(def2, data.table::as.data.table(udatatab)))


}
```

2.4 Simulation ICC study design

In order to examine the impact of the ICC, different τ^2 values were applied. For each of this value, a separate analysis was done, respectively. The results of all analyses were then compared graphically. The impact of ICC on the usefulness of dynamic approaches was then quantified applying linear mixed models. More about this point is presented in 2.10. These analyses are also referred to as "ICC scenarios" throughout the thesis.

2.4.1 Variance of random intercept τ^2

Considering a certain known or estimated variance of random intercepts, ICC of a GLMM can be calculated by assuming the theoretical variance of a GLMM = $\frac{\pi^2}{3}$ (Nakagawa et al., 2017), which gives: ICC = $\frac{\tau^2}{\tau^2+\frac{\pi^2}{3}}$.The following table presents different τ^2 values applied in the simulation ICC study and their corresponding ICC.

Table 11. Hypothetical τ^2values defined and their corresponding ICC

τ	τ^2	ICC
0.00	0.000	0.000
0.50	0.250	0.071
0.75	0.562	0.146
1.00	1.000	0.233
1.25	1.562	0.322
1.50	2.250	0.406
1.75	3.062	0.482
2.00	4.000	0.549
2.50	6.250	0.655
2.75	7.562	0.697
3.00	9.000	0.732
4.00	16.000	0.829
5.00	25.000	0.884
7.00	49.000	0.937
10.00	100.000	0.968

ICC measures the relativeness of variance explained by differences among observation units to all observed variance. In PEACHES study, these units are children included in the PEACHES analysis. In the simulation studies, these units are individuals in the simulation dataset(s). If the ICC is higher than 0.5, observed variance in the data is rather due to differences among individuals. If ICC is lower than 0.5, the observed variance can be rather explained by random errors within one individual over time, including measurement errors. In a binomial setting, this variance is assumed to be the theoretical variance, which is a fixed value $\frac{\pi^2}{3}$ (Nakagawa et al., 2017). In PEACHES dataset, an ICC around 0.5 was observed. This indicates comparable amount of contribution in explaining variability in the data due to between-individual vs. within-individual variances. With a wider range of ICC in this simulation study, other scenarios are examined. With an ICC close to 0, variability in data is explained mostly by random errors. With an ICC close to 1, rather variances between individuals contribute to the overall variability in data.

2.4.2 Model specification

The GLMM that underlies the simulation design is specified as follows:

$$LP_{i,j} = \alpha_i + \widehat{\beta_0} + \widehat{\beta_1} * t_{1_{i,j}} + \widehat{\beta_2} * t_{2_{i,j}} + \widehat{\beta_3} * t_{3_{i,j}} + \widehat{\beta_4} * t_{4_{i,j}} + \widehat{\beta_6} * t_{6_{i,j}} + \widehat{\beta_7} * t_{7_{i,j}}, \text{ where}$$

$$\alpha_i \sim N(0, \tau^2)$$

$LP_{i,j}$: $linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$ and $j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

The following logit link function was applied: $P[BMIz > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$

Variance τ^2 was a set of hypothetical values (see above in 2.4.1).

The coefficients of the fixed effects $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}, \widehat{\beta_4}, \widehat{\beta_6}, \widehat{\beta_7}$ were estimated from the training dataset (see 2.2.3.2). More details about the output of the underlying model can be found in Appendix B2 (Appendix B: Supplementary results).

Table 12. Parameters used for simulation design

Parameter	Estimated value that entered the simulation model (SE)
$\widehat{\beta_0}$	-4.65186 (0.51158)
$\widehat{\beta_1}$	0.10092 (2.94906)
$\widehat{\beta_2}$	0.56928 (3.06273)
$\widehat{\beta_3}$	-0.37908 (0.54987)
$\widehat{\beta_4}$	-0.19887 (0.21051)
$\widehat{\beta_6}$	-0.03126 (0.07224)
$\widehat{\beta_7}$	-0.07443 (0.03589)

2.4.3 Distributions of time variables

The time window for each visit includes the earliest possible and the latest possible date suggested by the Bundeszentrale für gesundheitliche Aufklärung (BzgA) (Bundeszentrale für gesundheitliche Aufklärung, 2020) (see Table 9). These time windows served as parameters of the uniform distribution for each visit, respectively.

2.4.4 Calculation of simulated outcome

The presence of risk of overweight at a certain visit for a certain child was simulated by using the logit link function and the simulation model presented in 2.4.2, whereas estimates of fixed effects obtained from fitting a GLMM with the training PEACHES data set and a set of hypothetical variances of random intercept (τ^2) were used. The simulated outcome is binary.

2.4.5 Sample size of the main simulation study

In order to reduce computing time, 100 IDs were simulated in each scenario of the simulation ICC study.

2.5 Prediction models

2.5.1 Overview of performed prediction models

In the main analyses, the following prediction models were specified, where linear and additive time effects from birth were included in the models:

- Bayesian dynamic model 1 (BDM1): individual random intercepts (RI), precision of RI distribution[1], and outcome are updated using outcome of the past visits and RIs from the last visits.
- Bayesian dynamic model 2 (BDM2): **fixed effects (FEs),** individual RIs, precision of RI distribution, and outcome are updated using outcome of the past visits and estimated RIs and FEs from the last visits.
- Bayesian static model 1 (BSM1): future outcome is updated using outcome of the past visits, other estimates are not updated. RIs are assumed to equal 0.
- Bayesian static model 2 (BSM2): prediction of outcome is made once at U1 using estimates from model fitted with cross-validation training data set. No updates of prediction over time. RIs are assumed to equal 0.
- Generalized linear mixed-effect regression model 1 (GLMER1): prediction is made once at U1 using estimates from model fitted with cross-validation training data set. No updates of prediction over time. RIs are assumed to equal 0.

It is to be expected that GLMER1 and BSM2 are the most static model and provide rigid prediction over time. These two models are supposed to provide similar results. Differences might be caused by different estimation methods and Monte-Carlo sampling. With greater sample size of the CV training dataset, the differences between predicted BDM2 and GLMER1 are supposed to become smaller. BDM2 is expected to provide the most flexible results.

The prediction performance was assessed by applying leave-one-out cross validation (LOOCV). After fitting a certain model with the training data set (with 999 children), validation done with the one child in the test data set to assess the prediction performance of the model. In total, 1,000 LOOCVs were done.

For BDM1 and BDM2, prediction was made for a child's outcome at each future visit separately, using all available accumulated data. This procedure results in maximal nine predictions for one child within one cross-validation. Specifically, accumulated data of all IDs at U1, U2, U3, U4, U5, U6, U7, U7a, and U8 were used to predict outcome of one child at U2, U3, U4, U5, U6, U7, U7a, U8, and U9, respectively. Whenever more data is available, BDM1 updated the random intercepts and the precision parameter, while BDM2 updates random intercepts, the precision parameter, and fixed exffects.

The GLMER1 and BSM2 models use the fixed effects that were estimated using the LOOCV-training data set. These estimates stay the same at all future time points. Random intercepts are

[1] Precision = $\frac{1}{\tau^2}$, where τ^2 is the variance of the distribution of random intercepts.

assumed to be 0. As a result, for these two models, outcome predicted using a certain model at a certain future time point are supposed to be the same, regardless of the accumulated data used. Like for GLMER1 and BSM2, for BSM1, random intercepts are assumed to be 0 and fixed effects are not updated over time. BSM1 differs from GLMER1 and BSM2 in a way that it accounts for available outcome of all available past visits and makes use of this information for prediction. As a result, outcome of the same future visit made at different prediction visits might show deviations. The specifications of the prediction models described above are summarized in the following table.

Table 13. Overview of the prediction models

Model	Fixed components over time	Components updated over time	Components used for updates	Bayesian approach	Dynamic approach
BDM1	• Fixed effects	• RI • precision of RI distribution • outcome	• Outcome of last visit • RI of last visit	Yes	Yes
BDM2	-	• RI • Precision of RI distribution • Outcome • Fixed effects	• Outcome from last visit • RI from last visit • Fixed effects estimated from last visit	Yes	Yes
BSM1	• Fixed effects estimated at U1 • RI = 0	• Outcome	• Outcome from last visit	Yes	No
BSM2	• Fixed effects estimated at U1 • RI = 0 • Outcome for all visits estimated at U1	-	-	Yes	No
GLMER1	• Fixed effects • RI = 0 • Outcome for all visits estimated at U1	-	-	No	No

2.5.2 Bayesian dynamic models

At the first visit, a child with ID *i* entered the study. His/her random intercept (RI) is assumed to be 0.

Risk for overweight of child *i* at U2, U3, U4, U5, U6, U7, U7a, U8, and U9 will then be predicted, respectively, using this RI (equal 0), RIs of children *i** from the trained data set, the precision of random intercept distribution, and the fixed effects estimated from the LOOCV-trained model.

In both models, the random intercepts of all children (*i** and i) and the precision parameter are updated at this step. Fixed effects are updated only in BMD2. At U1, both models BDM1 and BDM2 give the same predicted outcome. The following sections will describe the updating and predicting procedure at later prediction visits (U2 to U8) of BDM1 (2.5.2.1) and BMD2 (2.5.2.2) in details.

2.5.2.1 Bayesian dynamic model 1 (BDM1)

At U2, the BDM1 model contains the fixed effects estimated from the LOOCV-trained model, random intercepts of all children (*i** and i) estimated from the U1-BDM1 model as well as the precision parameter for the distribution of random intercepts ($\frac{1}{\tau^2}$). At this step, the random intercept of all children (*i** and i) are updated, using the precision parameter estimated from U1-BDM1 model. At the same time, the precision parameter is also updated. This updated model is called U2-BDM1 model.

At U3, the BDM1 model contains the fixed effects estimated from the LOOCV-trained model, random intercepts of all children (*i** and i) estimated from the U2-BDM1 model as well as the precision parameter for the distribution of random intercepts ($\frac{1}{\tau^2}$). At this step, the random intercept of all children (*i** and i) are updated, using the precision parameter estimated from U2-BDM1 model. At the same time, the precision parameter is also updated. This updated model is called U3-BDM1 model.

The same procedure applies for U4, U5, U6, U7, U7a, U8, and U9.

At U9, the BDM1 model contains the fixed effects estimated from the LOOCV-trained model, random intercepts of all children (*i** and i) estimated from the U8-BDM1 model as well as the precision parameter for the distribution of random intercepts ($\frac{1}{\tau^2}$). At this step, the random intercept of all children (*i** and i) are updated, using the precision parameter estimated from U8-BDM1 model. At the same time, the precision parameter is also updated. This updated model is called U9-BDM1 model, which is our final model. Theoretically, this final model could be used to predict outcome of another child(ren) that would enter the study in the future.

Table 14. Model specification of model BDM1

BDM1
Step 1 – Training with available IDs *i**
Prior: $\beta_0 \sim Normal(0, 10^{-5})$ $\beta_1 \sim Normal(0, 10^{-5})$

BDM1

$$\beta_2 \sim Normal(0, 10^{-5})$$

$$\beta_3 \sim Normal(0, 10^{-5})$$

$$\beta_4 \sim Normal(0, 10^{-5})$$

$$\beta_6 \sim N(0, 10^{-5})$$

$$\beta_7 \sim N(0, 10^{-5})$$

$$\frac{1}{\tau^2} \sim Gamma(10^{-4}, 10^{-4})$$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample.

Likelihood:

$$Y_{i^*,j^*} \sim Bernoulli\left(\left(P[zBMI > 1]_{i^*,j^*}\right)\right)$$

$$P[zBMI > 1]_{i^*,j^*} = \frac{e^{LP_{i^*,j^*}}}{1 + e^{LP_{i^*,j^*}}}$$

$$LP_{i^*,j^*} = \alpha_{i^*} + \beta_0 + \beta_1 * t_{1_{i^*,j^*}} + \cdots + \beta_7 * t_{7_{i^*,j^*}}$$

$$Random\ intercepts: \alpha_{i^*} \sim N(0, \tau^2)$$

$$LP_{i^*,j^*}: linear\ predictor\ of\ ID\ i^*\ at\ timepoint\ j^*$$

$$j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Estimated posterios:

Fixed effects: $\widehat{\beta_0}$, $\widehat{\beta_1}$, $\widehat{\beta_2}$, $\widehat{\beta_3}$, $\widehat{\beta_4}$, $\widehat{\beta_6}$, $\widehat{\beta_7}$

$\widehat{LP_{i^*,j^*}}$, *where* $j^* = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

Random intercepts: $\widehat{\alpha_{i^*}}$

Precision parameter: $\widehat{\frac{1}{\tau^2}}$

$\widehat{\frac{1}{\tau^2}}$ follows a Gamma distribution with parameters $\hat{r}$ and $\widehat{mu}$. These two parameters are calculated with the mean and variance of $\widehat{\frac{1}{\tau^2}}$ (MRC Biostatistics Unit, 2003).

BDM1

$$\widehat{mu} = \frac{mean\left(\widehat{\frac{1}{\tau^2}}\right)}{Var\left(\widehat{\frac{1}{\tau^2}}\right)}$$

$$\hat{r} = \frac{[mean\left(\widehat{\frac{1}{\tau^2}}\right)]^2}{Var\left(\widehat{\frac{1}{\tau^2}}\right)}$$

Step 2 – Updating and predicting at U1

Updated priors:

$\frac{1}{\tau_{U1}^2} \sim Gamma(\hat{r}, \widehat{mu})$

Random intercepts of children i from Step 1:* $\widehat{\alpha_{i^*}}$

Random intercepts of child i from the test sample: $\alpha_i = 0$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample, and

$$\begin{bmatrix} \mathrm{Y_{i,j}} \\ \mathrm{t_{1_{i,j}}} \\ \mathrm{t_{2_{i,j}}} \\ \mathrm{t_{3_{i,j}}} \\ \mathrm{t_{4_{i,j}}} \\ \mathrm{t_{6_{i,j}}} \\ \mathrm{t_{7_{i,j}}} \end{bmatrix},$$

where $j = \{U1\}$ and *i* is the child in the test sample.

Likelihood:

$$Y_{i,j} \sim Bernoulli((P[BMIz > 1]_{i,j}))$$

$$P[zBMI > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \alpha_i + \widehat{\beta_0} + \widehat{\beta_1} * t_{1_{i,j}} + \cdots + \widehat{\beta_7} * t_{7_{i,j}}$$

$$Random\ intercepts: \alpha_{i_{U1}} \sim N(0, {\tau_{U1}}^2),\ \alpha_{i_{U1}^*} \sim N(0, {\tau_{U1}}^2)$$

BDM1

$$LP_{i,j}: linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$$

$$j = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Posterios:

$$Linear\ predictor: \widehat{LP_{i,j}},\ where\ j = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Random intercepts: $\widehat{\alpha_{i_{U1}^*}}$, $\widehat{\alpha_{i_{U1}}}$

Precision parameter: $\widehat{\frac{1}{\tau_{U1}^2}}$

$\widehat{\frac{1}{\tau_{U1}^2}}$ follows a Gamma distribution with parameters $\widehat{r_{U1}}$ and $\widehat{mu_{U1}}$. These two parameters are calculated with the mean and variance of $\widehat{\frac{1}{\tau_{U1}^2}}$.

$$\widehat{mu_{U1}} = \frac{mean\left(\widehat{\frac{1}{\tau_{U1}^2}}\right)}{Var\left(\widehat{\frac{1}{\tau_{U1}^2}}\right)}$$

$$\widehat{r_{U1}} = \frac{[mean\left(\widehat{\frac{1}{\tau_{U1}^2}}\right)]^2}{Var\left(\widehat{\frac{1}{\tau_{U1}^2}}\right)}$$

Step 3 – Updating and predicting at U2

Updated priors:

$$\frac{1}{\tau_{U2}^2} \sim Gamma(\widehat{r_{U1}}, \widehat{mu_{U1}})$$

$$Random\ intercepts: \widehat{\alpha_{i_{U1}^*}}, \widehat{\alpha_{i_{U1}}}$$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample, and

BDM1

$$\begin{bmatrix} Y_{i,j} \\ t_{1_{i,j}} \\ t_{2_{i,j}} \\ t_{3_{i,j}} \\ t_{4_{i,j}} \\ t_{6_{i,j}} \\ t_{7_{i,j}} \end{bmatrix},$$

where $j = \{U1, U2\}$ and *i* is the child in the test sample

Likelihood:

$$Y_{i,j} \sim Bernoulli((P[zBMI > 1]_{i,j}))$$

$$P[zBMI > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \widehat{\alpha_{i_{U1}}} + \widehat{\beta_0} + \widehat{\beta_1} * t_{1_{i,j}} + \cdots + \widehat{\beta_7} * t_{7_{i,j}}$$

$$Random\ intercepts: \alpha_{i_{U2}} \sim N(0, {\tau_{U2}}^2)\ ,\ \alpha^*_{i_{U2}} \sim N(0, {\tau_{U2}}^2)$$

$$LP_{i,j}: linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$$

$$j = \{U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Posterio:

$\widehat{LP_{i,j}}$, *where* $j = \{U3, U4, U5, U6, U7, U7a, U8, U9\}$

Random intercepts: $\widehat{\alpha^*_{i_{U2}}}$, $\widehat{\alpha_{i_{U2}}}$

Precision parameter: $\widehat{\frac{1}{\tau^2_{U2}}}$.

$\widehat{\frac{1}{\tau^2_{U2}}}$ follows a Gamma distribution with parameters $\widehat{r_{U2}}$ and $\widehat{mu_{U2}}$. These two parameters are calculated with the mean and variance of $\widehat{\frac{1}{\tau^2_{U2}}}$

$$\widehat{mu_{U2}} = \frac{mean\left(\widehat{\frac{1}{\tau^2_{U2}}}\right)}{Var\left(\widehat{\frac{1}{\tau^2_{U2}}}\right)}$$

$$\widehat{r_{U2}} = \frac{[mean\left(\widehat{\frac{1}{\tau^2_{U2}}}\right)]^2}{Var\left(\widehat{\frac{1}{\tau^2_{U2}}}\right)}$$

Repeat Step 3 for the visit U3, U4, U5, U6, U7a, and U8.
Specifically, at U8, we have:

Updated priors:

BDM1

$$\frac{1}{\tau_{U8}^2} \sim Gamma(\widehat{r_{U7a}}, \widehat{mu_{U7a}})$$

$$Random\ intercepts: \widehat{\alpha_{i^*_{U7a}}}, \widehat{\alpha_{\iota_{U7a}}}$$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample, and

$$\begin{bmatrix} \mathrm{Y_{i,j}} \\ \mathrm{t_{1_{i,j}}} \\ \mathrm{t_{2_{i,j}}} \\ \mathrm{t_{3_{i,j}}} \\ \mathrm{t_{4_{i,j}}} \\ \mathrm{t_{6_{i,j}}} \\ \mathrm{t_{7_{i,j}}} \end{bmatrix},$$

where $j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\}$ and *i* is the child in the test sample

Likelihood:

$$Y_{i,j} \sim Bernoulli(\big(P[zBMI > 1]_{i,j}\big))$$

$$P[zBMI > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \widehat{\alpha_{\iota_{U7a}}} + \widehat{\beta_0} + \widehat{\beta_1} * t_{1_{i,j}} + \cdots + \widehat{\beta_7} * t_{7_{i,j}}$$

$$Random\ intercepts: \alpha_{i_{U8}} \sim N\ (0, {\tau_{U8}}^2)\ ,\ \alpha_{i^*_{U8}} \sim N\ (0, {\tau_{U8}}^2)$$

$$LP_{i,j}: linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$$

$$j = \{U9\}$$

Posterios:

$\widehat{LP_{i,j}}$, *where* $j = \{U9\}$

Random intercepts: $\widehat{\alpha_{i^*_{U8}}}$, $\widehat{\alpha_{\iota_{U8}}}$

Precision parameter: $\widehat{\frac{1}{\tau_{U8}^2}}$.

BDM1
$\widehat{\frac{1}{\tau_{U8}^2}}$ follows a Gamma distribution with parameters $\widehat{r_{U8}}$ and $\widehat{mu_{U8}}$. These two parameters are calculated with the mean and variance of $\widehat{\frac{1}{\tau_{U8}^2}}$. $$\widehat{mu_{U8}} = \frac{mean\left(\widehat{\frac{1}{\tau_{U8}^2}}\right)}{Var\left(\widehat{\frac{1}{\tau_{U8}^2}}\right)}$$ $$\widehat{r_{U8}} = \frac{[mean\left(\widehat{\frac{1}{\tau_{U8}^2}}\right)]^2}{Var\left(\widehat{\frac{1}{\tau_{U8}^2}}\right)}$$ These posterios estimated at U9 can be considered to predict future outcome later than U9 of this child.

2.5.2.2 Bayesian dynamic model 2 (BDM2)

At U2, BDM2 contains the updated RIs of all children (*i** and i) estimated with the U1-BDM2 model. Updated estimates for fixed effects and precision parameter also come from the U1-BDM2 model. At this step, the random intercept of all children (*i** and i) are once again updated using the precision parameter. At the same time, the precision parameter and the fixed effects are updated. This updated model is called U2-BDM2 model.

At U3, BDM2 contains the updated RIs of all children (*i** and i) estimated with the U2-BDM2 model. Updated estimates for fixed effects and precision parameter also come from the U2-BDM2 model. At this step, the random intercept of all children (*i** and i) are once again updated using the precision parameter. At the same time, the precision parameter and the fixed effects are updated. This updated model is called U3-BDM2 model.

The same procedure applies for U4, U5, U6, U7, U7a, U8, and U9.

At U9, BDM2 contains the updated RIs of all children (*i** and i) estimated with the U8-BDM2 model. Updated estimates for fixed effects and precision parameter also come from the U8-BDM2 model. At this step, the random intercept of all children (*i** and i) are once again updated using the precision parameter. At the same time, the precision parameter and the fixed effects are updated. This updated model is called U9-BDM2 model, which is our final model. Theoretically, this model would be used to predict outcome of another child(ren) that entered the study in the future.

Table 15. Model specification of model BDM2

BDM2
Step 1 – Training with available IDs *i** *Prior:* $$\beta_0 \sim Normal(0, 10^{-5})$$

BDM2

$$\beta_1 \sim Normal(0, 10^{-5})$$

$$\beta_2 \sim Normal(0, 10^{-5})$$

$$\beta_3 \sim Normal(0, 10^{-5})$$

$$\beta_4 \sim Normal(0, 10^{-5})$$
$$\beta_6 \sim Normal(0, 10^{-5})$$

$$\beta_7 \sim Normal(0, 10^{-5})$$

$$\frac{1}{\tau^2} \sim Gamma(10^{-4}, 10^{-4})$$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample.

Likelihood:

$$Y_{i^*,j^*} \sim Bernoulli\left(\left(P[zBMI > 1]_{i^*,j^*}\right)\right)$$
$$P[zBMI > 1]_{i^*,j^*} = \frac{e^{LP_{i^*,j^*}}}{1 + e^{LP_{i^*,j^*}}}$$

$$LP_{i^*,j^*} = \alpha_{i^*} + \beta_0 + \beta_1 * t_{1_{i^*,j^*}} + \cdots + \beta_7 * t_{7_{i^*,j^*}}$$

$$Random\ intercepts: \alpha_{i^*} \sim N(0, \tau^2)$$

$$LP_{i^*,j^*}: linear\ predictor\ of\ ID\ i^*\ at\ timepoint\ j^*$$

$$j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Estimated posterios:

Fixed effects: $\widehat{\beta_0}$, $\widehat{\beta_1}$, $\widehat{\beta_2}$, $\widehat{\beta_3}$, $\widehat{\beta_4}$, $\widehat{\beta_6}$, $\widehat{\beta_7}$

$\widehat{LP_{i^*,j^*}}$, *where* $j^* = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

Random intercepts: $\widehat{\alpha_{i^*}}$

Precision parameter: $\widehat{\frac{1}{\tau^2}}$

BDM2

$\widehat{\frac{1}{\tau^2}}$ follows a Gamma distribution with parameters $\hat{r}$ and $\widehat{mu}$. These two parameters are calculated with the mean and variance of $\widehat{\frac{1}{\tau^2}}$.

$$\widehat{mu} = \frac{mean\left(\widehat{\frac{1}{\tau^2}}\right)}{Var\left(\widehat{\frac{1}{\tau^2}}\right)}$$

$$\hat{r} = \frac{[mean\left(\widehat{\frac{1}{\tau^2}}\right)]^2}{Var\left(\widehat{\frac{1}{\tau^2}}\right)}$$

Step 2 – Updating and predicting at U1

Updated priors:

$\frac{1}{\tau^2_{U1}} \sim Gamma(\hat{r}, \widehat{mu})$

Random intercepts of children i from Step 1:* $\widehat{\alpha_{i^*}}$

Random intercept of child i form the test sample: $\alpha_i = 0$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample, and

$$\begin{bmatrix} \mathrm{Y_{i,j}} \\ \mathrm{t_{1_{i,j}}} \\ \mathrm{t_{2_{i,j}}} \\ \mathrm{t_{3_{i,j}}} \\ \mathrm{t_{4_{i,j}}} \\ \mathrm{t_{6_{i,j}}} \\ \mathrm{t_{7_{i,j}}} \end{bmatrix},$$

where $j = \{U1\}$ and *i* is the child in the test sample.

Likelihood:

$$Y_{i,j} \sim Bernoulli(\left(P[zBMI > 1]_{i,j}\right))$$

$$P[zBMIz > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

BDM2

$$LP_{i,j} = \alpha_i + \beta_0 + \beta_1 * t1_{i,j} + \cdots + \beta_7 * t7_{i,j}$$

$$Random\ intercepts: \alpha_{i_{U1}} \sim N(0, {\tau_{U1}}^2)\ ,\ \alpha_{i^*_{U1}} \sim N(0, {\tau_{U1}}^2)$$

$$LP_{i,j}: linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$$

$$j = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Posterio:

$Linear\ predictor: \widehat{LP_{i,j}}$, *where* $j = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

Random intercetps: $\widehat{\alpha_{i^*_{U1}}}$, $\widehat{\alpha_{i_{U1}}}$

Precision parameter: $\widehat{\frac{1}{\tau^2_{U1}}}$

$\widehat{\frac{1}{\tau^2_{U1}}}$ follows a Gamma distribution with parameters $\widehat{r_{U1}}$ and $\widehat{mu_{U1}}$. These two parameters are calculated with the mean and variance of $\widehat{\frac{1}{\tau^2_{U1}}}$.

$$\widehat{mu_{U1}} = \frac{mean\left(\widehat{\frac{1}{\tau^2_{U1}}}\right)}{Var\left(\widehat{\frac{1}{\tau^2_{U1}}}\right)}$$

$$\widehat{r_{U1}} = \frac{[mean\left(\widehat{\frac{1}{\tau^2_{U1}}}\right)]^2}{Var\left(\widehat{\tau_{U1}\frac{1}{\tau^2_{U1}}}\right)}$$

Fixed effects: $\widehat{\beta_{0_{U1}}}, \widehat{\beta_{1_{U1}}}, \widehat{\beta_{2_{U1}}}, \widehat{\beta_{3_{U1}}}, \widehat{\beta_{4_{U1}}}, \widehat{\beta_{6_{U1}}}, \widehat{\beta_{7_{U1}}}$

Step 3 – Updating and predicting at U2

Updated priors:

$$\frac{1}{\tau^2_{U2}} \sim Gamma(\widehat{r_{U1}}, \widehat{mu_{U1}})$$

$$Random\ intercepts: \widehat{\alpha_{i^*_{U1}}}, \widehat{\alpha_{i_{U1}}}$$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t1_{i^*,j^*} \\ t2_{i^*,j^*} \\ t3_{i^*,j^*} \\ t4_{i^*,j^*} \\ t6_{i^*,j^*} \\ t7_{i^*,j^*} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample, and

BDM2

$$\begin{bmatrix} Y_{i,j} \\ t_{1_{i,j}} \\ t_{2_{i,j}} \\ t_{3_{i,j}} \\ t_{4_{i,j}} \\ t_{6_{i,j}} \\ t_{7_{i,j}} \end{bmatrix},$$

where $j = \{U1, U2\}$ and *i* is the child in the test sample

Likelihood:

$$Y_{i,j} \sim Bernoulli((P[BMIz > 1]_{i,j}))$$

$$P[BMIz > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \widehat{\alpha_{i_{U1}}} + \widehat{\beta_{0_{U1}}} + \widehat{\beta_{1_{U1}}} * t_{1_{i,j}} + \cdots + \widehat{\beta_{7_{U1}}} * t_{7_{i,j}}$$

$$Random\ intercepts: \alpha_{i_{U2}} \sim N(0, {\tau_{U2}}^2)\ ,\ \alpha_{i^*_{U2}} \sim N(0, {\tau_{U2}}^2)$$

$$LP_{i,j}: linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$$

$$j = \{U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Posterio:

$\widehat{LP_{i,j}}$, *where* $j = \{U3, U4, U5, U6, U7, U7a, U8, U9\}$

Random intercepts: $\widehat{\alpha_{i^*_{U2}}}$, $\widehat{\alpha_{i_{U2}}}$

Precision parameter: $\widehat{\frac{1}{\tau^2_{U2}}}$.

$\widehat{\frac{1}{\tau^2_{U2}}}$ follows a Gamma distribution with parameters $\widehat{r_{U2}}$ and $\widehat{mu_{U2}}$. These two parameters are calculated with the mean and variance of $\widehat{\frac{1}{\tau^2_{U2}}}$.

$$\widehat{mu_{U2}} = \frac{mean\left(\widehat{\frac{1}{\tau^2_{U2}}}\right)}{Var\left(\widehat{\frac{1}{\tau^2_{U2}}}\right)}$$

$$\widehat{r_{U2}} = \frac{[mean\left(\widehat{\frac{1}{\tau^2_{U2}}}\right)]^2}{Var\left(\widehat{\frac{1}{\tau^2_{U2}}}\right)}$$

Fixed effects: $\widehat{\beta_{0_{U2}}}, \widehat{\beta_{1_{U2}}}, \widehat{\beta_{2_{U2}}}, \widehat{\beta_{3_{U2}}}, \widehat{\beta_{4_{U2}}}, \widehat{\beta_{6_{U2}}}, \widehat{\beta_{7_{U2}}}$

Repeat Step 3 for the visit U3, U4, U5, U6, U7a, and U8.
Specifically, at U8, we have:

BDM2

Updated priors:

$$\frac{1}{\tau_{U8}^2} \sim Gamma(\widehat{r_{U7a}}, \widehat{mu_{U7a}})$$

$$Random\ intercepts: \widehat{\alpha_{i^*_{U7a}}}, \widehat{\alpha_{\iota_{U7a}}}$$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample, and

$$\begin{bmatrix} \mathrm{Y_{i,j}} \\ \mathrm{t_{1_{i,j}}} \\ \mathrm{t_{2_{i,j}}} \\ \mathrm{t_{3_{i,j}}} \\ \mathrm{t_{4_{i,j}}} \\ \mathrm{t_{6_{i,j}}} \\ \mathrm{t_{7_{i,j}}} \end{bmatrix},$$

where $j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\}$ and *i* is the child in the test sample

Likelihood:

$$Y_{i,j} \sim Bernoulli((P[zBMI > 1]_{i,j}))$$

$$P[zBMI > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \widehat{\alpha_{\iota_{U7a}}} + \widehat{\beta_{0_{U7a}}} + \widehat{\beta_{1_{U7a}}} * t_{1_{i,j}} + \cdots + \widehat{\beta_{7_{U7a}}} * t_{7_{i,j}}$$

$$Random\ intercepts: \alpha_{i_{U8}} \sim N(0, {\tau_{U8}}^2),\ \alpha_{i^*_{U8}} \sim N(0, {\tau_{U8}}^2)$$

$$LP_{i,j}: linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$$

$$j = \{U9\}$$

Posterios:

$\widehat{LP_{i,j}}$, *where* $j = \{U9\}$

Random intercepts: $\widehat{\alpha_{i^*_{U8}}}$, $\widehat{\alpha_{\iota_{U8}}}$

Precision parameter: $\widehat{\frac{1}{\tau_{U8}^2}}$.

BDM2

$\widehat{\frac{1}{\tau^2_{U8}}}$ follows a Gamma distribution with parameters $\widehat{r_{U8}}$ and $\widehat{mu_{U8}}$. These two parameters are calculated with the mean and variance of $\widehat{\frac{1}{\tau^2_{U8}}}$.

$$\widehat{mu_{U8}} = \frac{mean\left(\widehat{\frac{1}{\tau^2_{U8}}}\right)}{Var\left(\widehat{\frac{1}{\tau^2_{U8}}}\right)}$$

$$\widehat{r_{U8}} = \frac{mean\left(\widehat{\frac{1}{\tau^2_{U8}}}\right)^2}{Var\left(\widehat{\frac{1}{\tau^2_{U8}}}\right)}$$

Fixed effects: $\widehat{\beta_{0_{U8}}}, \widehat{\beta_{1_{U8}}}, \widehat{\beta_{2_{U8}}}, \widehat{\beta_{3_{U8}}}, \widehat{\beta_{4_{U8}}}, \widehat{\beta_{6_{U8}}}, \widehat{\beta_{7_{U8}}}$

These posterios estimated at U9 can be considered to predict future outcome later than U9 of this child.

2.5.3 Bayesian static models

2.5.3.1 Bayesian static model 1 (BSM1)

At U1, child with ID *i* entered the study. His/her random intercept is assumed to be 0.

Risk for overweight of child *i* at U2, U3, U4, U5, U6, U7, U7a, U8, and U9 will then be predicted, respectively, using this assumed random intercept (equal 0) and the fixed effects as well as the RIs of children *i** estimated from the LOOCV-trained model. This is the initial model for BSM1 and BSM2.

At U2, outcome of child *i* at U1 and U2 is known and will be used as new data for the prediction of future outcome (at U3, U4, U5, U6, U7, U7a, U8, and U9). Random intercept of child *i* is assumed to be 0. Fixed effects as well as the RIs of children *i** estimated from the initial model are used for prediction and these estimates are not updated.

At U3, outcome of child *i* at U1, U2, and U3 is known and will be used as new data for the prediction of future outcome (at U4, U5, U6, U7, U7a, U8, and U9). Random intercept of child *i* is assumed to be 0. Fixed effects as well as the RIs of children *i** estimated from the initial model are used for prediction and these estimates are not updated.

The same procedure applies for U4, U5, U6, U7, U7a, and U8.

At U8, outcome of child *i* at U1, U2, U3, U4, U5, U6, U7, U7a, and U8 is known and will be used as new data for the prediction of future outcome (at U9). Random intercept of child *i* is assumed to be 0. Fixed effects as well as the RIs of children *i** estimated from the initial model are used for prediction and these estimates are not updated.

Theoretically, the same procedure can be done at U9 to predict future outcome.

Table 16. Model specification of model BSM1

BSM1
Step 1 – Training with available IDs *i**
Priors: $\beta_0 \sim Normal(0, 10^{-5})$ $\beta_1 \sim Normal(0, 10^{-5})$ $\beta_2 \sim Normal(0, 10^{-5})$ $\beta_3 \sim Normal(0, 10^{-5})$ $\beta_4 \sim Normal(0, 10^{-5})$ $\beta_6 \sim N(0, 10^{-5})$ $\beta_7 \sim N(0, 10^{-5})$ $\frac{1}{\tau^2} \sim Gamma(10^{-4}, 10^{-4})$
Data:

BSM1

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample.

Likelihood:

$$Y_{i^*,j^*} \sim Bernoulli\left(\left(P[BMIz > 1]_{i^*,j^*}\right)\right)$$

$$P[BMIz > 1]_{i^*,j^*} = \frac{e^{LP_{i^*,j^*}}}{1 + e^{LP_{i^*,j^*}}}$$

$$LP_{i^*,j^*} = \alpha_{i^*} + \beta_0 + \beta_1 * t_{1_{i^*,j^*}} + \cdots + \beta_7 * t_{7_{i^*,j^*}}$$

$$Random\ intercepts: \alpha_{i^*} \sim N(0, \tau^2)$$

$$LP_{i^*,j^*}: linear\ predictor\ of\ ID\ i^*\ at\ timepoint\ j^*$$

$$j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Estimated posterios:

Fixed effects: $\widehat{\beta_0}$, $\widehat{\beta_1}$, $\widehat{\beta_2}$, $\widehat{\beta_3}$, $\widehat{\beta_4}$, $\widehat{\beta_6}$, $\widehat{\beta_7}$

$\widehat{LP_{i^*,j^*}}$, *where* $j^* = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

Random intercepts: $\widehat{\alpha_{i^*}}$

Precision parameter: $\widehat{\frac{1}{\tau^2}}$

$\widehat{\frac{1}{\tau^2}}$ follows a Gamma distribution with parameters $\hat{r}$ and $\widehat{mu}$. These two parameters are calculated with the mean and variance of $\widehat{\frac{1}{\tau^2}}$.

$$\widehat{mu} = \frac{mean\left(\widehat{\frac{1}{\tau^2}}\right)}{Var\left(\widehat{\frac{1}{\tau^2}}\right)}$$

$$\hat{r} = \frac{[mean\left(\widehat{\frac{1}{\tau^2}}\right)]^2}{Var\left(\widehat{\frac{1}{\tau^2}}\right)}$$

Step 2 – Predicting at U1

Prior:

Random intercept of child i form the test sample: $\alpha_i = 0$

BSM1

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample,

random intercepts of children i from Step 1:* $\widehat{\alpha_{i^*}}$, and

$$\begin{bmatrix} \mathrm{Y_{i,j}} \\ \mathrm{t_{1_{i,j}}} \\ \mathrm{t_{2_{i,j}}} \\ \mathrm{t_{3_{i,j}}} \\ \mathrm{t_{4_{i,j}}} \\ \mathrm{t_{6_{i,j}}} \\ \mathrm{t_{7_{i,j}}} \end{bmatrix},$$

where $j = \{U1\}$ and *i* is the child in the test sample

Likelihood:

$$Y_{i,j} \sim Bernoulli\left(\left(P[BMIz > 1]_{i,j}\right)\right)$$

$$P[BMIz > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \alpha_i + \widehat{\beta_0} + \widehat{\beta_1} * t_{1_{i,j}} + \cdots + \widehat{\beta_7} * t_{7_{i,j}}$$

$$LP_{i,j}: linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$$

$$j = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Posterio:

$$Linear\ predictor: \widehat{LP_{i,j}},\ where\ j = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Step 3 – Predicting at U2

Prior:

Random intercepts of child i from the test sample: $\alpha_i = 0$

Data:

BSM1

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample,

random intercepts of IDs from Step 1: $\widehat{\alpha_{i^*}}$, and

$$\begin{bmatrix} \mathrm{Y_{i,j}} \\ \mathrm{t_{1_{i,j}}} \\ \mathrm{t_{2_{i,j}}} \\ \mathrm{t_{3_{i,j}}} \\ \mathrm{t_{4_{i,j}}} \\ \mathrm{t_{6_{i,j}}} \\ \mathrm{t_{7_{i,j}}} \end{bmatrix},$$

where $j = \{U1, U2\}$ and *i* is the child in the test sample

Likelihood:

$$Y_{i,j} \sim Bernoulli\left(\left(P[BMIz > 1]_{i,j}\right)\right)$$

$$P[BMIz > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \alpha_i + \widehat{\beta_0} + \widehat{\beta_1} * t_{1_{i,j}} + \cdots + \widehat{\beta_7} * t_{7_{i,j}}$$

$$LP_{i,j}: linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$$

$$j = \{U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Posterio:

$$\widehat{LP_{i,j}},\ where\ j = \{U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Repeat Step 3 for the visit U3, U4, U5, U6, U7a, and U8.
At U8, we have:

Prior:

Random intercepts of child i from the test sample: $\alpha_i = 0$

Data:

BSM1

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample,

random intercepts of IDs from the LOOCV training sample: $\widehat{\alpha_{i^*}}$, and

$$\begin{bmatrix} \mathrm{Y_{i,j}} \\ \mathrm{t_{1_{i,j}}} \\ \mathrm{t_{2_{i,j}}} \\ \mathrm{t_{3_{i,j}}} \\ \mathrm{t_{4_{i,j}}} \\ \mathrm{t_{6_{i,j}}} \\ \mathrm{t_{7_{i,j}}} \end{bmatrix},$$

where $j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\}$ and *i* is the child in the test sample

Likelihood:

$$Y_{i,j} \sim Bernoulli\left(\left(P[zBMI > 1]_{i,j}\right)\right)$$

$$P[zBMI > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \alpha_i + \widehat{\beta_0} + \widehat{\beta_1} * t_{1_{i,j}} + \cdots + \widehat{\beta_7} * t_{7_{i,j}}$$

$$LP_{i,j}: linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$$

$$j = \{U9\}$$

Posterio:

$\widehat{LP_{i,j}}$, *where* $j = \{U9\}$

2.5.3.2 Bayesian static model 2 (BSM2)

Prediction is made once at U1 using estimates from model fitted with LOOCV-training data set. No updates of prediction over time were made.

Table 17. Model specification of model BSM2

BSM2

Step 1 – Training with available IDs *i**

Prior:

BSM2

$$\beta_0 \sim Normal(0, 10^{-5})$$

$$\beta_1 \sim Normal(0, 10^{-5})$$

$$\beta_2 \sim Normal(0, 10^{-5})$$

$$\beta_3 \sim Normal(0, 10^{-5})$$

$$\beta_4 \sim Normal(0, 10^{-5})$$

$$\beta_6 \sim N(0, 10^{-5})$$

$$\beta_7 \sim N(0, 10^{-5})$$

$$\frac{1}{\tau^2} \sim Gamma(10^{-4}, 10^{-4})$$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample.

Likelihood:

$$Y_{i^*,j^*} \sim Bernoulli((P[zBMI > 1]_{i^*,j^*}))$$

$$P[zBMI > 1]_{i^*,j^*} = \frac{e^{LP_{i^*,j^*}}}{1 + e^{LP_{i^*,j^*}}}$$

$$LP_{i^*,j^*} = \alpha_{i^*} + \beta_0 + \beta_1 * t_{1_{i^*,j^*}} + \cdots + \beta_7 * t_{7_{i^*,j^*}}$$

$$Random\ intercepts: \alpha_{i^*} \sim N(0, \tau^2)$$

$$LP_{i^*,j^*}: linear\ predictor\ of\ ID\ i^*\ at\ timepoint\ j^*$$

$$j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Estimated posterios:

Fixed effects: $\widehat{\beta_0}$, $\widehat{\beta_1}$, $\widehat{\beta_2}$, $\widehat{\beta_3}$, $\widehat{\beta_4}$, $\widehat{\beta_6}$, $\widehat{\beta_7}$

$\widehat{LP_{i^*,j*}}$, *where* $j^* = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

Random intercepts: $\widehat{\alpha_{i^*}}$

Precision parameter: $\hat{\tau}$

$\widehat{\frac{1}{\tau^2}}$ follows a Gamma distribution with parameters $\hat{r}$ and $\widehat{mu}$. These two parameters are calculated with the mean and variance of $\widehat{\frac{1}{\tau^2}}$.

$$\widehat{mu} = \frac{mean\left(\widehat{\frac{1}{\tau^2}}\right)}{Var\left(\widehat{\frac{1}{\tau^2}}\right)}$$

$$\hat{r} = \frac{[mean\left(\widehat{\frac{1}{\tau^2}}\right)]^2}{Var\left(\widehat{\frac{1}{\tau^2}}\right)}$$

Step 2 – Predicting at U1

Prior:

Random intercepts of child i from the test sample: $\alpha_i = 0$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample,

random intercepts of IDs from the LOOCV training sample: $\widehat{\alpha_{i^*}}$*, and*

$$\begin{bmatrix} t_{1_{i,j}} \\ t_{2_{i,j}} \\ t_{3_{i,j}} \\ t_{4_{i,j}} \\ t_{6_{i,j}} \\ t_{7_{i,j}} \end{bmatrix},$$

where $j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i* is the child in the test sample

Likelihood:

$$Y_{i,j} \sim Bernoulli\left(\left(P[zBMI > 1]_{i,j}\right)\right)$$

$$P[zBMI > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \alpha_i + \widehat{\beta_0} + \widehat{\beta_1} * t_{1_{i,j}} + \cdots + \widehat{\beta_7} * t_{7_{i,j}}$$

$$LP_{i,j}: linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$$

BSM2
$j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ *Posterios:* $Linear\ predictor$: $\widehat{LP_{i,j}}$, *where* $j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ **At visit U2, U3, U4, U5, U6, U7a, and U8, no predictions are made but prediction results from visit U1 are adopted.**

2.5.4 Generalized linear mixed regression model 1 (GLMER1)

Generalized linear mixed regression model with linear and additive time effects from birth was examined. Prediction is made once at U1 using estimates from model fitted with LOOCV-training data set. No updates of prediction over time were made. At visit U2, U3, U4, U5, U6, U7a, and U8, no predictions are made but prediction results from visit U1 are adopted.

Table 18. Model specification of model GLMER1

GLMER1
Step 1 – Training with available IDs *i**
Data: $\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix}$, where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample. *Model specification:* $P[zBMI > 1]_{i^*,j^*} = \frac{e^{LP_{i^*,j^*}}}{1 + e^{LP_{i^*,j^*}}}$ *Linear predictor:* $LP_{i^*,j^*} = \alpha_{i^*} + \beta_0 + \beta_1 * t_{1_{i^*,j^*}} + \cdots + \beta_7 * t_{7_{i^*,j^*}}$ $Random\ intercepts$: $\alpha_{i^*} \sim N(0, \tau^2)$ LP_{i^*,j^*}: $linear\ predictor\ of\ ID\ i^*\ at\ timepoint\ j^*$ $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

GLMER1

Model estimates:

Fixed effects: $\widehat{\beta_0}$, $\widehat{\beta_1}$, $\widehat{\beta_2}$, $\widehat{\beta_3}$, $\widehat{\beta_4}$, $\widehat{\beta_6}$, $\widehat{\beta_7}$

$\widehat{LP_{i^*,j^*}}$, *where* $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

Random intercepts: $\widehat{\alpha_{i^*}}$

Step 2 – Predicting at U1

Assumption:

Random intercepts of child i from the test sample: $\alpha_i = 0$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample,

random intercepts of IDs from the LOOCV training sample: $\widehat{\alpha_{i^*}}$, and

$$\begin{bmatrix} t_{1_{i,j}} \\ t_{2_{i,j}} \\ t_{3_{i,j}} \\ t_{4_{i,j}} \\ t_{6_{i,j}} \\ t_{7_{i,j}} \end{bmatrix},$$

where $j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i* is the child in the test sample

Prediction model:

$$P[zBMI > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \alpha_i + \widehat{\beta_0} + \widehat{\beta_1} * t_{1_{i,j}} + \cdots + \widehat{\beta_7} * t_{7_{i,j}}$$

$$j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Predicted values:

$\widehat{LP_{i,j}}$, *where* $j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

At visit U2, U3, U4, U5, U6, U7a, and U8, no predictions are made but prediction results from visit U1 are adopted.

2.6 Applied measures

2.6.1 Measures comparing prediction performance at the individual level

2.6.1.1 Prediction error at each time point (for each ID)

For a certain ID, Prediction Error (PE) was calculated at each timepoint for each model, respectively.

$$PE_{i,j,k,m} = \left(P[zBMI > 1]_{i,j,k,m} - I[zBMI_{observed} > 1]\right)^2$$

$$i = \{1,2,3,\dots,n\}, where\ n\ is\ the\ total\ number\ of\ individuals\ in\ the\ dataset$$

$$prediction\ visit\ j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\};$$

$$future\ visit\ k = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\};\ j < k;\ \text{and}$$

$$m = \{GLMER1, BSM1, BSM2, BDM1, BDM2\}$$

2.6.1.2 Relative prediction error at each time point (for each ID), compared with GLMER1

For a certain ID, Relative Prediction Error (RPE) between a model versus GLMER1 was calculated at each timepoint, respectively.

$$RPE_{i,j,k,m}^{GLMER1} = \frac{PE_{i,j,k,GLMER1}}{PE_{i,j,k,m}}$$

RPE on the natural logarithmic scale was also calculated, this measure was used in the regression models quantifying the association between RPE and the amount of information available from past months.

$$LRPE_{i,j,k,m}^{GLMER1} = ln\left(RPE_{i,j,k,m}\right) = ln\left(\frac{PE_{i,j,k,GLMER1}}{PE_{i,j,k,m}}\right)$$

$$i = \{1,2,3,\dots,n\}, where\ n\ is\ the\ total\ number\ of\ individuals\ in\ the\ dataset$$

$$prediction\ visit\ j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\};$$

$$future\ visit\ k = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\};\ j < k;\ \text{and}$$

$$m = \{BSM1, BSM2, BDM1, BDM2\}$$

A LMM including fractional polynomial to the 4th grade[1] of time effect was selected[2] and fitted to quantify the LRPE over time. LRPE < 0 indicates better prediction performance of GLMER1 than BSM1 at a certain visit. A LRPE > 0 indicates that on average, at a certain visit, GLMER1 performs better. Linear time effect (in months) was also entered into the model.

The full LMM was specified as follows:

$$LRPE_{i,j,k,m}^{GLMER1} = \beta_0 + \beta_1 * t_{1_{i,j,k,m}} + \beta_2 * {t_{1_{i,j,k,m}}}^2 + \beta_3 * {t_{1_{i,j,k,m}}}^3 + \beta_4 * {t_{1_{i,j,k,m}}}^4 + \varepsilon_{i,j,k,m}$$

[1] Fractional polynomial was applied to capture the non-linearity of the time effect, if any exists.

[2] Model selection was done with backward selection using AIC. The function `step()` in package `lmerTest` KUZNETSOVA, A., BROCKHOFF, P. B. & CHRISTENSEN, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *2017,* 82**,** 26. was used.

$$i = \{1,2,3,\dots,n\}, where\ n\ is\ the\ total\ number\ of\ individuals\ in\ the\ dataset$$

$$prediction\ visit\ j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\};$$

$$future\ visit\ k = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\};\ j < k;$$

$m = \{BSM1, BSM2, BDM1, BDM2\}$, and $t_{1_{i,j,k,m}}$: $number\ of\ months\ from\ U1$

2.6.1.3 Relative prediction error at each time point (for each ID), compared with GLMER4

For a certain ID, Relative Prediction Error between a model versus GLMER4 was calculated at each timepoint, respectively.

$$RPE_{i,j,k,m}^{GLMER4} = \frac{PE_{i,j,k,GLMER4}}{PE_{i,j,k,m}}$$

RPE on the natural logarithmic scale was also calculated, this measure was used in the regression models quantifying the association between RPE and the amount of information aka number of past visits.

$$LRPE_{i,j,k,m}^{GLMER4} = \ln\left(RPE_{i,j,k,m}\right) = ln\left(\frac{PE_{i,j,k,GLMER4}}{PE_{i,j,k,m}}\right)$$

$$i = \{1,2,3,\dots,n\}, where\ n\ is\ the\ total\ number\ of\ individuals\ in\ the\ dataset$$

$$prediction\ visit\ j = \{U4, U5, U6, U7, U7a, U8\};$$

$$future\ visit\ k = \{U5, U6, U7, U7a, U8, U9\};\ j < k;\ \text{and}$$

$$m = \{BSM1, BSM2, BDM1, BDM2\}$$

2.6.2 Measures comparing prediction performance at the aggregated level

For each timepoint, PE and ER was averaged over all IDs, respectively. This measure was used to compare the prediction performance of the models in a descriptive way.

2.6.2.1 Prediction error averaged over all IDs - Brier score (BS)

$$BS_{j,k,m} = \frac{1}{n}\sum_{i=1}^{n} PE_{i,j,k,m}$$

$$i = \{1,2,3,\dots,n\}, where\ n\ is\ the\ total\ number\ of\ individuals\ in\ the\ dataset$$

$$prediction\ visit\ j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\};$$

$$future\ visit\ k = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\};\ j < k;\ \text{and}$$

$$m = \{BSM1, BSM2, BDM1, BDM2\}$$

2.6.2.2 Scaled Brier Score

Brier Score can be scaled by the maximum score for a non-informative model (Steyerberg et al., 2010), resulting in the scaled Brier Score (sBS). sBS can be calculated as below (Steyerberg et al., 2010):

$$sBS_{j,k,m} = 1 - \frac{BS_{j,k,m}}{maxBS_{j,k}}$$

$$prediction\ visit\ j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\};$$

$$future\ visit\ k = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\};\ j < k;\ \text{and}$$

$$m = \{BSM1, BSM2, BDM1, BDM2\}$$

Where the maximum score for a non-informative model can be calculated as:

$$maxBS_{j,k} = mean\ (P[zBMI > 1])_{j,k} * (1 - mean\ (P[zBMI > 1])_{j,k})$$

2.6.3 Measures describing the relative usefulness of models

2.6.3.1 Brier Skill score

This section presents approaches applied in this thesis, which attempt to describe the usefulness of a certain dynamic model compared with GLMER1, GLMER4, or overtime (compared with model at U1).

The concept of Brier Skill Score (BSS) seems to originate in non-medical literature (Wikipedia contributors). This measure describes the relative improvement of a certain model compared with a reference model. For the PEACHES study, I propose two comparisons, one taking model GLMER1 and one taking model GLMER4 as the reference model, respectively. For the simulation study, one comparison was performed taking GLMER1 as the reference model.

Additionally, BSS over time was also calculated for the PEACHES study, the main simulation study as well as the ICC scenarios, respectively. In these analyses, reference model is the one, which is fitted at U1 – from now on called “U1-model”. BSS over time was calculated for GLMER1, BSM1, BSM2, BDM1, and BDM2, respectively.

BSS can be calculated as (World Climate Research Programme, 2009, Glen, 2016):

$$BSS_{j,k,m} = 1 - \frac{BS_{j,k,m}}{BS_{j,k,ref}}$$

$$ref: the\ reference\ model, e.g. GLMER1, GLMER4$$

$$prediction\ visit\ j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\}\ if\ ref = GLMER1,$$

$$j = \{U4, U5, U6, U7, U7a, U8\}\ if\ ref = GLMER4;$$

$$future\ visit\ k = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}\ if\ ref = GLMER1,$$

$$k = \{U5, U6, U7, U7a, U8, U9\}\ if\ ref = GLMER4;\ j < k;\ \text{and}\ m = \{BSM1, BSM2, BDM1, BDM2\}$$

- **Brier Skill score (reference: GLMER1)**

BSS for the comparison with GLMER1 can be calculated for all time points as the followings:

$$BSS_{j,k,m} = 1 - \frac{BS_{j,k,m}}{BS_{j,k,GLMER1}}$$

$$jprediction\ visit\ j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\};$$

$$future\ visit\ k = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\};\ j < k;\ \text{and}$$

$$m = \{BSM1, BSM2, BDM1, BDM2\}$$

This measure was calculated for the PEACHES study, main simulation analysis, as well as the ICC scenarios.

- **Brier Skill score (reference: GLMER4)**

BSS for the comparison with GLMER4 can be calculated for visit U4 or later as the followings:

$$BSS_{j,k,m} = 1 - \frac{BS_{j,k,m}}{BS_{j,k,GLMER4}}$$

$$prediction\ visit\ j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\};$$

$$future\ visit\ k = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\};\ j < k;\ \text{and}$$

$$m = \{BSM1, BSM2, BDM1, BDM2\}$$

This measure was calculated for the PEACHES study.

- **Brier Skill score (reference: model at U1 – "U1-model")**

BSS for the comparison with the "U1-model" can be calculated as the followings:

$$BSS_{j,k,m} = 1 - \frac{BS_{j,k,m}}{BS_{U1,k,m}}$$

$$prediction\ visit\ j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\};$$

$$future\ visit\ k = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\};\ j < k;\ \text{and}$$

$$m = \{BSM1, BSM2, BDM1, BDM2\}$$

This measure was calculated for the PEACHES study, main simulation analysis, as well as the ICC scenarios. For all models, BSS_{U1} is equal 0.

2.6.3.2 Mean logarithmized relative prediction error

For the simulation ICC scenarios, $LRPE^{GLMER1}$ (see 2.6.1.2) was averaged over all LOOCVs to obtain the mean logarithmized relative prediction error (MLRPE^{GLMER1}). This measure was then used to examine the influence of time and ICC on the difference in prediction performance of the Bayesian models and GLMER1 (see 2.10). MLRPE^{GLMER1} was calculated as:

$$MLRPE_{j,k,m}^{GLMER1} = \frac{1}{n}\sum_{i=1}^{n} LRPE_{i,j,k,m}^{GLMER1}$$

$$i = \{1,2,3, \dots, n\}, where\ n\ is\ the\ total\ number\ of\ individuals\ in\ the\ dataset$$

$$prediction\ visit\ j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\};$$

$$future\ visit\ k = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\};\ j < k;\ \text{and}$$

$$m = \{BSM1, BSM2, BDM1, BDM2\}$$

2.7 Graphical presentation of the applied measures

The following plots were created in order to visualize the results of the analyses of PEACHES data (presented in 2.2) and the main simulation study (presented in 2.3).

2.7.1 Logarithmized relative prediction errors (LRPE) on individual level

For each model, boxplots for $LRPE^{GLMER1}$ (see 2.6.1.2) and $LRPE^{GLMER4}$ (see 2.6.1.3) were created to depict the distribution of these measures among individuals for each future visit, at a certain prediction visit. Mean values of these measures are also given within the boxplots as red dots.

Because prediction for all future visits was made at U1, model BSM2 should show the same (therefore the same distributions of the) LRPE, regardless of the number of past visits.

Even though the prediction made by using model BSM1, which considered the same model estimates for the same ID over time, sampling of outcome was done at each prediction visit. This led to different values of predicted outcome for the same prediction visit, which was made at different prediction visits. However, one would expect similar distributions.

Model BDM1 and BDM2 are expected show improvement of model performance by increasing the number of past visits.

LRPE values that are greater than 0 indicates improvement in prediction performance of the alternative model, compared with GLMER1 or GLMER4. The higher the LRPE, the larger is this improvement.

2.7.2 Brier Score over time

At a certain prediction visit, Brier Score of each model for all possible future visits was plotted. For example, at the prediction visit U1, Brier Score of each model was plotted for each future visit, which are U2, U3, U4, U5, U6, U7, U7a, U8, and U9. At the prediction visit U8, Brier Score of each model was plotted for U9. This plot serves as visual comparison between the course of Brier Score over time of different models.

For each of the model GLMER1, GLMER2, GLMER3, GLMER4, and BSM2 and for each future visit, Brier Score should be constant, independent of the prediction visit. In other words, Brier Score of the prediction model GLMER1 for visit U4 is supposed to be the same, if the prediction is made at U1, U2, or U3.

For model BSM1 and for each future visit, the Brier Score is not necessarily constant over time, because the prediction is made by Monte Carlo sampling, which happens independently at each prediction visit. However, a remarkable change over time is not expected.

One can expect that the two models BDM1 and BDM2 would show a changing course of Brier Score over time. The later the prediction visit, the better the model for future visits should perform.

2.7.3 Average prediction error rate over time

Similarly, as presented in the section 2.7.2 above, APER was plotted for each model, each future visit, and each prediction visit, respectively.

One can expect that the differences between models would be smaller, compared with Brier Score.

2.7.4 Scaled Brier Score over time

Scaled Brier Score was plotted for models BSM1, BSM2, BDM1, and BDM2, respectively. The features of this plot are the same as those of the one described in the section 2.7.2 above. This was done for the comparison with GLMER1 and GLMER4, separately. Negative values of scaled Brier Score refer to models with worse prediction performance, compared with a non-informative model.

2.7.5 Brier Skill Score vs. GLMM over time

Brier Skill Score was plotted for models BSM1, BSM2, BDM1, and BDM2, respectively. The features of this plot are the same as those of the one described in the section 2.7.2 above. This was done for the comparison with GLMER1 and GLMER4, separately. Negative values of Brier Skill Score refer to models with worse prediction performance, compared with GLMER1 or GLMER4, respectively. If overperformance of a certain model is observed, updating this model by using updated data improves prediction performance compared with static approaches.

2.7.6 Brier Skill Score vs. “U1-model” over time

Brier Skill Score was plotted for models GLMER1, BSM1, BSM2, BDM1, and BDM2. This was done for the comparison with the “U1-model” of each approach, respectively. For a future visit, BSS of the models at each prediction visit was plotted. Negative values of BSS refer to models with worse prediction performance than the “U1-model”. Overperformance over time of a certain model indicates that updating this model by using updated data improve predictions performance.

2.7.7 Scatter plots of Brier Score for different ICC scenarios

The following plots were created to depict the results of the simulation ICC study.

Scatter plots were created to depict the course of Brier Score for each prediction visit, as well as future visit, similarly as described in 2.7.2. The plots were made for each model and each ICC scenario, respectively.

The higher ICC indicates higher inter-individual, or in other words, higher variance of random intercepts distribution. Updating random intercepts using information from past visit(s) would then improve prediction performance of the dynamic models (BDM1 and BDM2), compared to the static models. The more visits in the past, the more information from the data is gained. It is then expected that the improvement in prediction performance of the dynamic models would then be more visible over time.

Within a certain ICC scenario, between the models BDM1 and BDM2 as well as between models BSM1 and BSM2, no large differences in prediction performance are expected. Prediction performance of the Bayesian static models (BSM1 and BSM2) is expected to be more similar to performance of GLMER1 than those of the dynamic Bayesian models BDM1 and BDM2.

2.7.8 Calibration plots

Calibration plots were created to plot the observed values against the predicted values at a certain future visit, using data from certain prediction visit(s). Perfect agreement of all observations, which means perfect prediction would be a regression line with intercept 0 and slope 1. A slope smaller

than 1 indicates risk underestimation, a slope greater than 1 indicates risk overestimation. The change of calibration over time of BDM1, BDM2, and BSM1 is also demonstrated in these plots, while calibration of BSM2 and the GLMM, where no updates are made, should stay the same over time.

Calibration plots of future visit U1 and U9 are shown in this thesis. Calibration plots of all other visits can be found in the supplementary material.

2.8 Quantification of the association between prediction performance and overall time effect

The association between relative prediction performance and number of past months was quantified by applying a mixed-effects regression model including random intercepts. The dependent variable of the model was $\ln\left(RPE_{i,j,m}\right)$ (see 0). The independent variables included number of months as well as its fractional polynomials up to grade 4. The best model was then chosen by likelihood ratio tests and backward elimination of fixed effects. For this purpose, function `step()` from the package `lmerTest` was applied (Kuznetsova et al., 2017).

Overall time effect can be calculated using coefficient estimates from the mixed model described above. It is the amount of improvement in prediction performance by adding one or more month(s) from the past. In the context of PEACHES study and the main simulation study, overall time effect is rather a vector of values corresponding to the number of past months. Overall time effect can be interpreted as the value of information from the past that contributes to the improvement of prediction performance, relative to the reference model.

2.9 Results on the individual level

For 10 IDs, models are graphically compared at the individual level regarding estimates and predicted outcome at each timepoint. The aim of this analysis is to demonstrate the updating process of the dynamic models and to compare the estimates obtained from this process with those obtained from the static model on the individual level.

Additionally, in the simulation main study, the following aspects were considered in the comparison between models: estimate bias of fixed effects, change of random intercept over time, and predicted outcome compared with simulated values. In case of biased estimate at U1, the change of random intercept over time could be more extreme for BDM1, since the fixed effects stay the same and any attempt to adapt the model would then affect the random effects.

2.10 Examining factors that associate with prediction error

In a linear mixed model, ICC and time as influencing factors were considered, in order to quantify the effect of these factors on the Brier Score (see 2.6.2.1) and $MLRPE^{GLMER1}$ (see 2.6.3.2), respectively. Backward selection was done with the full LMMs to select best model with respect to AIC[1].

[1] Model selection was done with backward selection using AIC. The function `step()` in package `lmerTest` ibid. was used.

The full LMM for the analysis with BS is specified as followings:

$$BS_{j,k,m} = \beta_0 + \beta_1 * t_{k,m} + \beta_2 * ICC_{j,k,m} + \beta_3 * \mathrm{m}_{j,k} + \beta_4 * ICC_{j,k,m} * \mathrm{m}_{j,k} + \beta_5 * ICC_{j,k,m} * t_{k,m} + \beta_6 * ICC_{j,k,m} * t_{k,m} * \mathrm{m}_{j,k} + \varepsilon_{j,k,m}$$

The full LMM for the analysis with MLRPE$^{\text{GLMER1}}$ is specified as followings:

$$MLRPE_{j,k,m}^{GLMER1} = \beta_0 + \beta_1 * t_{k,m} + \beta_2 * ICC_{j,k,m} + \beta_3 * \mathrm{m}_{j,k} + \beta_4 * ICC_{j,k,m} * \mathrm{m}_{j,k} + \beta_5 * ICC_{j,k,m} * t_{k,m} + \beta_6 * ICC_{j,k,m} * t_{k,m} * \mathrm{m}_{j,k} + \varepsilon_{j,k,m}$$

$where$: $t_{k,m}$ = average time from U1 in months

$$prediction\ visit\ j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\};$$

$$future\ visit\ k = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\};\ j < k;\ \text{and}$$

$$m = \{BSM1, BSM2, BDM1, BDM2\}$$

2.11 Supplementary analyses

2.11.1 Extended GLMM models

In a supplementary analysis, the following extended GLMM models were examined:

- GLMER2: additive time effects from birth, mother's preconception BMI status (preBMI), and child's largeness for gestational age (LGA), as well as their interaction with later time (U6, U7). Prediction is made once at U1 using estimates from model fitted with cross-validation training data set. No updates of prediction over time.
- GLMER3: additive time effects from U4 (3 months after birth). Prediction is made once at U5 using estimates from model fitted with cross-validation training data set. No updates of prediction over time.
- GLMER4: additive time effects from U4 (3 months after birth), mother's preconception BMI status (preBMI), and child's largeness for gestational age (LGA), as well as their interaction with time. Prediction is made once at U4 using estimates from model fitted with cross-validation training data set. No updates of prediction over time.

2.11.1.1 Extended model GLMER2

Table 19. Model specification of model GLMER2

GLMER2
Step 1 – Training with available IDs *i**
Data:

GLMER2

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \\ X_{1_{i^*}} \\ X_{2_{i^*}} \\ X_{3_{i^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample.

Model specification:

$$P[BMIz > 1]_{i^*,j^*} = \frac{e^{LP_{i^*,j^*}}}{1 + e^{LP_{i^*,j^*}}}$$

$$LP_{i^*,j^*} = \alpha_{i^*} + \beta_0 +$$

$$\beta_1 * t_{1_{i^*,j^*}} + \beta_2 * t_{2_{i^*,j^*}} + \beta_3 * t_{3_{i^*,j^*}} + \beta_4 * t_{4_{i^*,j^*}} + \beta_6 * t_{6_{i^*,j^*}} + \beta_7 * t_{7_{i^*,j^*}} +$$

$$\gamma_1 * X_{1_{i^*}} + \gamma_2 * X_{2_{i^*}} + \gamma_3 * Z_{U1_{i^*}} +$$

$$\gamma_4 * t_{6_{i^*,j^*}} * X_{2_{i^*}} + \gamma_5 * t_{7_{i^*,j^*}} * X_{2_{i^*}} + \gamma_6 * t_{6_{i^*,j^*}} * Z_{U1_{i^*}} + \gamma_7 * t_{7_{i^*,j^*}} * Z_{U1_{i^*}}$$

$Random\ intercepts: \alpha_{i^*} \sim N(0, \tau^2)$

$LP_{i^*,j^*}: linear\ predictor\ of\ ID\ i^*\ at\ timepoint\ j^*,$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

Covariates:

X_1: mother's pre-conceptional obesity status

X_2: child's largeness for gestational age

Z_{U1}: child's BMI Z-score at U1

Model estimates:

Fixed effects: $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}, \widehat{\beta_4}, \widehat{\beta_6}, \widehat{\beta_7}, \widehat{\gamma_1}, \widehat{\gamma_2}, \widehat{\gamma_3}, \widehat{\gamma_4}, \widehat{\gamma_5}, \widehat{\gamma_6}, \widehat{\gamma_7}$

$\widehat{LP_{i^*,J}}$, *where* $j = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

Random intercepts: $\widehat{\alpha_{i^*}}$

Step 2 – Predicting at U1

Assumption:

Random intercepts of child i form the test sample: $\alpha_i = 0$

Data:

GLMER2

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \\ X_{1_{i^*}} \\ X_{2_{i^*}} \\ X_{3_{i^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample,

random intercepts of IDs from the LOOCV training sample: $\widehat{\alpha_{i^*}}$, *and*

$$\begin{bmatrix} t_{1_{i,j}} \\ t_{2_{i,j}} \\ t_{3_{i,j}} \\ t_{4_{i,j}} \\ t_{6_{i,j}} \\ t_{7_{i,j}} \\ X_{1_i} \\ X_{2_i} \\ X_{3_i} \end{bmatrix}$$

Prediction model:

$$P[BMIz > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \alpha_i + \widehat{\beta_0} +$$

$$\widehat{\beta_1} * t_{1_{i,j}} + \widehat{\beta_2} * t_{2_{i,j}} + \widehat{\beta_3} * t_{3_{i,j}} + \widehat{\beta_4} * t_{4_{i,j}} + \widehat{\beta_6} * t_{6_{i,j}} + \widehat{\beta_7} * t_{7_{i,j}} +$$

$$\widehat{\gamma_1} * X_{1_i} + \widehat{\gamma_2} * X_{2_i} + \widehat{\gamma_3} * Z_{U1_i} +$$

$$\widehat{\gamma_4} * t_{6_{i,j}} * X_{2_i} + \widehat{\gamma_5} * t_{7_{i,j}} * X_{2_i} + \widehat{\gamma_6} * t_{6_{i,j}} * Z_{U1_i} + \widehat{\gamma_7} * t_{7_{i,j}} * Z_{U1_i}$$

$LP_{i,j}$: $linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$

$j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

Predicted values:

$\widehat{LP_{i,j}}$, *where* $j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

At visit U2, U3, U4, U5, U6, U7a, and U8, no predictions are made but prediction results from visit U1 are adopted.

2.11.1.2 Extended model GLMER3

Table 20. Model specification of model GLMER3

GLMER3

Step 1 – Training with available IDs *i

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \\ Z_{U4_{i^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample.

Model specification:

$$P[BMIz > 1]_{i^*,j^*} = \frac{e^{LP_{i^*,j^*}}}{1 + e^{LP_{i^*,j^*}}}$$

$$LP_{i^*,j^*} = \alpha_{i^*} + \beta_0 + \beta_5 * t_{5_{i^*,j^*}} + \beta_6 * t_{6_{i^*,j^*}} + \beta_7 * t_{7_{i^*,j^*}} + \beta_{7a} * t_{7a_{i^*,j^*}} +$$

$$\gamma_1 * Z_{U4_{i^*}} + \gamma_2 * t_{5_{i^*,j^*}} * Z_{U4_{i^*}} + \gamma_3 * t_{6_{i^*,j^*}} * Z_{U4_{i^*}} + \gamma_4 * t_{7_{i^*,j^*}} * Z_{U4_{i^*}} + \gamma_5 * t_{7a_{i^*,j^*}} * Z_{U4_{i^*}}$$

$$Random\ intercepts: \alpha_{i^*} \sim N(0, \tau^2)$$

$$LP_{i^*,j^*}: linear\ predictor\ of\ ID\ i^*\ at\ timepoint\ j^*$$

$$j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Model estimates:

Fixed effects: $\widehat{\beta_0}$, $\widehat{\beta_1}$, $\widehat{\beta_2}$, $\widehat{\beta_3}$, $\widehat{\beta_4}$, $\widehat{\beta_6}$, $\widehat{\beta_7}$

$\widehat{LP_{i^*,J}}$, *where* $j = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

Random intercepts: $\widehat{\alpha_{i^*}}$

Covariates:

Z_{U4}: child's risk of overweight at U4

Step 2 – Predicting at U4

GLMER3

Assumption:

Random intercepts of child i form the test sample: $\alpha_i = 0$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \\ Z_{U4_{i^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample,

random intercepts of IDs from the LOOCV training sample: $\widehat{\alpha_{i^*}}$, and

$$\begin{bmatrix} t_{1_{i,j}} \\ t_{2_{i,j}} \\ t_{3_{i,j}} \\ t_{4_{i,j}} \\ t_{6_{i,j}} \\ t_{7_{i,j}} \\ Z_{U4_i} \end{bmatrix},$$

where $j = \{U5, U6, U7, U7a, U8, U9\}$ and *i* is the child in the test sample

Prediction model:

$$P[BMIz > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \alpha_i + \hat{\beta}_0 + \widehat{\beta_5} * t_{5_{i,j}} + \widehat{\beta_6} * t_{6_{i,j}} + \widehat{\beta_7} * t_{7_{i,j}} + \widehat{\beta_{7a}} * t_{7a_{i,j}} + \gamma_1 * Z_{U4_i} + \gamma_2 * t_{5_{i,j}} * Z_{U4_i} + \gamma_3 * t_{6_{i,j}} * Z_{U4_i} + \gamma_4 * t_{7_{i,j}} * Z_{U4_i} + \gamma_5 * t_{7a_{i,j}} * Z_{U4_i}$$

$$LP_{i,j}: linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$$

$$j = \{U5, U6, U7, U7a, U8, U9\}$$

Predicted values:

$$\widehat{LP_{i,j}},\ where\ j = \{U5, U6, U7, U7a, U8, U9\}$$

At visit U5, U6, U7a, and U8, no predictions are made but prediction results from visit U4 are adopted.

2.11.1.3 Extended model GLMER4

Table 21. Model specification of model GLMER4

GLMER4

Step 1 – Training with available IDs *i**

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \\ Z_{U4_{i^*}} \\ X_{1_{i^*}} \\ X_{2_{i^*}} \\ X_{3_{i^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample.

Model specification:

$$P[BMIz > 1]_{i^*,j^*} = \frac{e^{LP_{i^*,j^*}}}{1 + e^{LP_{i^*,j^*}}}$$

$$LP_{i^*,j^*} = \alpha_{i^*} + \beta_0 + \beta_5 * t_{5_{i^*,j^*}} + \beta_6 * t_{6_{i^*,j^*}} + \beta_7 * t_{7_{i^*,j^*}} + \beta_{7a} * t_{7a_{i^*,j^*}} +$$
$$\gamma_1 * Z_{U4_{i^*}} + \gamma_2 * t_{5_{i^*,j^*}} * Z_{U4_{i^*}} + \gamma_3 * t_{6_{i^*,j^*}} * Z_{U4_{i^*}} + \gamma_4 * t_{7_{i^*,j^*}} * Z_{U4_{i^*}} +$$
$$\gamma_5 * t_{7a_{i^*,j^*}} * Z_{U4_{i^*}} + \gamma_6 * Z_{U1_{i^*}} + \gamma_7 * Z_{U4_{i^*}} + \gamma_8 * X_{1_{i^*}} + \gamma_9 * X_{2_{i^*}} +$$
$$\gamma_{10} * X_{1_{i^*}} * t_{5_{i^*,j^*}} + \gamma_{11} * X_{2_{i^*}} * t_{5_{i^*,j^*}} + \gamma_{12} * X_{1_{i^*}} * t_{6_{i^*,j^*}} + \gamma_{13} * X_{2_{i^*}} * t_{6_{i^*,j^*}} +$$
$$\gamma_{14} * X_{1_{i^*}} * t_{7_{i^*,j^*}} + \gamma_{15} * X_{2_{i^*}} * t_{7_{i^*,j^*}} + \gamma_{16} * X_{1_{i^*}} * t_{7a_{i^*,j^*}} + \gamma_{17} * X_{2_{i^*}} * t_{7a_{i^*,j^*}}$$

$$Random\ intercepts: \alpha_{i^*} \sim N(0, \tau^2)$$

$$LP_{i^*,j^*}: linear\ predictor\ of\ ID\ i^*\ at\ timepoint\ j^*$$

$$j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

Model estimates:

Fixed effects: $\widehat{\beta_0}$, $\widehat{\beta_1}$, $\widehat{\beta_2}$, $\widehat{\beta_3}$, $\widehat{\beta_4}$, $\widehat{\beta_6}$, $\widehat{\beta_7}$

$\widehat{LP_{i^*,j}}$, *where* $j = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$

GLMER4

Random intercepts: $\widehat{\alpha_{i^*}}$

Covariates:

X_1: mother's pre-conceptional obesity status

X_2: child's largeness for gestational age

Z_{U4}: child's risk of overweight at U4

Step 2 – Predicting at U4

Assumption:

Random intercepts of child i form the test sample: $\alpha_i = 0$

Data:

$$\begin{bmatrix} Y_{i^*,j^*} \\ t_{1_{i^*,j^*}} \\ t_{2_{i^*,j^*}} \\ t_{3_{i^*,j^*}} \\ t_{4_{i^*,j^*}} \\ t_{6_{i^*,j^*}} \\ t_{7_{i^*,j^*}} \\ Z_{U4_{i^*}} \\ X_{1_{i^*}} \\ X_{2_{i^*}} \\ X_{3_{i^*}} \end{bmatrix},$$

where $j^* = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$ and *i** are IDs from the LOOCV training sample,
random intercepts of IDs from the LOOCV training sample: $\widehat{\alpha_{i^*}}$, and

$$\begin{bmatrix} \mathrm{t}_{1_{\mathrm{i,j}}} \\ \mathrm{t}_{2_{\mathrm{i,j}}} \\ \mathrm{t}_{3_{\mathrm{i,j}}} \\ \mathrm{t}_{4_{\mathrm{i,j}}} \\ \mathrm{t}_{6_{\mathrm{i,j}}} \\ \mathrm{t}_{7_{\mathrm{i,j}}} \\ Z_{U4_i} \\ X_{1_i} \\ X_{2_i} \\ X_{3_i} \end{bmatrix},$$

where $j = \{U5, U6, U7, U7a, U8, U9\}$ and *i* is the child in the test sample

Prediction model:

$$P[BMIz > 1]_{i,j} = \frac{e^{LP_{i,j}}}{1 + e^{LP_{i,j}}}$$

$$LP_{i,j} = \alpha_i + \widehat{\beta_0} + \widehat{\beta_5} * t_{5_{i,j}} + \widehat{\beta_6} * t_{6_{i,j}} + \widehat{\beta_7} * t_{7_{i,j}} + \widehat{\beta_{7a}} * t_{7a_{i,j}} +$$

GLMER4

$$\widehat{\gamma_1} * Z_{U4_i} + \widehat{\gamma_2} * t_{5_{i,j}} * Z_{U4_i} + \widehat{\gamma_3} * t_{6_{i,j}} * Z_{U4_i} + \widehat{\gamma_4} * t_{7_{i,j}} * Z_{U4_i} +$$

$$\widehat{\gamma_5} * t_{7a_{i,j}} * Z_{U4_i} + \widehat{\gamma_6} * Z_{U1_i} + \widehat{\gamma_7} * Z_{U4_i} + \widehat{\gamma_8} * X_{1_i} + \widehat{\gamma_9} * X_{2_i} +$$

$$\widehat{\gamma_{10}} * X_{1_i} * t_{5_{i,j}} + \widehat{\gamma_{11}} * X_{2_i} * t_{5_{i,j}} + \widehat{\gamma_{12}} * X_{1_i} * t_{6_{i,j}} + \widehat{\gamma_{13}} * X_{2_i} * t_{6_{i,j}} +$$

$$\widehat{\gamma_{14}} * X_{1_i} * t_{7_{i,j}} + \widehat{\gamma_{15}} * X_{2_i} * t_{7_{i,j}} + \widehat{\gamma_{16}} * X_{1_i} * t_{7a_{i,j}} + \widehat{\gamma_{17}} * X_{2_i} * t_{7a_{i,j}}$$

$$LP_{i,j}: linear\ predictor\ of\ ID\ i\ at\ timepoint\ j$$

$$j = \{U5, U6, U7, U7a, U8, U9\}$$

Predicted values:

$$\widehat{LP_{i,j}},\ where\ j = \{U5, U6, U7, U7a, U8, U9\}$$

At visit U5, U6, U7a, and U8, no predictions are made but prediction results from visit U4 are adopted.

2.11.2 Comparing models using error rate

In addition to Brier score, classification error at each time point was calculated and examined in a supplementary analysis.

For a certain ID, classification prediction error rate was calculated at each timepoint, respectively.

$$ER_{i,j,m} = \begin{cases} 1\ if\ P[zBMI > 1]_{i,j,m}\ | zBMI_{observed} > 1 \\ 0\ if\ P[zBMI > 1]_{i,j,m}\ | zBMI_{observed} \leq 1 \end{cases}$$

$$i = \{1,2,3, \dots, n\}, where\ n\ is\ the\ total\ number\ of\ individuals\ in\ the\ dataset$$

$$j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8, U9\}$$

$$m = \{GLMER1, BSM1, BSM2, BDM1, BDM2\}$$

Prediction error rate averaged for all IDs

$$APER_{j,k,m} = \frac{1}{n}\sum_{i=1}^{n} ER_{i,j,k,m}$$

$$i = \{1,2,3, \dots, n\}, where\ n\ is\ the\ total\ number\ of\ individuals\ in\ the\ dataset$$

$$prediction\ visit\ j = \{U1, U2, U3, U4, U5, U6, U7, U7a, U8\};$$

$$future\ visit\ k = \{U2, U3, U4, U5, U6, U7, U7a, U8, U9\};\ j < k;\ \text{and}$$

$$m = \{BSM1, BSM2, BDM1, BDM2\}$$

2.11.3 Analyses with smaller sample size

For the PEACHES study and simulation main study, sensitivity analyses were performed with sample size of 100 to examine the impact of a smaller sample size on the results of the study.

2.12 Literature search

A non-systematic literature search was performed to identify methods of dynamic prediction in clinical or epidemiological context. Snowball method and citation searching were performed in Google Scholar with Jenkins et al., which was published in 2018 (Jenkins et al., 2018) as key publication. Articles that cited this publication, “related” articles (according to Google Scholar) as well as references of this publication were screened for their relevance.

2.13 Statistical program

R version 4.0.4 (2021-02-15), on platform: x86_64-pc-linux-gnu (64-bit), running under Debian GNU/Linux bullseye/sid was used to perform analyses and to create results, tables, and plots. For further details about used packages, see Appendix A: Technical information.

3. Results

This part of the thesis presents the results of the 1) PEACHES study, 2) simulation main study with ICC estimated from the PEACHES study, 3) simulation study with different ICC, 4) results of the supplementary analyses, and 5) literature research. A summary of the results can be found at the end of the chapter (section 0).

3.1 PEACHES study

Logarithmized Relative Prediction Error (based on Brier's score) (LRPE) of BSM1, BSM2, BDM1, and BDM2 compared with that of GLMER1 using is presented in 3.1.1.1, 3.1.1.2, 3.1.1.3, and 3.1.1.4, respectively. In 3.1.2, comparison between these five models using the average RPE is depicted, followed by comparison regarding calibration (in 3.1.3), scaled Brier Score (in 3.1.4), and Brier Skill Score (in 3.1.5).

3.1.1 Relative prediction error

3.1.1.1 Relative prediction error of GLMER1 vs BSM1

The following table shows the output of the LMM, which quantifies the difference between GLMER1 and BSM1 over time. On average, the relative difference of prediction error (RPE) between the two models GLMER1 vs. BSM1 is $e^{0.158}$ = 1.171 at U1. This indicates that BSM1 predicts better than GLMER1 at U1. This difference stays the same, regardless of the number of previous months. Table 22 shows the estimated coefficients of the LMM. The proportion of variances that was contributed by different visits is $\frac{0.116^2}{(0.178^2+0.116^2+0.156^2)} * 100\%$ =19.368%. Variances among individuals contribute to 45.604% of the total variances.

Table 22. LRPE of GLMER1 vs. BSM1 in PEACHES study – output of LMM

group	term	estimate	std.error	statistic	df	p.value
	(Intercept)	0.158	0.039	4.069	7.990	3.60E-03
udata:PseudoID	sd__(Intercept)	0.178				
udata	sd__(Intercept)	0.116				
Residual	sd__Observation	0.156				

The distribution of LRPE in 1,000 LOOCVs test datasets over time is presented in the following box plots in relationship with number of past visits. Red dots represent mean of a distribution at a certain prediction visit. Nine panels represent the future visits. E.g., panel "U2" shows the distribution of LRPE for predicted outcome at U2, which was predicted using data at U1. For the prediction of outcome at U9, regardless of number of previous visits, LRPE of GLMER1 vs. BSM1 seem not to differ.

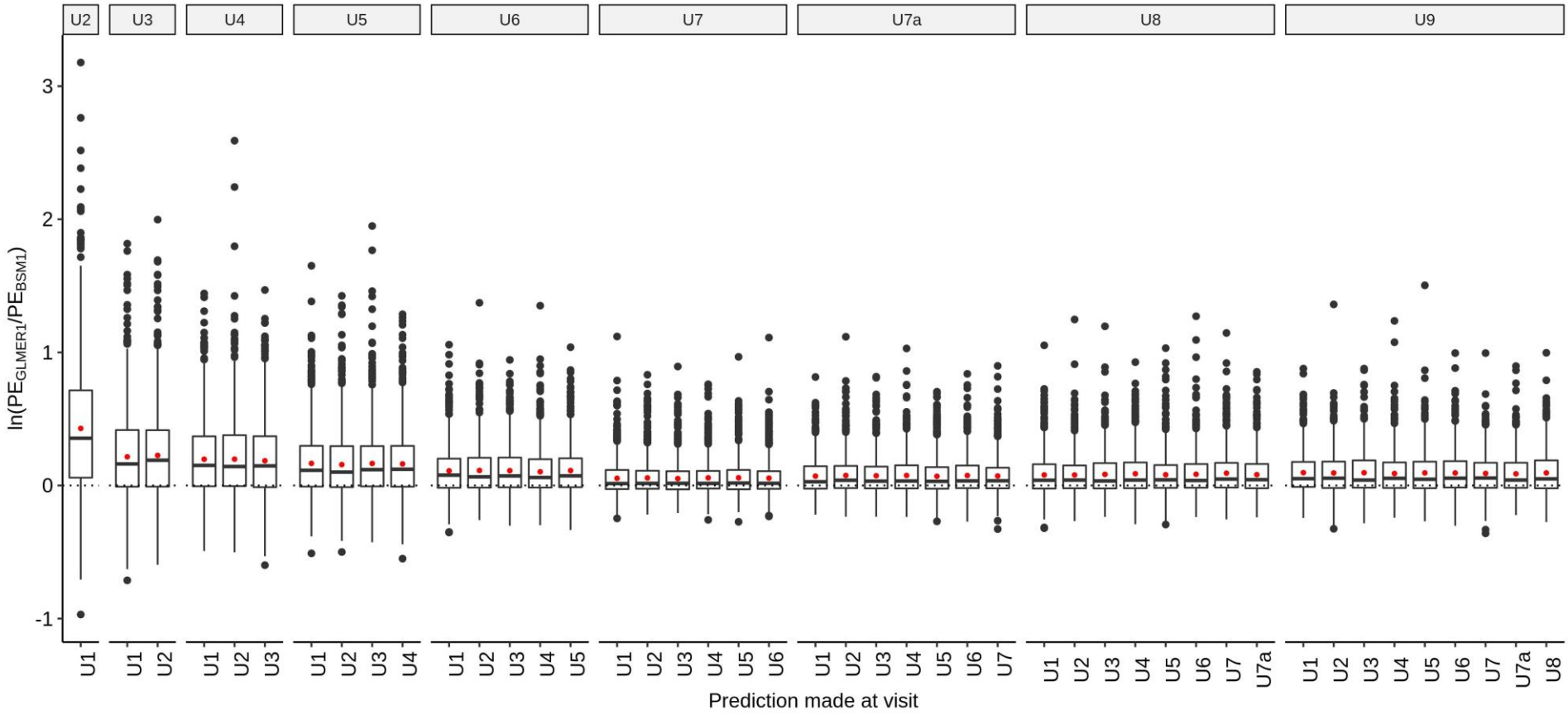


Figure 3. LRPE of GLMER1 vs. BSM1 over time in PEACHES study

3.1.1.2 Relative prediction error of GLMER1 vs. BSM2

Table 23. LRPE of GLMER1 vs. BSM2 in PEACHES study – output of LMM

group	term	estimate	std.error	statistic	df	p.value
	(Intercept)	0.164	0.004	38.451	945.459	1.6961E-195
PseudoID	sd__(Intercept)	0.091				
Residual	sd__Observation	0.280				

Results of the LMM, which quantifies the time effect on LRPE of GLMER1 vs. BSM2 are shown in Table 23. On average, the relative difference of prediction error between the two models GLMER1 vs. BSM1 is $e^{0.164}$ = 1.178 at U1. This indicates that BSM2 predicts better than GLMER1 at U1. Similarly, as results of the comparison between GLMER1 vs. BSM2, this difference stays the same, regardless of the number of previous months.

Box plots in the following figure presents the distribution of LRPE over time, with increasing number of previous visits. Prediction was made at U1 for all future visits at once and no model updating was done. Thus, no change in predicted outcome if passing more information into the model. For that reason, box plots within one panel are exact the same.

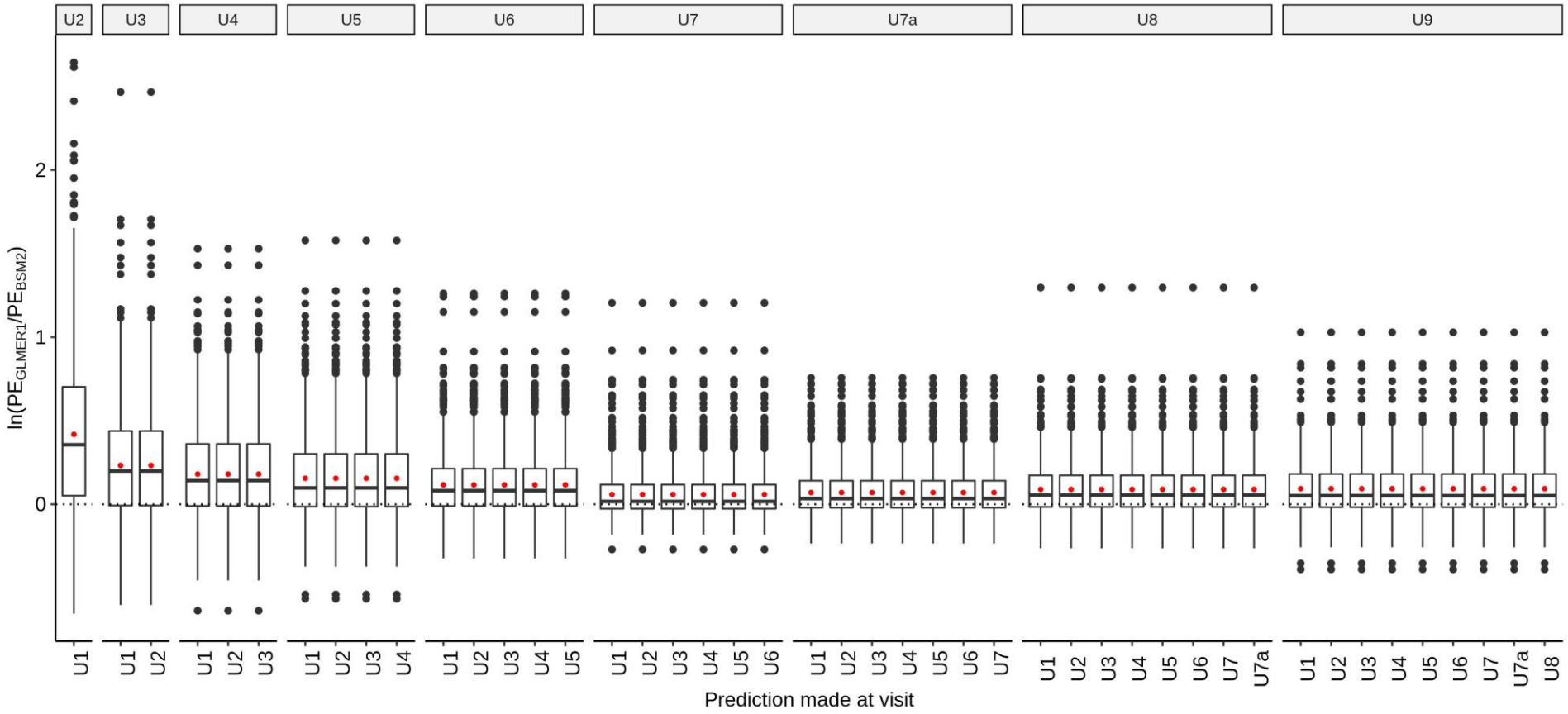


Figure 4. LRPE of GLMER1 vs. BSM2 over time in PEACHES study

3.1.1.3 Relative prediction error of GLMER1 vs. BDM1

Table 24. LRPE of GLMER1 vs. BDM1 in PEACHES study – output of LMM

group	term	estimate	std.error	statistic	df	p.value
	(Intercept)	-0.908	0.178	-5.096	8.001	9.34E-04
	months_from_u1	0.095	0.003	37.884	30467.862	6.32E-307
	I(months_from_u1^2)	-5.41E-03	2.92E-04	-18.541	30476.356	2.54E-76
	I(months_from_u1^3)	1.43E-04	1.08E-05	13.251	30482.428	5.77E-40
	I(months_from_u1^4)	-1.28E-06	1.21E-07	-10.576	30486.671	4.26E-26
udata:PseudoID	sd__(Intercept)	0.899				
udata	sd__(Intercept)	0.534				
Residual	sd__Observation	0.524				

On average, RPE of GLMER1 vs. BDM1 at U1 is $e^{-0.908}$= 0.403, which means GLMER1 overperforms BDM1 at this visit. However, this difference attenuates over time. Non-linearity of time effect is present. The proportion of variances that was contributed by different visits is $\frac{0.534^2}{(0.899^2+0.534^2+0.524^2)} * 100\%$ =20.846%. Figure 5 presents distribution of the LRPE of GLMER1 vs. BDM1. The box plots tend to move upward with increasing number of previous visits. Their mean/median lie above 0 when data from U6 or later was used for prediction. This observation supports the results of the LMM presented above (Table 24).

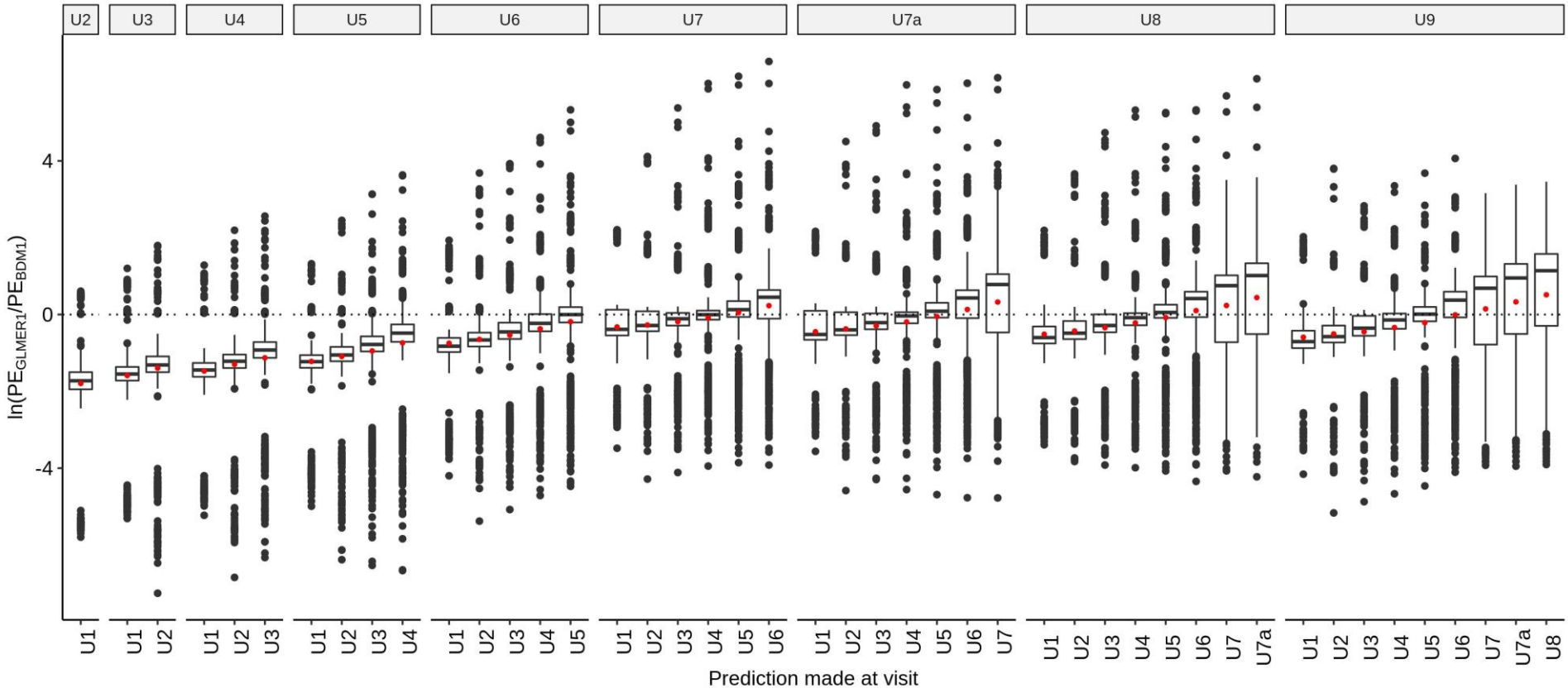


Figure 5. LRPE of GLMER1 vs. BDM1 over time in PEACHES study

3.1.1.4 Relative prediction error of GLMER1 vs. BDM2

Table 25. LRPE of GLMER1 vs. BDM2 in PEACHES study– output of LMM

group	term	estimate	std.error	statistic	df	p.value
	(Intercept)	-0.911	0.178	-5.127	8.015	8.95E-04
	months_from_u1	0.097	0.003	38.529	30474.364	0.00E+00
	I(months_from_u1^	-5.64E-03	2.94E-04	-19.171	30482.974	1.96E-81
	I(months_from_u1^	1.51E-04	1.09E-05	13.879	30489.127	1.16E-43
	I(months_from_u1^	-1.36E-06	1.22E-07	-11.202	30493.428	4.56E-29
udata:PseudoID	sd__(Intercept)	0.900				
udata	sd__(Intercept)	0.532				
Residual	sd__Observation	0.528				

On average, GLMER1 overperforms BDM2 at U1. However, this difference attenuates over time in a linear manner. Non-linear time effects are present. The proportion of variances that was contributed by different visits is $\frac{0.532^2}{(0.900^2+0.532^2+0.528^2)} * 100\% = 20.631\%$. Similarly, as when comparing BDM1 and GLMER1, in Figure 6, the box plots tend to move upward with increasing number of previous visits. Their mean/median lie above 0 when data from U6 or later was used for prediction. This observation supports the results of the LMM presented above (Table 25).

3.1.1.5 Overall time effect on RPE

Overall time effect (see 2.8) of the four Bayesian models is shown in the following table, when incorporating the non-linearity of time effects. RPE of GLMER1 versus BSM1 and BSM2 was shown to be constant and around 1 over time, while RPE of GLMER1 versus BDM1 and BDM2 seems to increase with increasing number of months from U1. At month 36 (U7) and 48 (U7a) after birth, BDM1 and BDM2 overperform GLMER1. This cannot be observed for timepoint 60 months (U9). At this timepoint, GLMER1 performs better than BDM1 and BDM2.

Table 26. RPE of GLMER1 vs. Bayesian models in PEACHES study

Number of months from U1	RPE of GLMER1 vs.			
	BSM1	BSM2	BDM1	BDM2
0	1.171	1.178	0.403	0.393
1	1.171	1.178	0.441	0.436
3	1.171	1.178	0.513	0.519
6	1.171	1.178	0.604	0.625
12	1.171	1.178	0.720	0.751
24	1.171	1.178	0.821	0.840
36	1.171	1.178	1.013	1.056
48	1.171	1.178	1.214	1.252
60	1.171	1.178	0.673	0.515

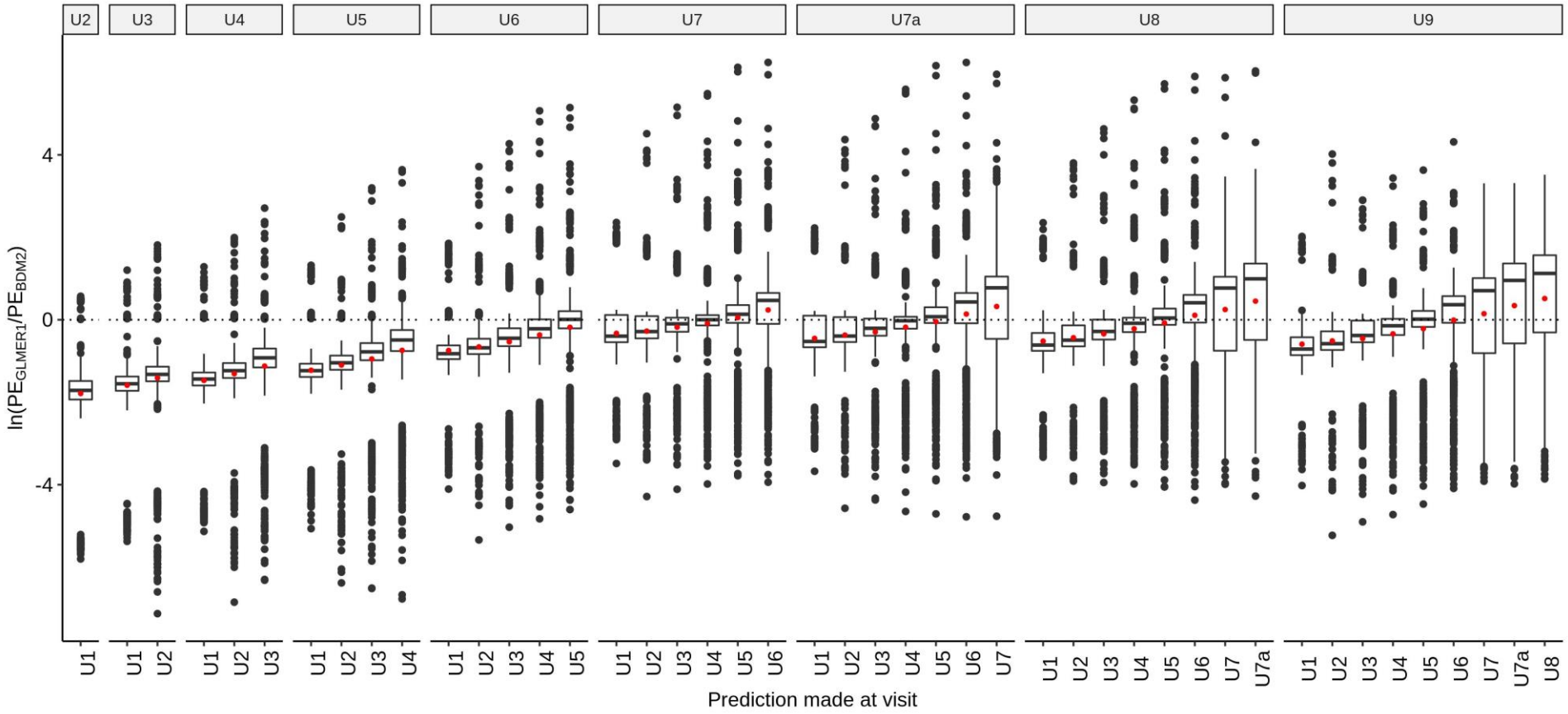


Figure 6. LRPE of GLMER1 vs. BDM2 over time in PEACHES study

3.1.2 Brier Score of Bayesian models vs. GLMER1

Brier Score of GLMER1, BSM1, BSM2, BDM1, and BDM2 for each prediction visit and each future visit was shown in Table 27 and Figure 7. The Brier Score of GLMER1 and BSM2 stay the same, regardless of at what visit the prediction was made. The Brier Score of BSM1 for the same future visit changes only little when considering different prediction visits.

At early visits (U1, U2, and U3), the Brier Score of all models seem to resemble. The improvement in prediction of BDM1 and BDM2 compared with GLMER1, BSM1, and BSM2 can be observed from prediction visit U4 (Figure 7).

Table 27. Brier Score of Bayesian models vs. GLMER1 over time in PEACHES study

Prediction made at visit	Future visit to be predicted	Brier Score				
		GLMER1	BSM2	BSM1	BDM1	BDM2
U1	U2	0.033	0.033	0.033	0.025	0.025
U1	U3	0.083	0.084	0.084	0.078	0.079
U1	U4	0.098	0.098	0.098	0.096	0.096
U1	U5	0.129	0.130	0.130	0.127	0.127
U1	U6	0.170	0.171	0.172	0.168	0.167
U1	U7	0.229	0.230	0.231	0.227	0.228
U1	U7a	0.199	0.200	0.200	0.200	0.201
U1	U8	0.192	0.193	0.193	0.187	0.188
U1	U9	0.193	0.194	0.194	0.182	0.182
U2	U3	0.083	0.084	0.084	0.080	0.080
U2	U4	0.098	0.098	0.098	0.099	0.099
U2	U5	0.129	0.130	0.130	0.133	0.133
U2	U6	0.170	0.171	0.172	0.172	0.172
U2	U7	0.229	0.230	0.230	0.234	0.234
U2	U7a	0.199	0.200	0.200	0.205	0.204
U2	U8	0.192	0.193	0.194	0.190	0.190
U2	U9	0.193	0.194	0.195	0.186	0.187
U3	U4	0.098	0.098	0.098	0.090	0.089
U3	U5	0.129	0.130	0.130	0.124	0.125
U3	U6	0.170	0.171	0.171	0.165	0.165
U3	U7	0.229	0.230	0.231	0.229	0.228
U3	U7a	0.199	0.200	0.200	0.203	0.203
U3	U8	0.192	0.193	0.193	0.188	0.188
U3	U9	0.193	0.194	0.195	0.191	0.191
U4	U5	0.129	0.130	0.130	0.109	0.109
U4	U6	0.170	0.171	0.172	0.149	0.149
U4	U7	0.229	0.230	0.230	0.222	0.222
U4	U7a	0.199	0.200	0.200	0.194	0.194
U4	U8	0.192	0.193	0.193	0.178	0.178
U4	U9	0.193	0.194	0.195	0.185	0.186
U5	U6	0.170	0.171	0.171	0.131	0.131
U5	U7	0.229	0.230	0.231	0.210	0.211
U5	U7a	0.199	0.200	0.201	0.184	0.184
U5	U8	0.192	0.193	0.193	0.168	0.169
U5	U9	0.193	0.194	0.195	0.179	0.179
U6	U7	0.229	0.230	0.231	0.196	0.196
U6	U7a	0.199	0.200	0.200	0.172	0.171
U6	U8	0.192	0.193	0.193	0.159	0.158
U6	U9	0.193	0.194	0.194	0.168	0.168
U7	U7a	0.199	0.200	0.200	0.149	0.150
U7	U8	0.192	0.193	0.193	0.146	0.146
U7	U9	0.193	0.194	0.195	0.156	0.156
U7a	U8	0.192	0.193	0.193	0.128	0.128
U7a	U9	0.193	0.194	0.195	0.143	0.143
U8	U9	0.193	0.194	0.195	0.129	0.130

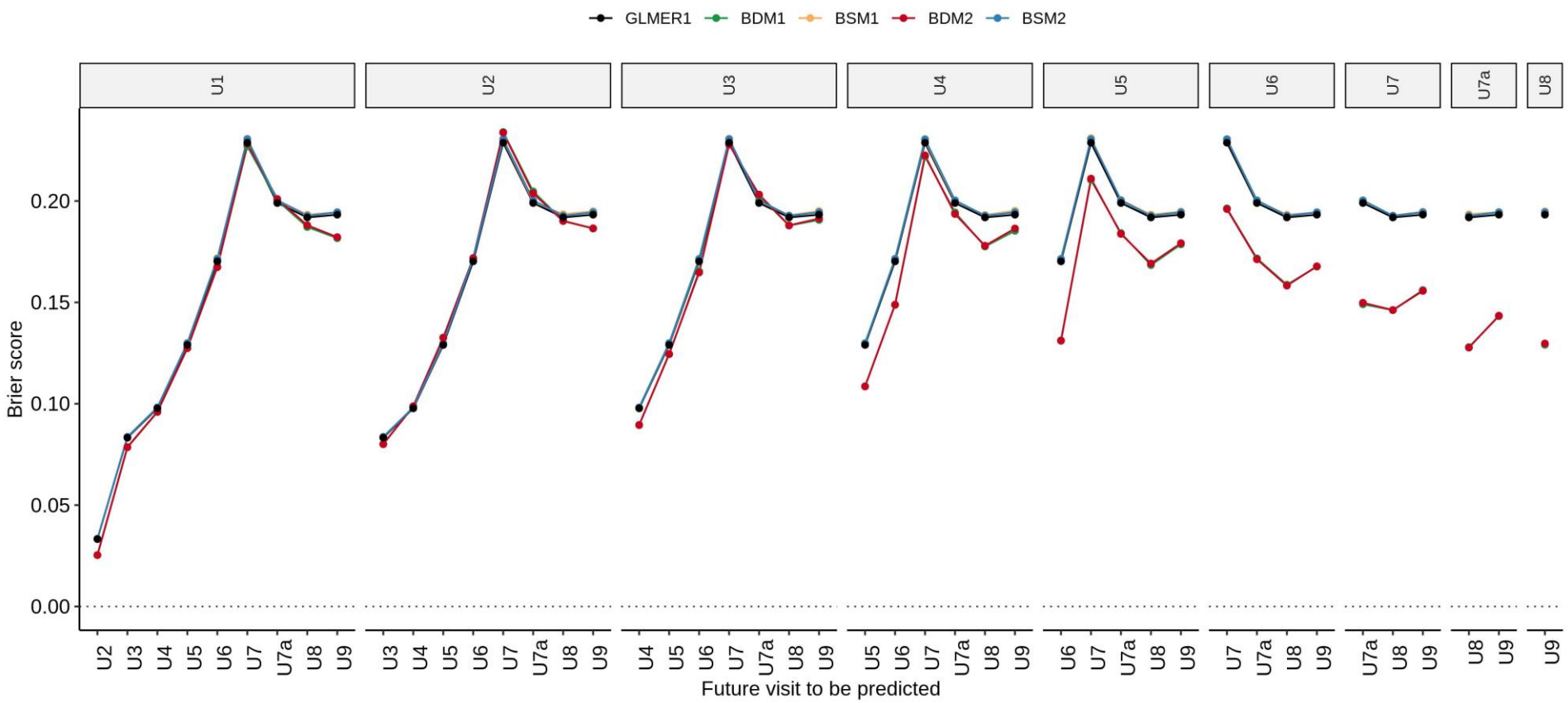


Figure 7. Brier Score of Bayesian models vs. GLMER1 over time in PEACHES study

3.1.3 Calibration of Bayesian models vs. GLMER1

Results of the calibration analysis for a certain future visit in PEACHES study are shown in the following table.

Table 28. Calibration slopes and intercepts of Bayesian models vs. GLMER1 in PEACHES study

Future visit	Prediction visit	GLMER1 Slope	GLMER1 Intercept	BSM1 Slope	BSM1 Intercept	BSM2 Slope	BSM2 Intercept	BDM1 Slope	BDM1 Intercept	BDM2 Slope	BDM2 Intercept
U2	U1	-6.195	0.124	-4.649	0.090	-2.794	0.068	2.032	-0.065	2.036	-0.064
U3	U1	0.200	0.082	0.160	0.083	0.004	0.087	0.627	0.034	0.613	0.035
U3	U2	0.200	0.082	-0.087	0.090	0.004	0.087	0.487	0.047	0.494	0.046
U4	U1	0.272	0.094	0.483	0.087	0.255	0.095	0.331	0.070	0.320	0.071
U4	U2	0.272	0.094	0.424	0.089	0.255	0.095	0.280	0.076	0.273	0.076
U4	U3	0.272	0.094	0.252	0.095	0.255	0.095	0.656	0.037	0.660	0.037
U5	U1	0.671	0.103	0.248	0.129	0.329	0.124	0.224	0.112	0.218	0.113
U5	U2	0.671	0.103	0.619	0.108	0.329	0.124	0.139	0.124	0.138	0.124
U5	U3	0.671	0.103	0.849	0.096	0.329	0.124	0.433	0.084	0.430	0.085
U5	U4	0.671	0.103	0.523	0.114	0.329	0.124	0.864	0.026	0.867	0.025
U6	U1	0.919	0.097	0.647	0.134	0.847	0.111	0.330	0.136	0.346	0.133
U6	U2	0.919	0.097	0.705	0.127	0.847	0.111	0.265	0.151	0.269	0.150
U6	U3	0.919	0.097	0.904	0.105	0.847	0.111	0.454	0.112	0.467	0.109
U6	U4	0.919	0.097	0.735	0.123	0.847	0.111	0.770	0.046	0.763	0.047
U6	U5	0.919	0.097	0.838	0.112	0.847	0.111	1.005	-0.006	1.005	-0.006
U7	U1	0.336	0.247	0.259	0.265	0.332	0.250	0.176	0.265	0.134	0.277
U7	U2	0.336	0.247	0.376	0.241	0.332	0.250	0.078	0.295	0.084	0.293
U7	U3	0.336	0.247	0.248	0.267	0.332	0.250	0.283	0.235	0.300	0.230
U7	U4	0.336	0.247	0.427	0.230	0.332	0.250	0.427	0.192	0.418	0.194
U7	U5	0.336	0.247	0.236	0.269	0.332	0.250	0.584	0.143	0.575	0.146
U7	U6	0.336	0.247	0.306	0.255	0.332	0.250	0.725	0.102	0.728	0.102
U7a	U1	0.674	0.139	0.594	0.157	0.600	0.156	0.274	0.187	0.248	0.194
U7a	U2	0.674	0.139	0.735	0.133	0.600	0.156	0.201	0.209	0.225	0.202
U7a	U3	0.674	0.139	0.651	0.147	0.600	0.156	0.319	0.176	0.309	0.179
U7a	U4	0.674	0.139	0.716	0.136	0.600	0.156	0.485	0.130	0.497	0.127
U7a	U5	0.674	0.139	0.542	0.167	0.600	0.156	0.620	0.091	0.624	0.090
U7a	U6	0.674	0.139	0.681	0.142	0.600	0.156	0.742	0.058	0.747	0.057
U7a	U7	0.674	0.139	0.643	0.149	0.600	0.156	0.970	-0.009	0.958	-0.006
U8	U1	0.611	0.147	0.515	0.166	0.710	0.136	0.424	0.138	0.392	0.146
U8	U2	0.611	0.147	0.462	0.174	0.710	0.136	0.349	0.161	0.353	0.159
U8	U3	0.611	0.147	0.629	0.149	0.710	0.136	0.438	0.137	0.437	0.138
U8	U4	0.611	0.147	0.585	0.156	0.710	0.136	0.627	0.090	0.620	0.092
U8	U5	0.611	0.147	0.520	0.165	0.710	0.136	0.734	0.062	0.720	0.065
U8	U6	0.611	0.147	0.538	0.163	0.710	0.136	0.818	0.041	0.823	0.040
U8	U7	0.611	0.147	0.691	0.140	0.710	0.136	0.948	0.004	0.944	0.005
U8	U7a	0.611	0.147	0.476	0.172	0.710	0.136	1.117	-0.037	1.116	-0.036
U9	U1	0.322	0.196	0.451	0.181	0.424	0.185	0.529	0.120	0.509	0.125
U9	U2	0.322	0.196	0.257	0.206	0.424	0.185	0.382	0.157	0.376	0.158
U9	U3	0.322	0.196	0.216	0.212	0.424	0.185	0.312	0.171	0.297	0.174
U9	U4	0.322	0.196	0.135	0.222	0.424	0.185	0.448	0.140	0.427	0.145
U9	U5	0.322	0.196	0.309	0.200	0.424	0.185	0.556	0.114	0.547	0.116
U9	U6	0.322	0.196	0.476	0.178	0.424	0.185	0.695	0.084	0.691	0.085
U9	U7	0.322	0.196	0.323	0.198	0.424	0.185	0.825	0.052	0.827	0.052
U9	U7a	0.322	0.196	0.362	0.193	0.424	0.185	0.959	0.024	0.954	0.025
U9	U8	0.322	0.196	0.264	0.205	0.424	0.185	1.085	-0.005	1.074	-0.002

For prediction at U2, all models show bad calibration and seem to underestimate the risk of overweight at U2. BDM1 and BDM2 seem to overperform other models.

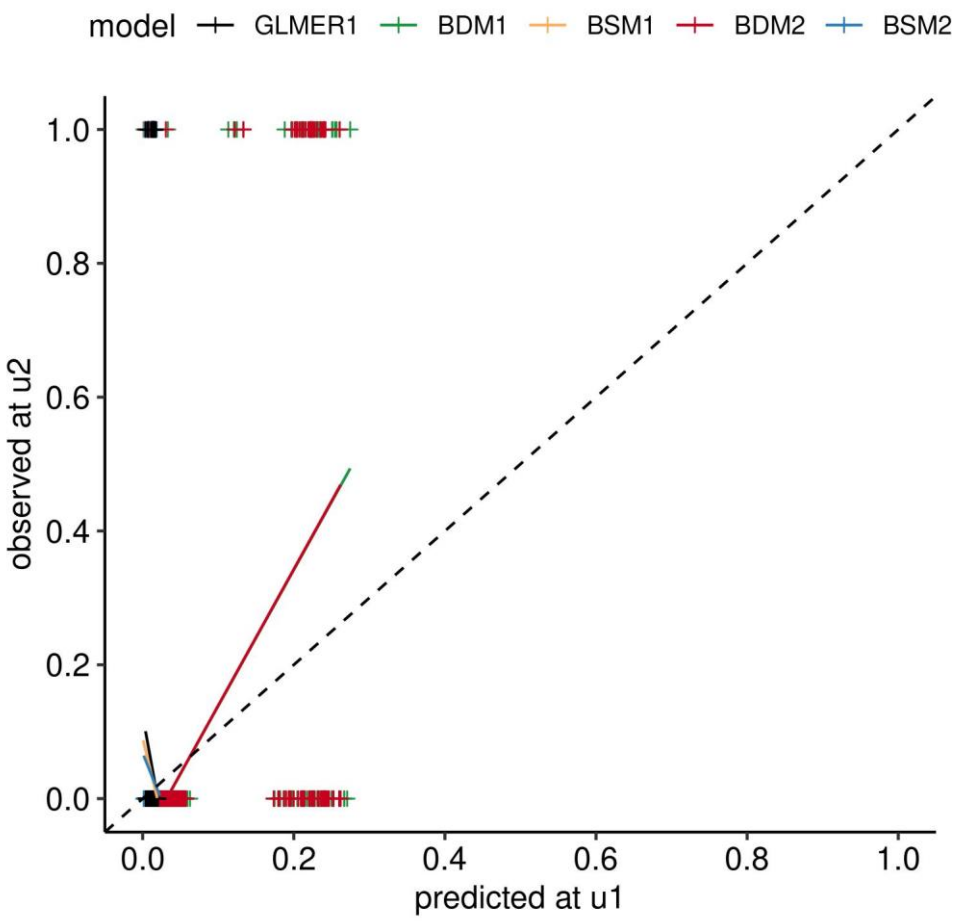


Figure 8. Calibration of Bayesian models vs. GLMER1 predicting U2 outcome using U1 data in PEACHES study

For prediction at later future visits, calibration of models fitted at different prediction visits is compared. Since the same GLMER1 and BSM2 models were used to predict outcome at all future visits, calibration of these two models is the same, regardless at which visit the prediction was made. Calibration of BSM1 seems to change little over time. BDM1 and BDM2 show improvement in calibration with increasing number of past visits. Figure 9 shows calibration of Bayesian models vs. GLMER1 predicting U9 outcome using data from U1 to U8, respectively.

Full results of this analysis can be found online in the folder `/supplementary_material/3.1.3`.

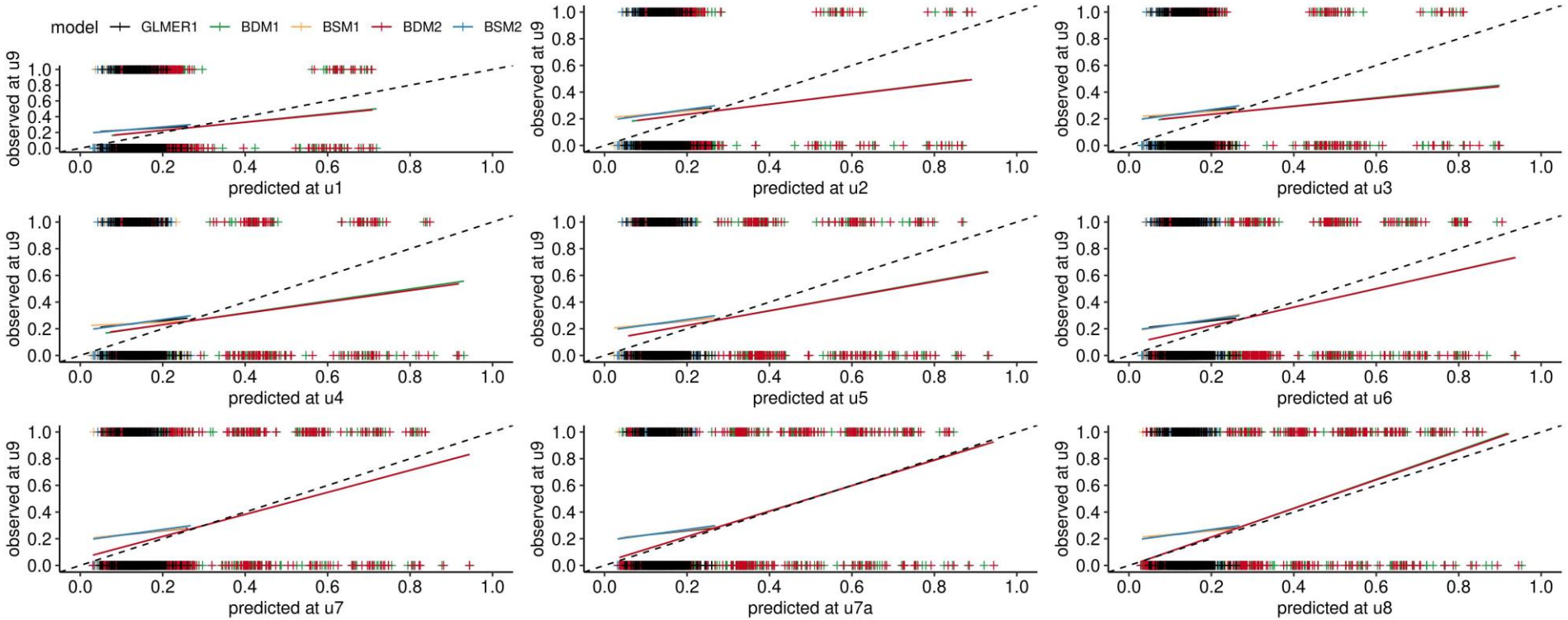


Figure 9. Calibration of Bayesian models vs. GLMER1 predicting U9 outcome using cumulated data from U1 until U8 in PEACHES study

3.1.4 Scaled Brier Score of Bayesian models vs. GLMER1

Scaled Brier Score of models predicting future outcome using data from past visit(s) (Figure 10) show that using accumulated data from visit U5 or later can improve the usefulness of the Bayesian models BDM1 and BDM2 vs. GLMER1. The Scaled Brier Score of other models BSM1, GLMER1 and BSM2 stay below 0, which indicates that these models are not more useful than an uninformative model in predicting future outcome.

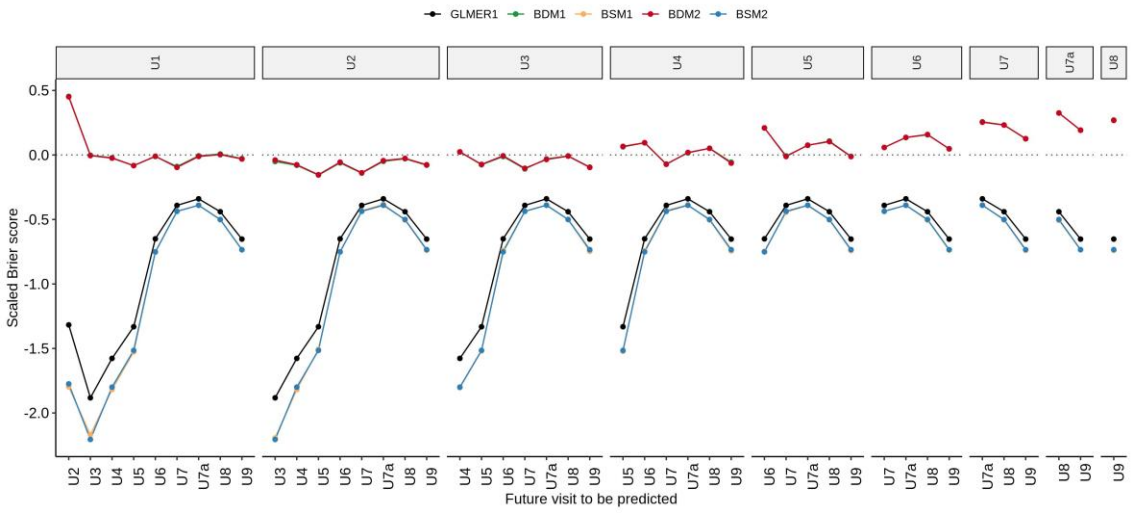


Figure 10. Scaled Brier Score of Bayesian models vs. GLMER1 over time in PEACHES study

3.1.5 Brier Skill Score

3.1.5.1 Brier Skill Score of Bayesian models vs. GLMER1

Brier Skill Score of the four Bayesian models over time (shown in

Figure 11) reveal that overall, BDM1 and BDM2 overperform GLMER1. This overperformance is not consistent if prediction was made at early visits (U1, U2, and U3), while prediction made at U4 or later with BDM1 and BDM2 shows consistent improvement comparing with GLMER1. BSM1 and BSM1 show consistently similar performance with GLMER1.

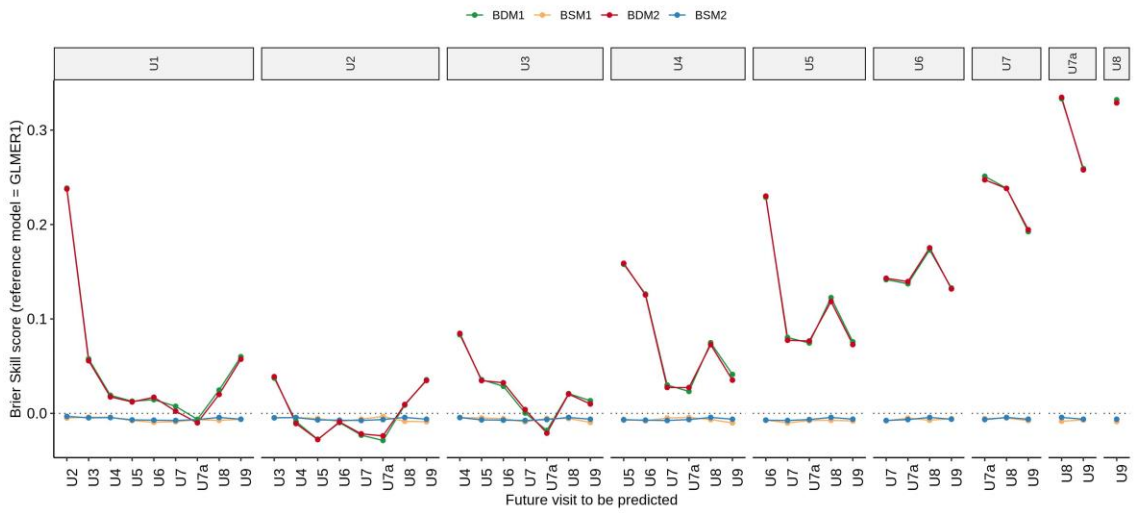


Figure 11. Brier Skill Score of Bayesian models vs. GLMER1 over time in PEACHES study

3.1.5.2 Brier Skill Score with “U1-model” as reference

Figure 11 shows Brier Skill Score of four Bayesian models and GLMER1 over time. Comparing with model trained at U1 – “the “U1-model”, both BDM1 and BDM2 models tend to show overperformance at later visits. For future visits U4, U5, and U6, overperformance can be observed when prediction was made at U3 or later. For future visits U7, U7a, and U8, overperformance starts with prediction made at U4. For future visit U9, models trained with data from U5 or later overperform the “U1-model”. GLMER1, BSM1, and BSM2 show comparatively no improvement in prediction performance over time.

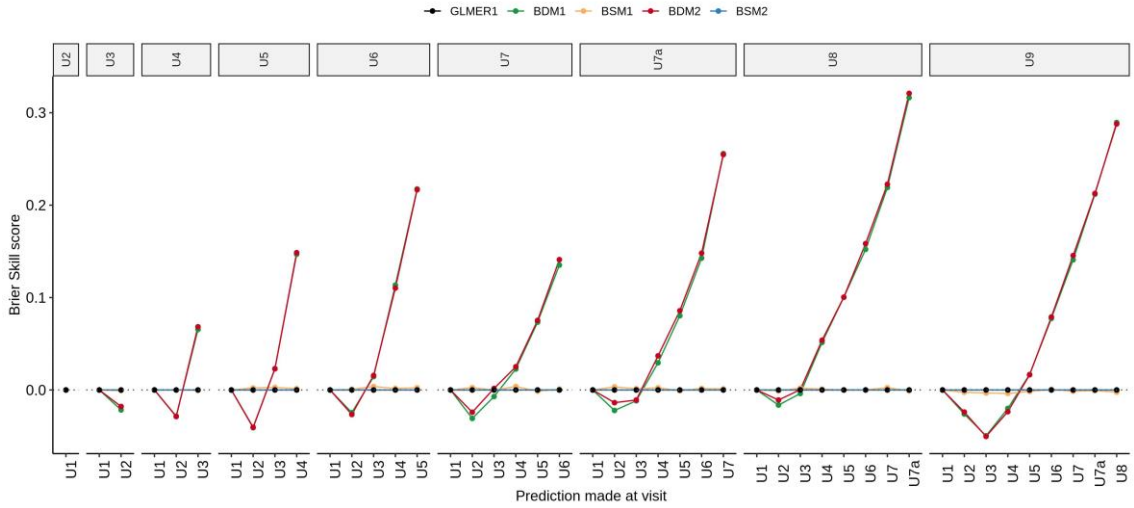


Figure 12. Brier Skill Score vs. “U1-model” over time in PEACHES study

3.1.6 Results at individual level - PEACHES study

While models BSM1, BSM2, and GLMER1 assume random intercept = 0 for each of the validation IDs, BDM1 and BDM2 estimate and update the random intercept of this ID at each prediction visit. Figure 13 shows the distribution of random intercepts in 1,000 LOOCV validation data sets. At earlier visits (U1, U2, U3, and U4), the distribution of the random intercepts estimated by BDM1 and BDM2 seem to be less varying. At later visits, variances of random intercepts become greater. From U5, even though the mean of this distributions is around 0, the random intercepts of different IDs seem to be more varying.

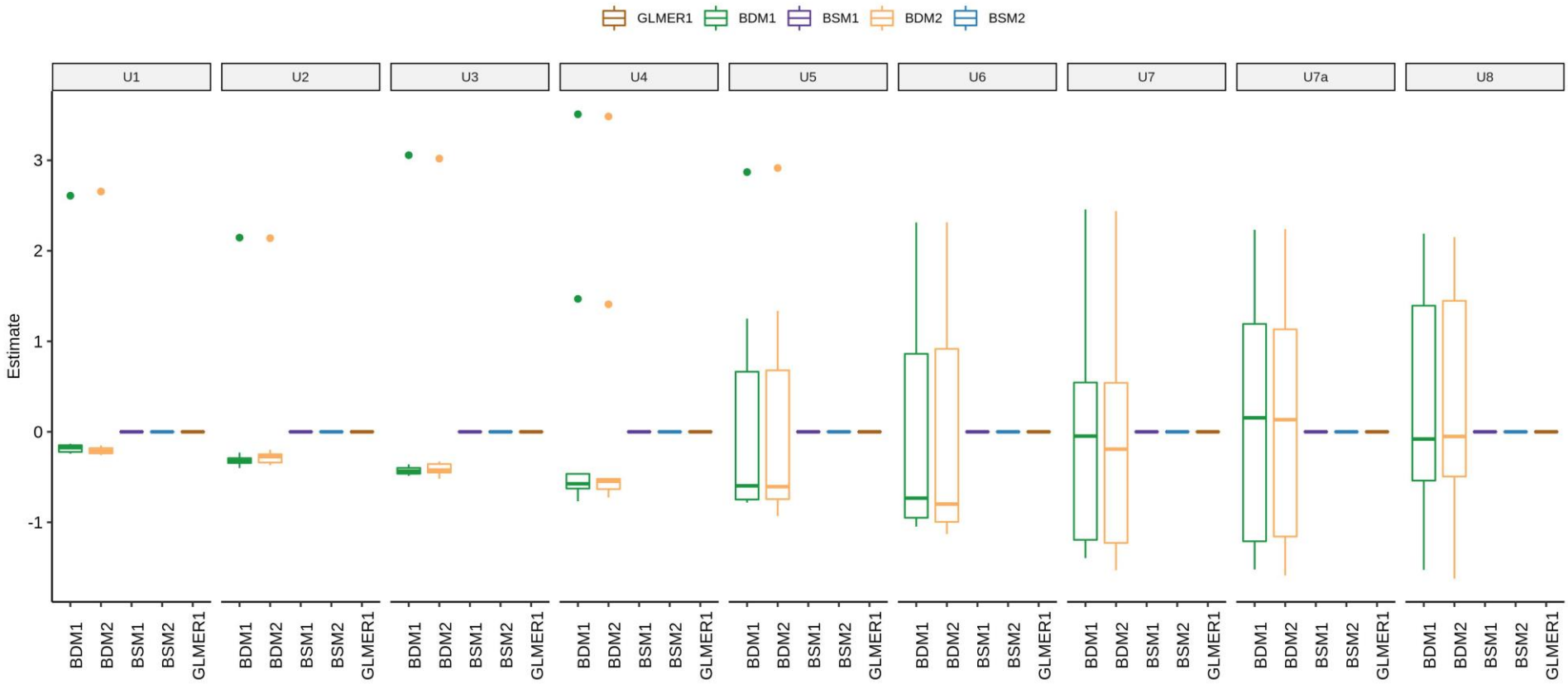


Figure 13. Distributions of estimated random intercepts over time –Bayesian models vs. GLMER1 in PEACHES study

The following plots show the development of the estimates for random intercept and fixed effects as well as predicted outcomes for the LOOCV with a specific ID, ID number 2, dependent of prediction visits. Results of other selected IDs can be found online in the folder `/supplementary_material/3.1.6`.

It can be seen in Figure 14 that random intercept of this ID estimated with GLMER1, BSM1, and BSM2, stays equal 0, while BDM1 and BDM2 updated this estimate at each new prediction visit. The expected value of this estimate starts near 0 and continually decreases from U1 to U7 and moves back near 0 at U7a and U8. The standard error for ID 2's random intercept was obtained from the GLMER1 model and serves as the starting point for BDM1 and BDM2. With increasing number of previous visits, the standard error seems to decrease.

Figure 15 shows the updating process of the fixed effects over time. Expected values of ID 2's fixed effects stay the same for all models but BDM2. These value for BSM1, BSM2, and BDM1 are exact the same (over time) as expected, due to their defined model specification. Slight differences between the expected values estimated by the Bayesian models and GLMER1 are expected. The expected values estimated by BDM2 seem to fluctuate around those estimated by BDM1, BSM1, BSM2. No timely trend can be seen for the expected values of BDM2.

Outcome predicted at different prediction visits are shown in Figure 16. For BSM2 and GLMER1 the courses of the predicted outcome stay the same over time, since no updating was done. BSM1 shows unremarkable changes of the predicted outcome over time. BDM1 and BDM2 update the predicted outcome according to available data at the prediction visit. The development of ID 2's random intercept seems to directly associate with the development of this ID's outcome over time.

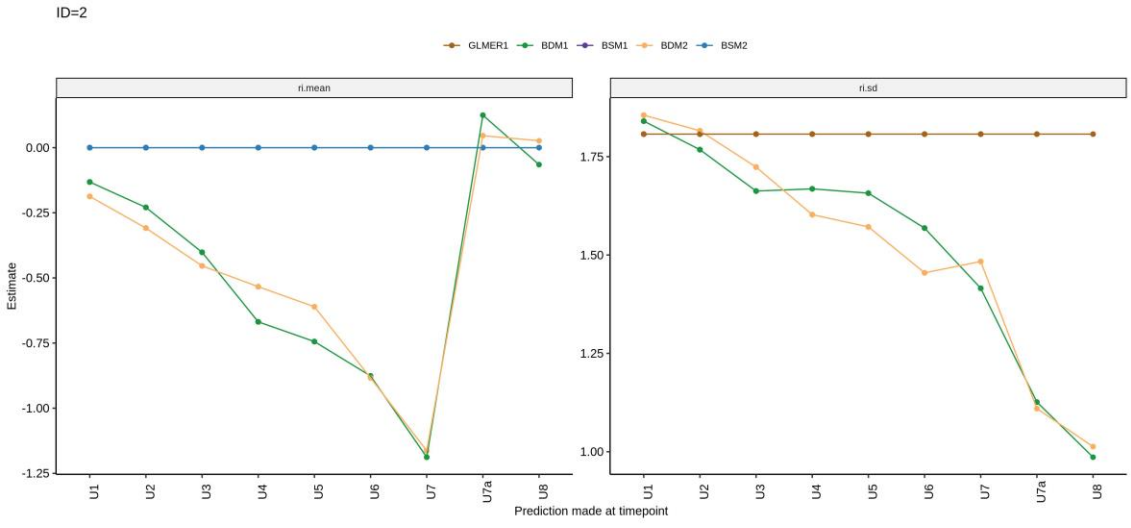


Figure 14. Random intercept estimated by Bayesian models vs. GLMER1 over time in LOOCV with validation ID = 2 (left: expected value, right: standard error)

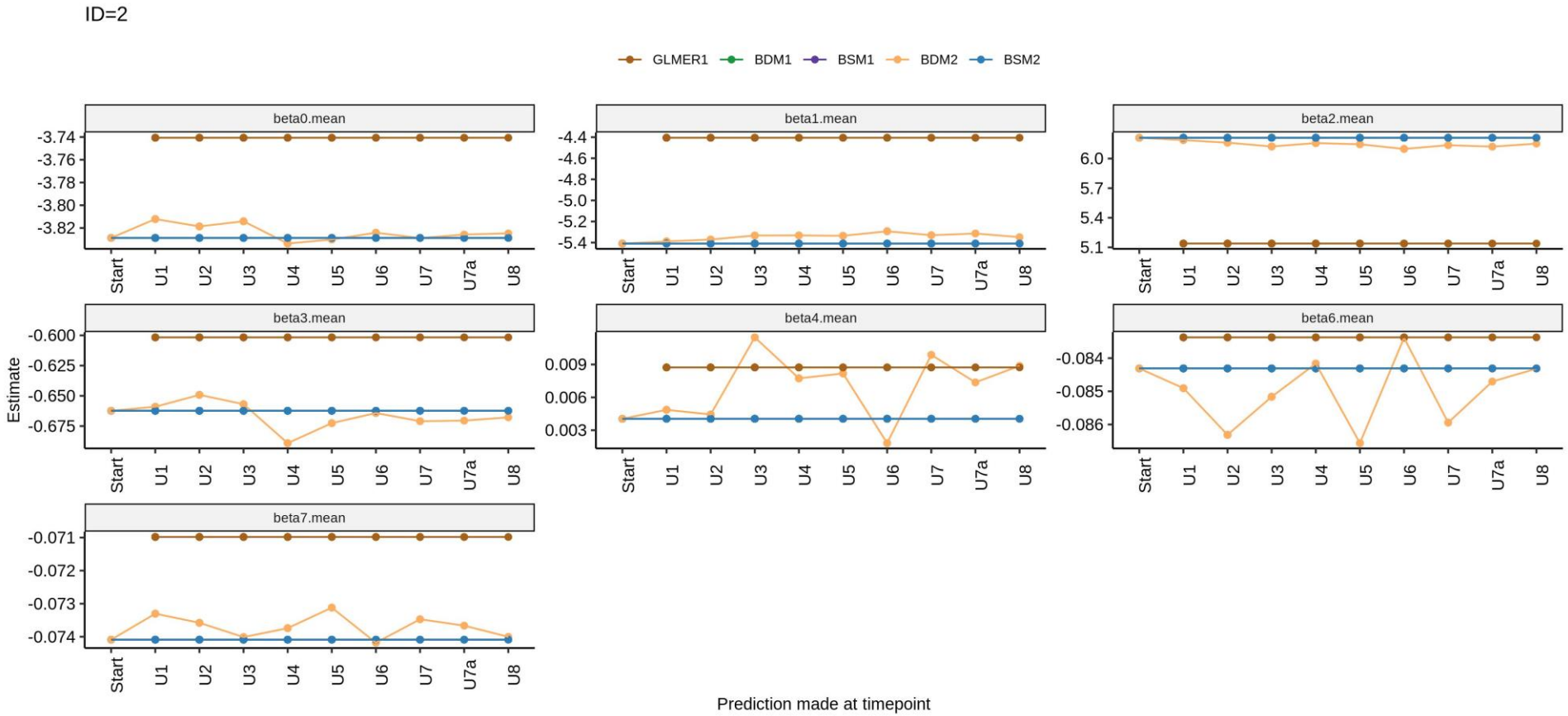


Figure 15. Fixed effects estimated by Bayesian models vs. GLMER1 over time in LOOCV with validation ID = 2

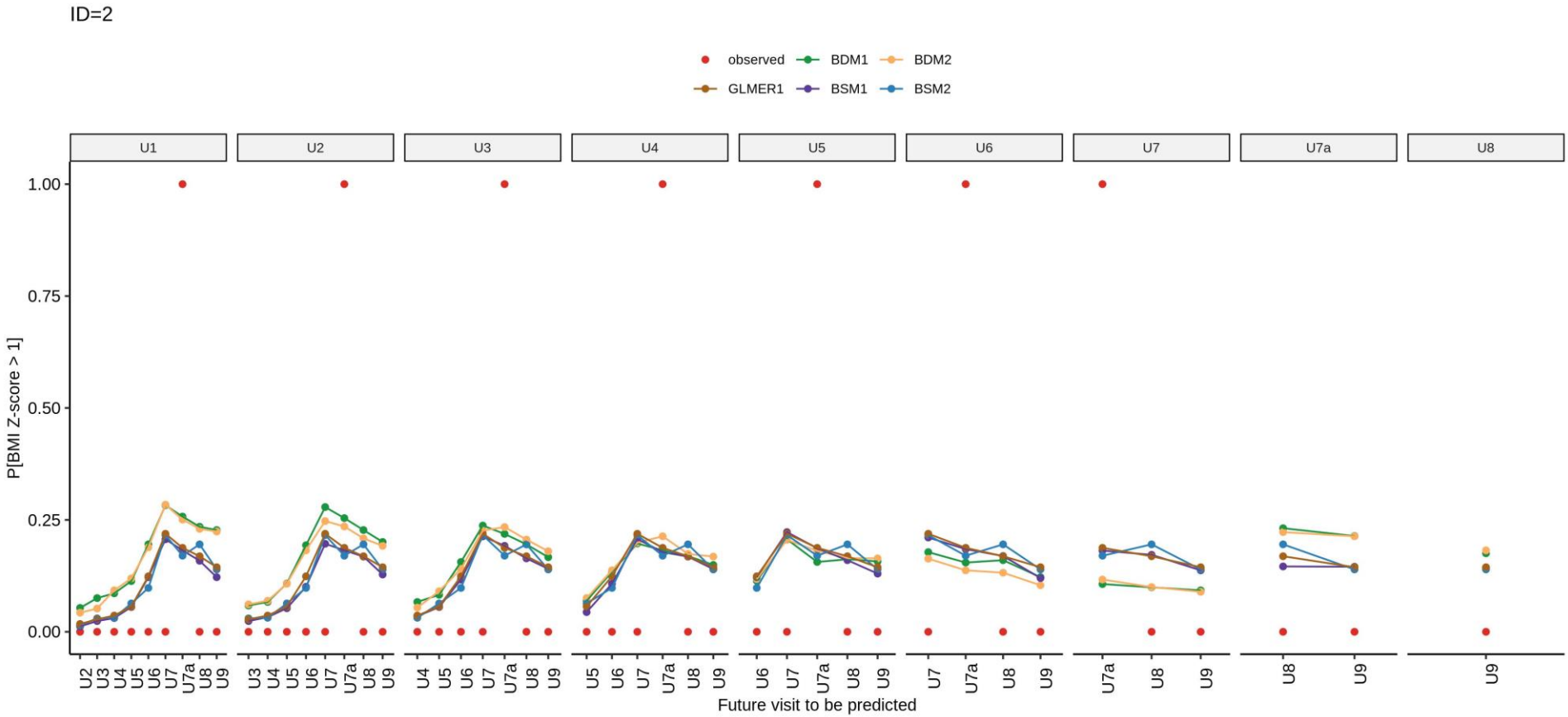


Figure 16. Outcome predicted by Bayesian models vs. GLMER1 over time in LOOCV with validation ID = 2

3.2 Simulation main study

3.2.1 Simulated data used for analyses

The following plot presents the mean risk of overweight in the simulated dataset with 1,000 individuals.

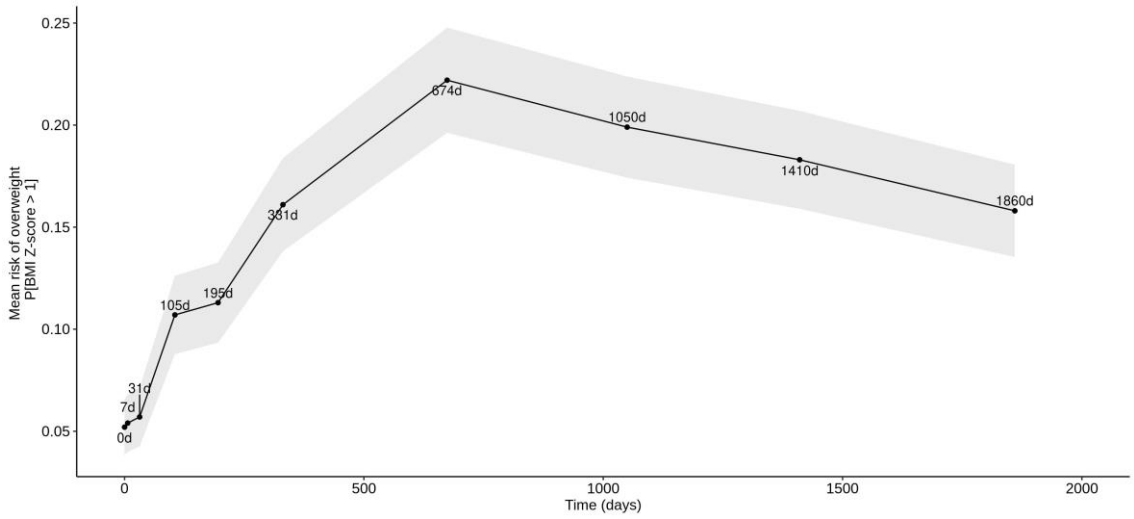


Figure 17. Mean risk of overweight in large simulation data set

3.2.2 Relative prediction error

3.2.2.1 Relative prediction error of GLMER1 vs. BSM1

On average, the relative difference of prediction error between these two models GLMER1 vs. BSM1 is $e^{0.211}$ = 1.235 at U1. This indicates that BSM1 predicts better than GLMER1 at U1. This difference stays the same, regardless of number of previous months. Table 29 shows the estimated coefficients of the LMM. Figure 18 shows the distribution of LRPE of GLMER1 vs. BSM1. The proportion of variance that can be explained by different visits is $\frac{0.087^2}{0.125^2+0.087^2+0.199^2} * 100\% = 12.054\%$. The distribution of LRPE over time, in relationship with number of past visits is presented in the box plots in Figure 18. The box plots show that number of previous visits seems not to influence the LRPE of GLMER1 vs. BSM1 over time.

Table 29. LRPE of GLMER1 vs. BSM1 in simulation main study – output of LMM

group	term	estimate	std.error	statistic	df	p.value
	(Intercept)	0.211	0.029	7.252	7.982	8.886E-05
udata:val.id	sd__(Intercept)	0.125				
udata	sd__(Intercept)	0.087				
Residual	sd__Observation	0.199				

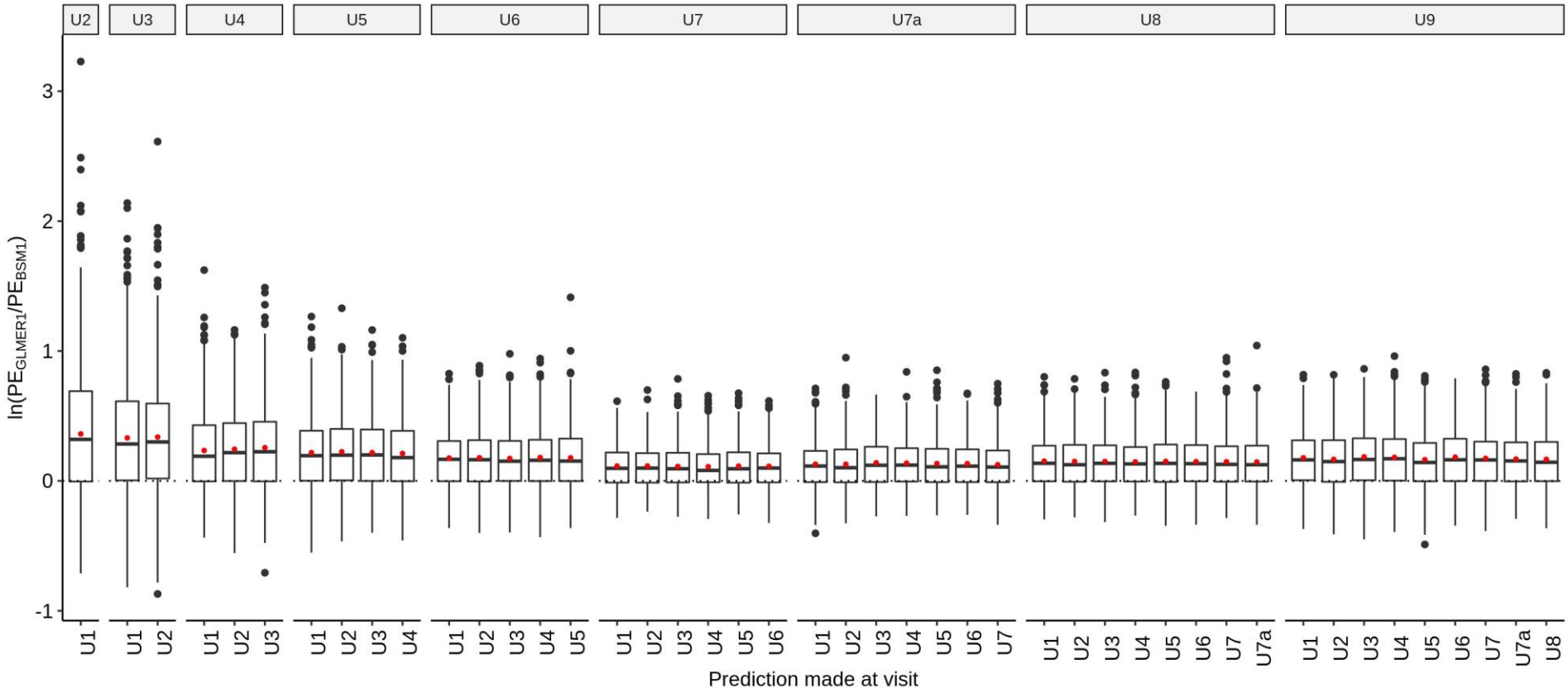


Figure 18. LRPE of GLMER1 vs. BSM1 over time in simulation main study

3.2.2.2 Relative prediction error of GLMER1 vs. BSM2

Similar as the results shown in 3.2.2, BSM2 predicts better than GLMER1 at U1. RPE of GLMER1 vs. BSM2 at this visit is $e^{0.339} = 1.404$. This difference stays the same, regardless of number of previous visits. The following table shows the estimated coefficients of the LMM. Figure 19 shows the distribution of LRPE between GLMER1 and BSM2.

Table 30. LRPE of GLMER1 vs. BSM2 in simulation study – output of LMM

group	term	estimate	std.error	statistic	df	p.value
	(Intercept)	0.339	0.010	33.743	0.974	2.057E-02
val.id	sd__(Intercept)	0.166				
Residual	sd__Observation	5.074E-08				

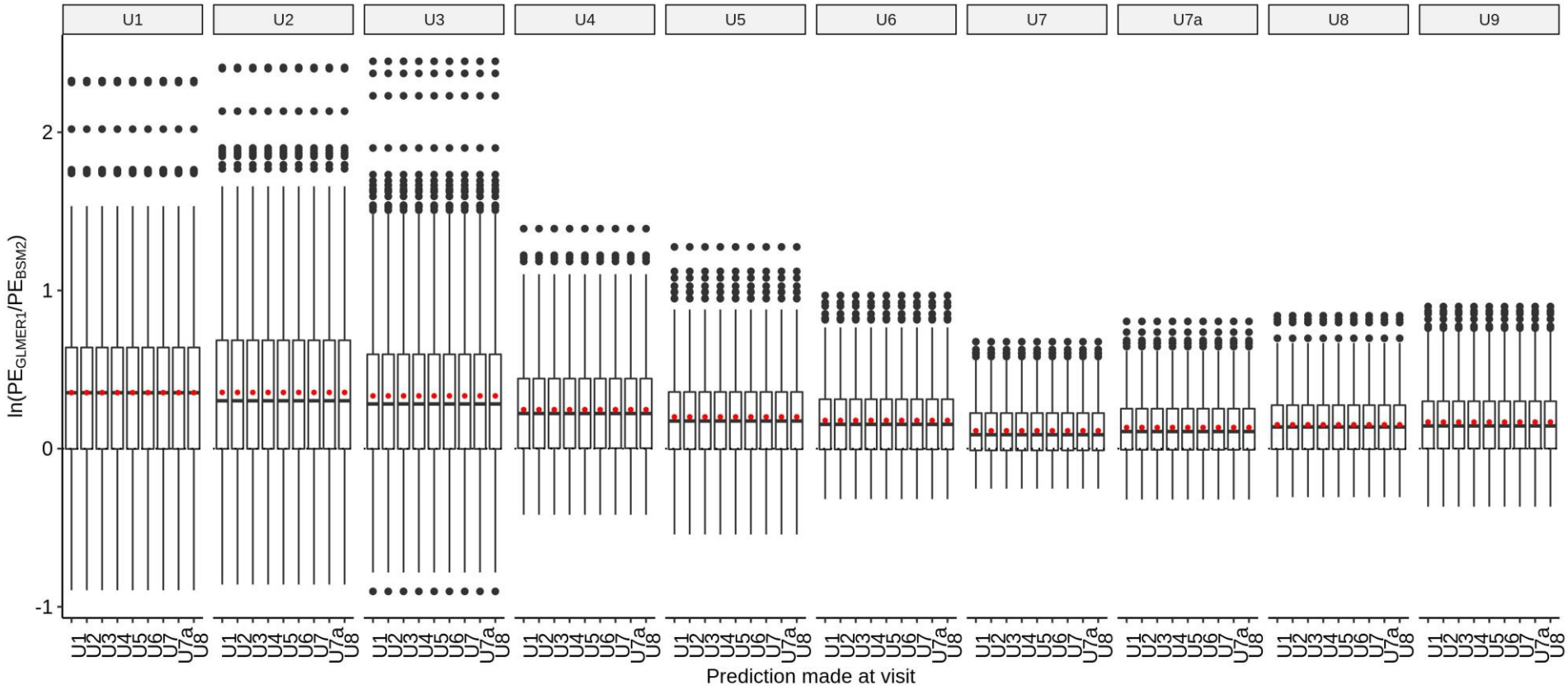


Figure 19. LRPE of GLMER1 vs. BSM2 over time in simulation main study

3.2.2.3 Relative prediction error of GLMER1 vs. BDM1

On average, GLMER1 overperforms BDM1 at U1. However, this difference attenuates over time. Non-linear time effects are present. The proportion of variance that can be explained by different visits is $\frac{0.510^2}{0.924^2+0.510^2+0.586^2} * 100\% = 18.848\%$. In Figure 20, the box plots move upward with increasing number of previous months, and their mean/median seem lie above 0 when data from U7 or later was used for prediction. This observation supports the results of the LMM presented above (Table 31).

Table 31. LRPE of GLMER1 vs. BDM1 in simulation main study – output of LMM

group	term	estimate	std.error	statistic	df	p.value
	(Intercept)	-1.235	0.170	-7.247	7.998E+00	8.839E-05
	months_from_u1	0.116	0.003	42.838	3.620E+04	0
	I(months_from_u1^2)	-7.250E-03	3.304E-04	-21.944	36197.9457	4.863E-106
	I(months_from_u1^3)	2.038E-04	1.272E-05	16.024	36198.7072	1.3783E-57
	I(months_from_u1^4)	-1.967E-06	1.482E-07	-13.269	3.620E+04	4.3498E-40
udata:val.id	sd__(Intercept)	0.924				
udata	sd__(Intercept)	0.510				
Residual	sd__Observation	0.586				

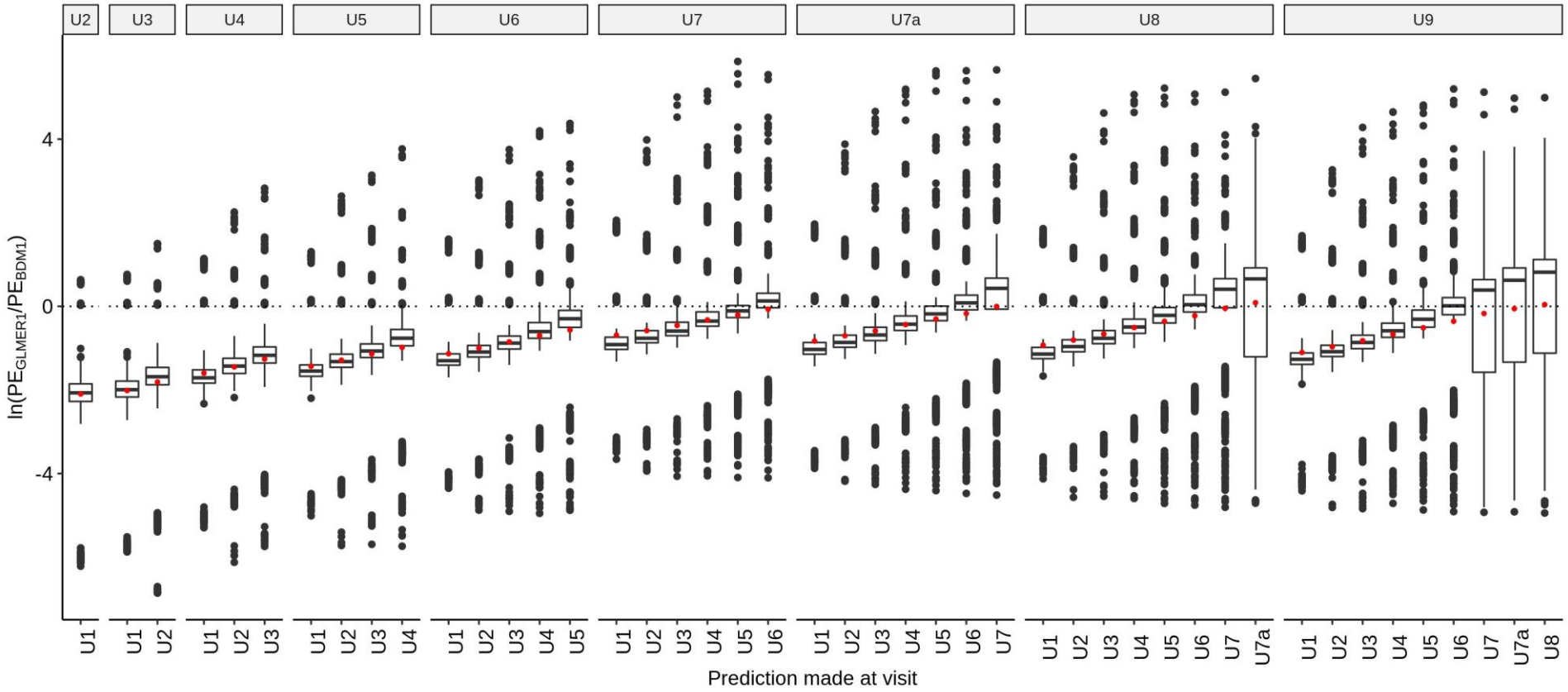


Figure 20. LRPE of GLMER1 vs. BDM1 over time in simulation main study

3.2.2.4 Relative prediction error of GLMER1 vs. BDM2

Similar as the results of BDM1 (3.2.2.3), at U1, GLMER1 overperforms BDM2 on average. This effect attenuates over time. Non-linear time effects are present. The proportion of variance that can be explained by different visits is $\frac{0.518^2}{0.927^2+0.518^2+0.590^2} * 100\% = 18.182\%$. In Figure 21, the box plots move upward with increasing number of previous months, and their mean/median seem lie above 0 when data from U7 or later was used for prediction. This observation supports the results of the LMM presented above (Table 32).

Table 32. LRPE of GLMER1 vs. BDM2 in simulation study – output of LMM

group	term	estimate	std.error	statistic	df	p.value
	(Intercept)	-1.246	0.173	-7.195	7.998	9.304E-05
	months_from_u1	0.119	0.003	43.541	36198.520	0.000E+00
	I(months_from_u1^2)	-0.007	3.324E-04	-22.554	36199.485	7.334E-112
	I(months_from_u1^3)	2.110E-04	1.279E-05	16.490	36200.267	7.1953E-61
	I(months_from_u1^4)	-2.032E-06	1.491E-07	-13.625	36200.806	3.6076E-42
udata:val.id	sd__(Intercept)	0.927				
udata	sd__(Intercept)	0.518				
Residual	sd__Observation	0.590				

3.2.2.5 Overall time effect on RPE

Overall time effect (see 2.8) of the four Bayesian models is shown in the following table, when incorporating the non-linearity of time effects. RPE of GLMER1 versus BSM1 and BSM2 was shown to be constant regardless of number of previous months, while until 48 months after birth, RPE of GLMER1 versus BDM1 and BDM2 seems to increase with increasing number of months from U1. BDM1 and BDM2 do not overperform GLMER1 at any timepoints.

Table 33. RPE of GLMER1 vs. Bayesian models in simulation main study

Number of months from U1	RPE of GLMER1 vs.			
	BSM1	BSM2	BDM1	BDM2
0	1.235	1.404	0.291	0.288
1	1.235	1.404	0.324	0.322
3	1.235	1.404	0.388	0.386
6	1.235	1.404	0.468	0.467
12	1.235	1.404	0.561	0.560
24	1.235	1.404	0.628	0.621
36	1.235	1.404	0.776	0.769
48	1.235	1.404	0.757	0.756
60	1.235	1.404	0.157	0.153

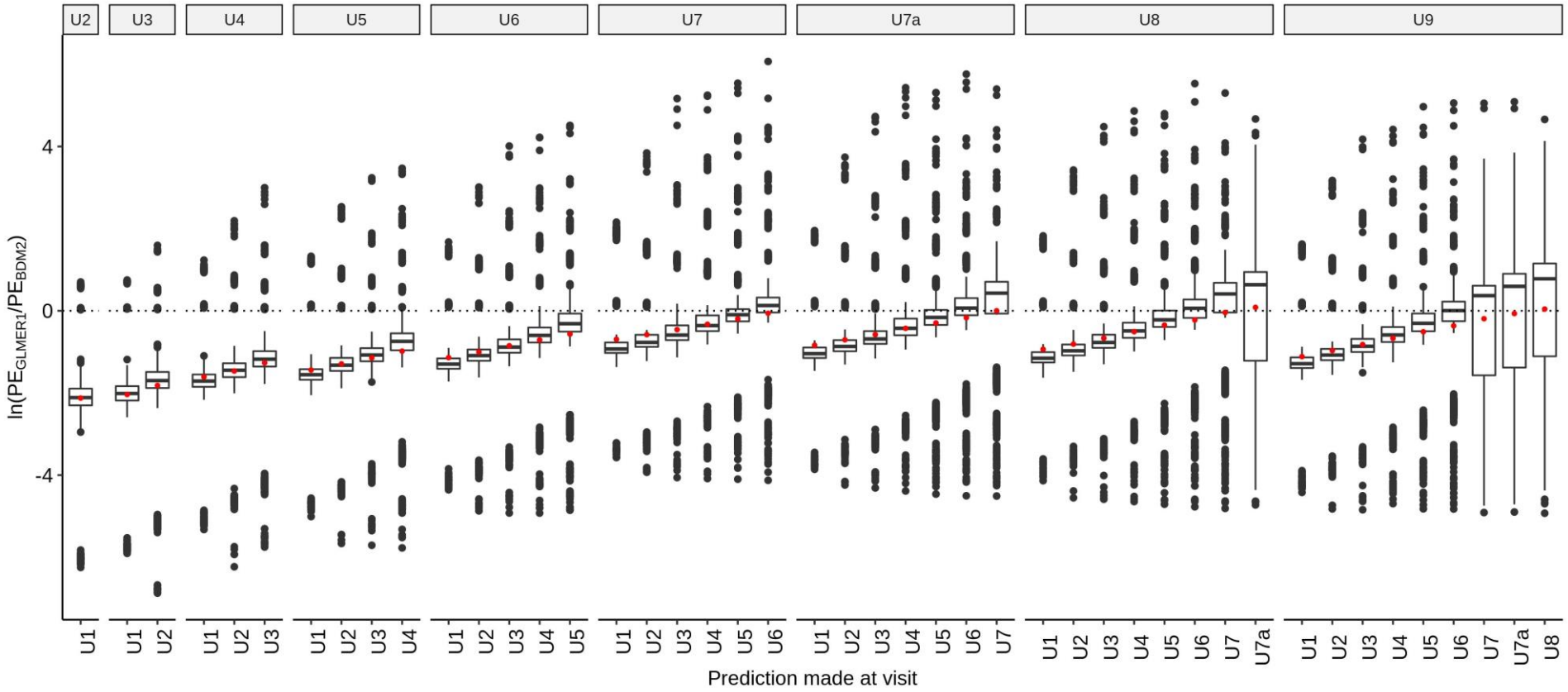


Figure 21. LRPE of BDM2 vs. GLMER1 over time in simulation study

3.2.3 Brier Score of Bayesian models vs. GLMER1

Table 34. Brier Score of dynamic models vs. GLMER1 over time

Prediction made at timepoint	Future timepoint to be predicted	Brier Score				
		GLMER1	BSM1	BSM2	BDM1	BDM2
U1	U2	0.053	0.053	0.053	0.049	0.049
U1	U3	0.055	0.055	0.055	0.053	0.053
U1	U4	0.100	0.101	0.101	0.089	0.089
U1	U5	0.105	0.105	0.106	0.093	0.093
U1	U6	0.143	0.144	0.144	0.131	0.131
U1	U7	0.182	0.184	0.184	0.161	0.161
U1	U7a	0.169	0.170	0.170	0.152	0.152
U1	U8	0.159	0.160	0.160	0.137	0.138
U1	U9	0.141	0.142	0.142	0.126	0.127
U2	U3	0.055	0.055	0.055	0.051	0.050
U2	U4	0.100	0.101	0.101	0.089	0.089
U2	U5	0.105	0.105	0.106	0.090	0.090
U2	U6	0.143	0.144	0.144	0.126	0.126
U2	U7	0.182	0.184	0.184	0.156	0.156
U2	U7a	0.169	0.170	0.170	0.144	0.144
U2	U8	0.159	0.160	0.160	0.132	0.133
U2	U9	0.141	0.142	0.142	0.119	0.119
U3	U4	0.100	0.101	0.101	0.083	0.083
U3	U5	0.105	0.105	0.106	0.088	0.089
U3	U6	0.143	0.144	0.144	0.119	0.119
U3	U7	0.182	0.184	0.184	0.150	0.151
U3	U7a	0.169	0.170	0.170	0.139	0.139
U3	U8	0.159	0.160	0.160	0.125	0.126
U3	U9	0.141	0.142	0.142	0.115	0.115
U4	U5	0.105	0.105	0.106	0.088	0.088
U4	U6	0.143	0.144	0.144	0.117	0.116
U4	U7	0.182	0.184	0.184	0.146	0.146
U4	U7a	0.169	0.170	0.170	0.134	0.134
U4	U8	0.159	0.160	0.160	0.120	0.120
U4	U9	0.141	0.142	0.142	0.109	0.110
U5	U6	0.143	0.144	0.144	0.115	0.115
U5	U7	0.182	0.184	0.184	0.143	0.143
U5	U7a	0.169	0.170	0.170	0.130	0.130
U5	U8	0.159	0.161	0.160	0.115	0.115
U5	U9	0.141	0.142	0.142	0.106	0.106
U6	U7	0.182	0.184	0.184	0.135	0.135
U6	U7a	0.169	0.170	0.170	0.125	0.126
U6	U8	0.159	0.160	0.160	0.112	0.112
U6	U9	0.141	0.142	0.142	0.102	0.102
U7	U7a	0.169	0.171	0.170	0.121	0.122
U7	U8	0.159	0.160	0.160	0.108	0.108
U7	U9	0.141	0.142	0.142	0.099	0.099
U7a	U8	0.159	0.160	0.160	0.105	0.105
U7a	U9	0.141	0.142	0.142	0.096	0.096
U8	U9	0.141	0.142	0.142	0.096	0.096

Brier Score of four Bayesian models and GLMER1 are shown in Table 34. Their courses over time are presented in Figure 22. From U1, two groups of models seem to distinguish in their prediction performance. BDM1 and BDM2 overperform BSM1, BSM2, and GLMER1. This over-performance seems to increase with increasing amount of data from past visits that was used for prediction.

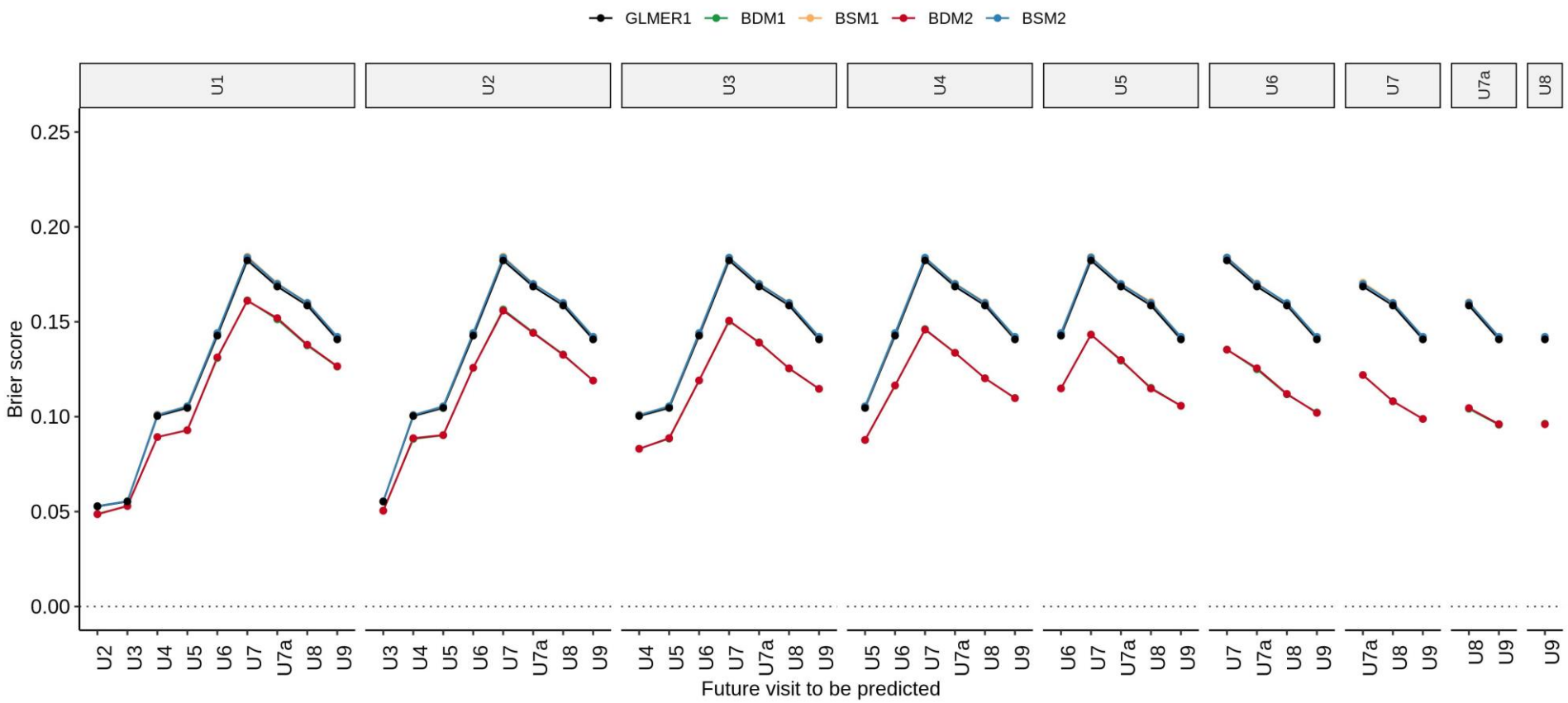


Figure 22. Brier Score of Bayesian models vs. GLMER1 over time in simulation main study

3.2.4 Calibration of Bayesian models vs. GLMER1

Calibration slopes and intercepts of Bayesian models vs. GLMER1 in the simulation main study are shown in the following table. Calibration slope close to 0 and intercept close to 1 indicate good calibration.

Table 35. Calibration slopes and intercepts of Bayesian models vs. GLMER1 – simulation main study

Future visit	Prediction visit	GLMER1 Slope	GLMER1 Intercept	BSM1 Slope	BSM1 Intercept	BSM2 Slope	BSM2 Intercept	BDM1 Slope	BDM1 Intercept	BDM2 Slope	BDM2 Intercept
U2	U1	-75.783	1.070	-4.500	0.105	-3.106	0.089	0.954	0.006	0.957	0.005
U3	U1	21.476	-0.323	2.195	0.024	1.351	0.036	0.646	0.017	0.623	0.018
U3	U2	21.476	-0.323	0.955	0.043	1.351	0.036	0.758	0.010	0.792	0.007
U4	U1	29.997	-0.928	5.269	-0.054	11.471	-0.243	1.028	0.005	1.031	0.004
U4	U2	29.997	-0.928	8.171	-0.142	11.471	-0.243	0.871	0.020	0.864	0.020
U4	U3	29.997	-0.928	6.546	-0.091	11.471	-0.243	1.041	0.004	1.038	0.004
U5	U1	9.938	-0.344	0.155	0.107	-1.597	0.179	1.041	-0.012	1.016	-0.010
U5	U2	9.938	-0.344	2.877	-0.005	-1.597	0.179	0.935	-0.001	0.937	-0.001
U5	U3	9.938	-0.344	1.521	0.051	-1.597	0.179	0.925	0.002	0.908	0.003
U5	U4	9.938	-0.344	2.773	-0.001	-1.597	0.179	0.853	0.009	0.854	0.009
U6	U1	10.847	-0.604	4.037	-0.097	3.692	-0.074	0.746	0.043	0.716	0.047
U6	U2	10.847	-0.604	3.769	-0.079	3.692	-0.074	0.809	0.032	0.822	0.029
U6	U3	10.847	-0.604	4.527	-0.129	3.692	-0.074	0.947	0.011	0.939	0.012
U6	U4	10.847	-0.604	5.744	-0.205	3.692	-0.074	0.890	0.018	0.903	0.015
U6	U5	10.847	-0.604	5.158	-0.168	3.692	-0.074	0.896	0.019	0.895	0.018
U7	U1	4.039	-0.270	1.456	0.057	3.107	-0.131	1.058	-0.012	1.036	-0.008
U7	U2	4.039	-0.270	1.342	0.070	3.107	-0.131	0.985	0.002	0.975	0.004
U7	U3	4.039	-0.270	2.776	-0.093	3.107	-0.131	1.005	0.000	0.992	0.002
U7	U4	4.039	-0.270	2.973	-0.116	3.107	-0.131	0.949	0.010	0.955	0.008
U7	U5	4.039	-0.270	1.856	0.012	3.107	-0.131	0.952	0.011	0.941	0.013
U7	U6	4.039	-0.270	2.327	-0.042	3.107	-0.131	1.000	-0.001	0.999	-0.001
U7a	U1	0.011	0.198	-0.455	0.242	1.385	0.067	0.860	0.026	0.858	0.026
U7a	U2	0.011	0.198	0.344	0.166	1.385	0.067	0.957	0.005	0.952	0.006
U7a	U3	0.011	0.198	0.574	0.144	1.385	0.067	0.974	0.004	0.980	0.002
U7a	U4	0.011	0.198	0.437	0.157	1.385	0.067	0.962	0.003	0.960	0.004
U7a	U5	0.011	0.198	0.763	0.126	1.385	0.067	0.982	0.001	0.975	0.002
U7a	U6	0.011	0.198	-0.043	0.203	1.385	0.067	0.992	-0.001	0.986	-0.001
U7a	U7	0.011	0.198	-1.969	0.387	1.385	0.067	0.990	-0.002	0.981	0.000
U8	U1	-0.668	0.242	-0.224	0.201	0.726	0.124	1.119	-0.021	1.086	-0.016
U8	U2	-0.668	0.242	-0.231	0.202	0.726	0.124	1.037	-0.007	1.037	-0.008
U8	U3	-0.668	0.242	-0.536	0.226	0.726	0.124	1.103	-0.018	1.096	-0.017
U8	U4	-0.668	0.242	-1.164	0.277	0.726	0.124	1.071	-0.014	1.064	-0.014
U8	U5	-0.668	0.242	-2.056	0.348	0.726	0.124	1.099	-0.017	1.098	-0.018
U8	U6	-0.668	0.242	1.188	0.087	0.726	0.124	1.095	-0.017	1.084	-0.017
U8	U7	-0.668	0.242	-0.466	0.221	0.726	0.124	1.091	-0.018	1.086	-0.017
U8	U7a	-0.668	0.242	-1.598	0.312	0.726	0.124	1.098	-0.018	1.097	-0.019
U9	U1	-4.565	0.487	-2.949	0.350	-3.297	0.374	0.870	0.018	0.857	0.020
U9	U2	-4.565	0.487	-3.557	0.391	-3.297	0.374	0.970	0.001	0.977	-0.001
U9	U3	-4.565	0.487	-1.711	0.269	-3.297	0.374	1.003	-0.003	1.001	-0.003
U9	U4	-4.565	0.487	-0.134	0.167	-3.297	0.374	1.007	-0.006	0.996	-0.005
U9	U5	-4.565	0.487	-2.441	0.318	-3.297	0.374	1.028	-0.007	1.029	-0.008
U9	U6	-4.565	0.487	-1.824	0.277	-3.297	0.374	1.037	-0.010	1.044	-0.011
U9	U7	-4.565	0.487	-2.632	0.330	-3.297	0.374	1.049	-0.012	1.048	-0.012
U9	U7a	-4.565	0.487	-2.566	0.326	-3.297	0.374	1.064	-0.014	1.060	-0.014
U9	U8	-4.565	0.487	-2.661	0.332	-3.297	0.374	1.026	-0.008	1.028	-0.009

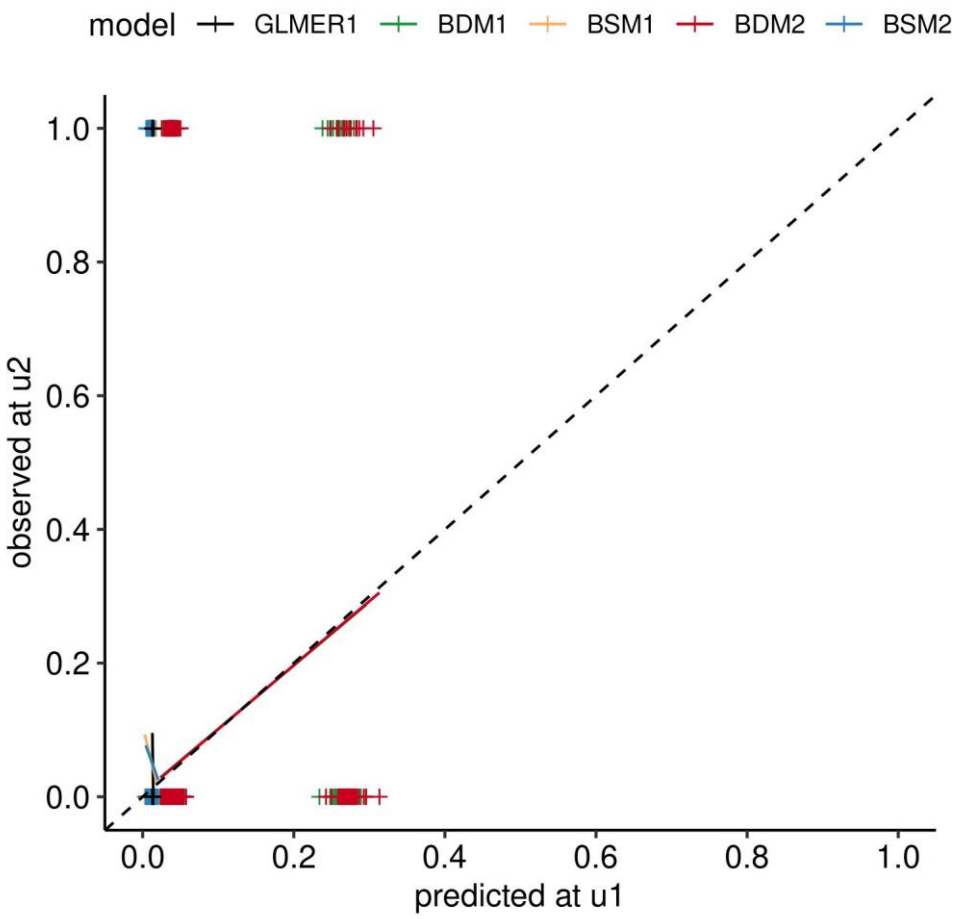


Figure 23. Calibration of Bayesian models vs. GLMER1 predicting U2 outcome using U1 data

Calibration of the prediction models that predicted future outcome at U2 using data of U1 is depicted in Figure 23. BDM1 and BDM2 shows good calibration, while other three models BSM1, BSM2, and GLMER1 show bad calibration. This can also be observed for prediction of outcome at U9, which used accumulated data from U1 to U8, respectively (Figure 23).

Full results of this analysis can be found online in the folder `/supplementary_material/3.2.4`

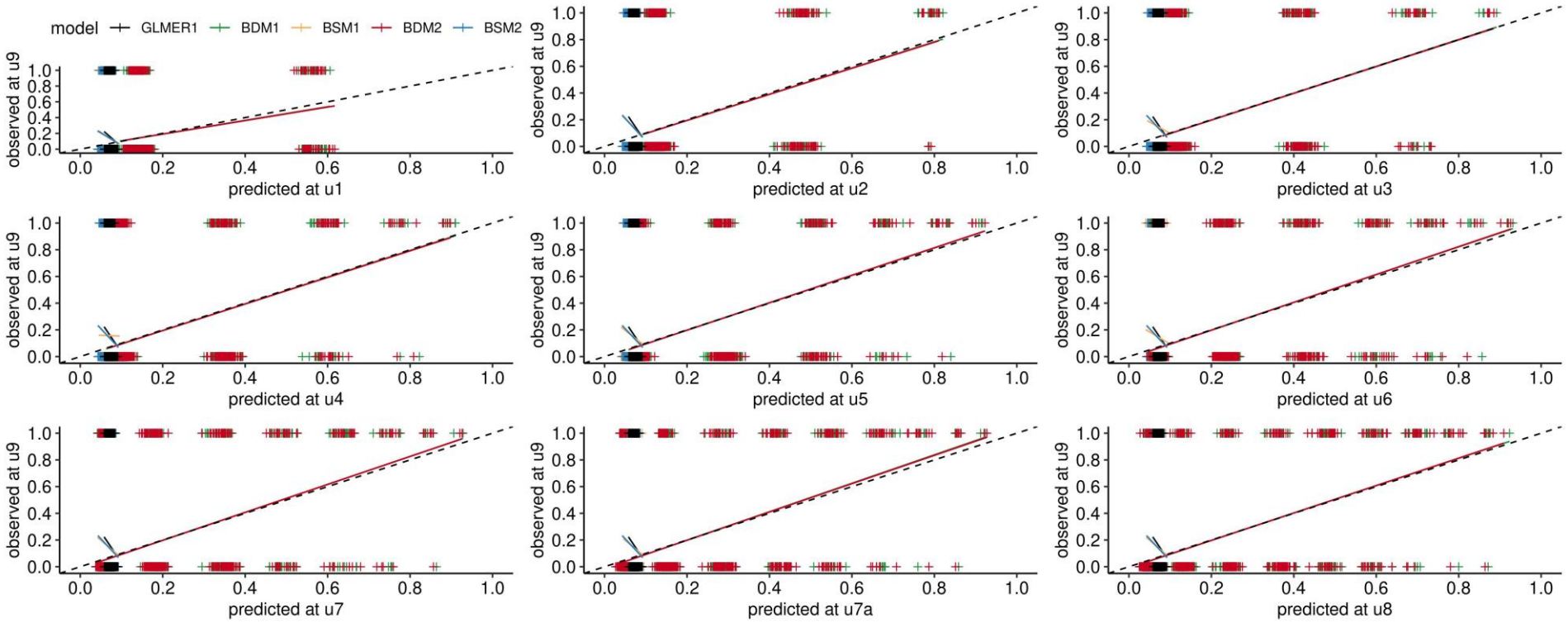


Figure 24. Calibration of Bayesian models vs. GLMER1 predicting U9 outcome using cumulated data from U1 until U8

3.2.5 Scaled Brier Score of Bayesian models vs. GLMER1

As shown in Figure 25, Scaled Brier Score of BDM1 and BDM2 reveal consistent overperformance of these models compared to BSM1, BSM2, and GLMER1 over time. The Scaled Brier Score of BDM1 and BDM2 also show their overperformance compared to a non-informative model, and this overperformance seem to increase over time.

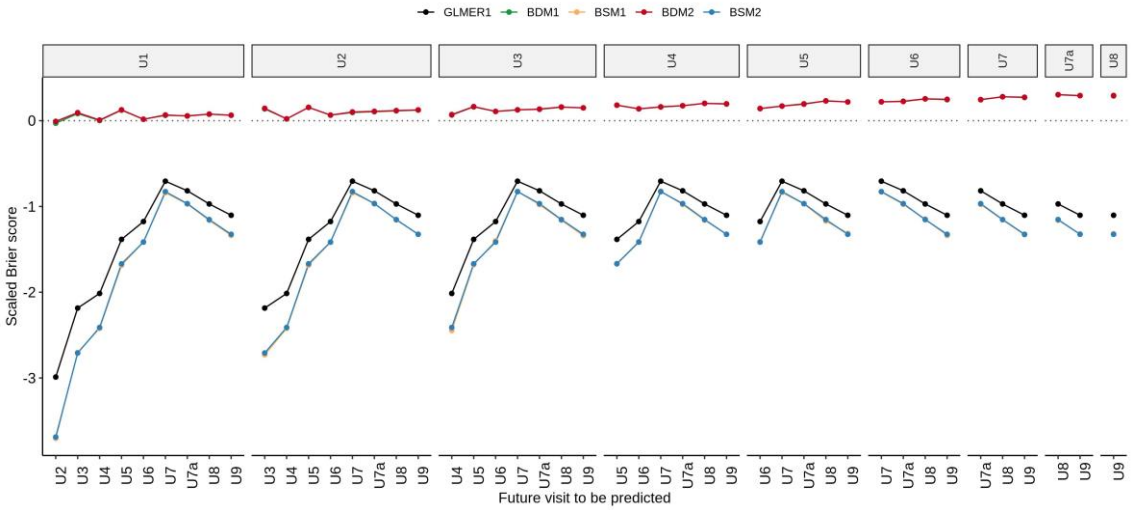


Figure 25. Scaled Brier Score of Bayesian models vs. GLMER1 in simulation main study

3.2.6 Brier Skill Score

3.2.6.1 Brier Skill Score of Bayesian models vs. GLMER1

The consistent overperformance of BDM1 and BDM2 compared to GLMER1 over time can also be seen in Figure 26. BSM1 and BSM2 show similar prediction performance as GLMER1.

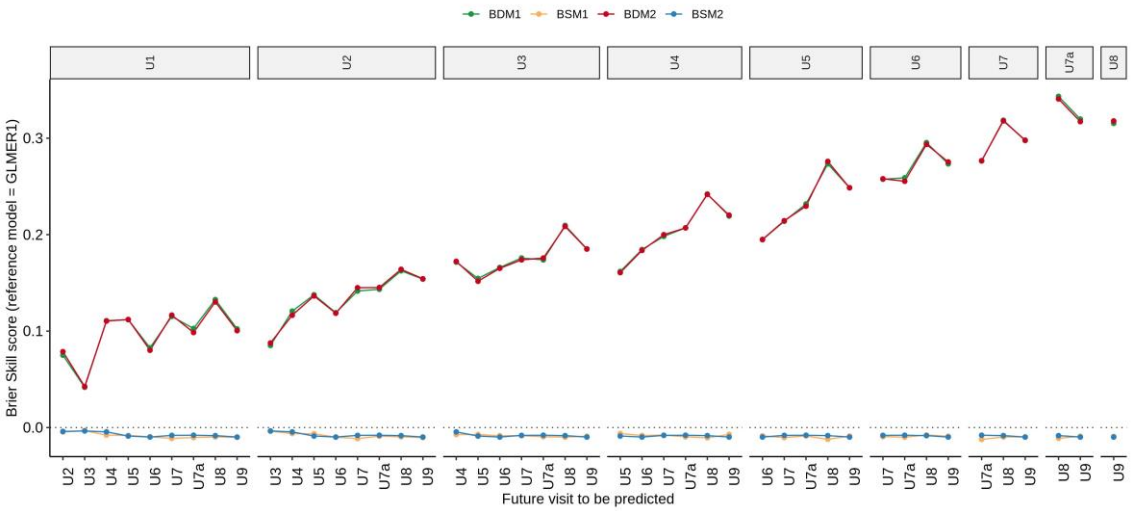


Figure 26. Brier Skill Score of Bayesian models vs. GLMER1 over time in simulation main study

3.2.6.2 Brier Skill Score with “U1-model” as reference

Figure 12 shows Brier Skill score of four Bayesian models and GLMER1 over time. Comparing with model trained at U1 – “the “U1-model”, both BDM1 and BDM2 models show consistent over-performance. GLMER1, BSM1, and BSM2 show comparatively no improvement in prediction performance.

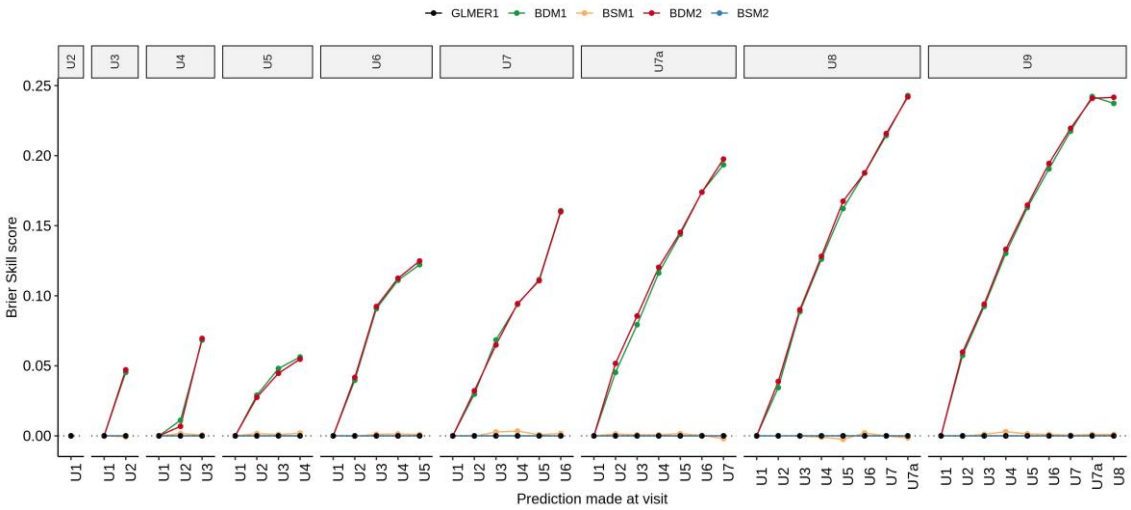


Figure 27. Brier Skill Score of Bayesian models vs. “U1-model” over time in simulation main study

3.2.7 Results at individual level - Simulation main study

While models BSM1, BSM2, and GLMER1 assume random intercept = 0 for each of the validation IDs, BDM1 and BDM2 estimate and update random intercept of this ID at each prediction visit. Figure 28 shows the distribution of random intercepts in 1,000 LOOCV validation data sets. At earlier visits (U1 and U2), the distribution of the random intercepts estimated by BDM1 and BDM2 seem to be less varying. From U3, variances of the random intercepts of different IDs seem to be greater. The means of both distributions for BDM1 and BDM2 lie around 0.

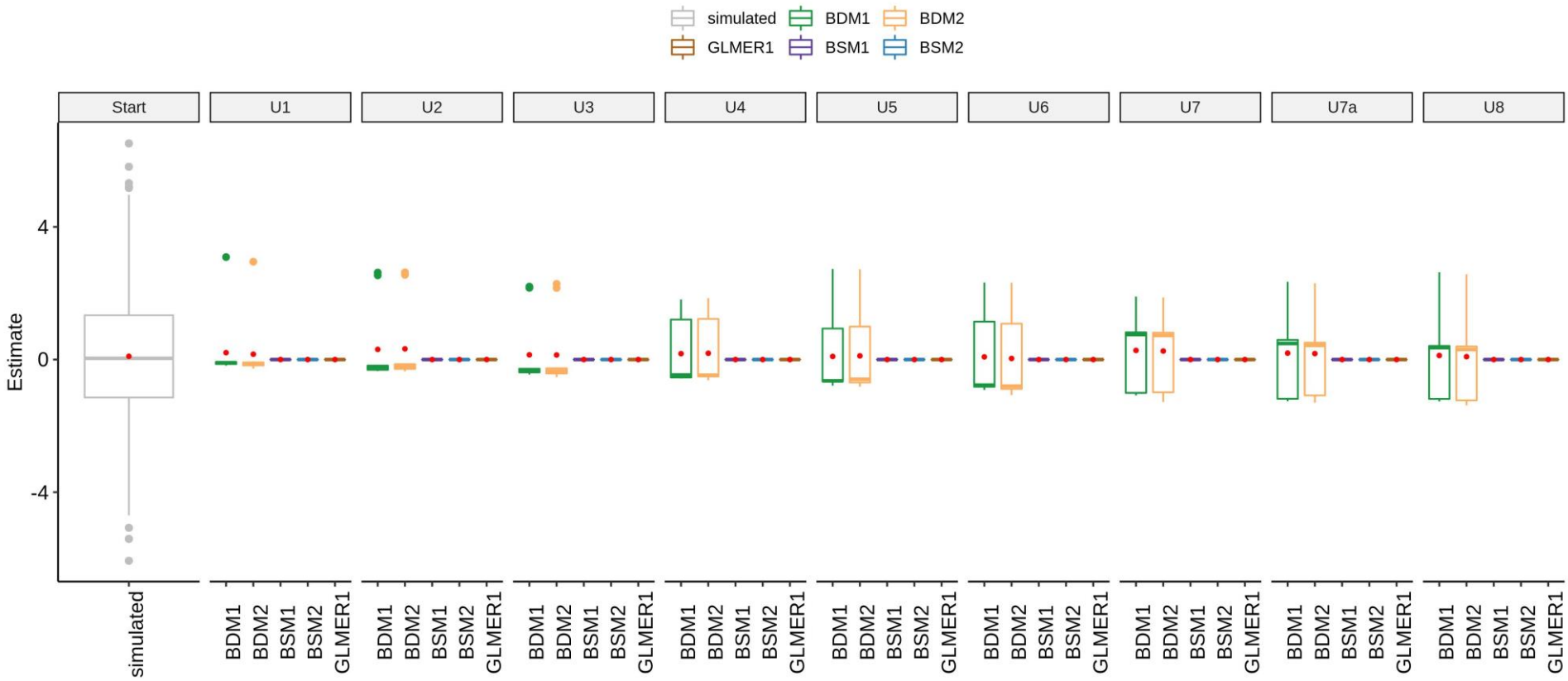


Figure 28. Distributions of random intercepts over time – compared between Bayesian vs. GLMER1 models in simulation main study

The following plots show the development of the estimated random intercepts and fixed effects as well as predicted outcomes for the LOOCV with a specific ID, ID number 1, dependent of prediction visits. Results of other selected IDs can be found online in the folder `/supplementary_material/3.2.7.`

It can be seen in Figure 29 that random intercept of this ID estimated with GLMER1, BSM1, and BSM2, stays equal 0, while BDM1 and BDM2 updated this estimate at each new prediction visit. The expected value of this estimate starts near 0 and continually decreases from U1 to U4 and then continually increases to the simulated value = 3.668. The standard error for ID 2's random intercept was obtained from the GLMER1 model and serves as the starting point for BDM1 and BDM2. The standard error seems to decrease with increasing amount of information from previous visits that was available for prediction.

Figure 30 shows the updating process of the fixed effects over time. Expected values of ID 1's fixed effects stay the same for all models but BDM2. These value for BSM1, BSM2, and BDM1 are exact the same (over time) as expected, due to their defined model specification. Small differences between the expected values estimated by the Bayesian models and GLMER1 are expected. The expected values estimated by BDM2 seem to fluctuate around those estimated by BDM1, BSM1, BSM2. However, no timely trend can be seen here. Expected values of all fixed effects with the exception of β_6, that were estimated by four Bayesian models and GLMER1 seem to be deviate from the simulated value.

Outcome predicted at different prediction visits are shown in Figure 31. For BSM2 and GLMER1 the courses of the predicted outcome stay the same over time, since no updating was done. BSM1 shows unremarkable changes of the predicted outcome over time. BDM1 and BDM2 update the predicted outcome according to available data at the prediction visit. The development of ID 1's random intercept and outcome of past visits seems to associate with the development of this ID's outcome over time.

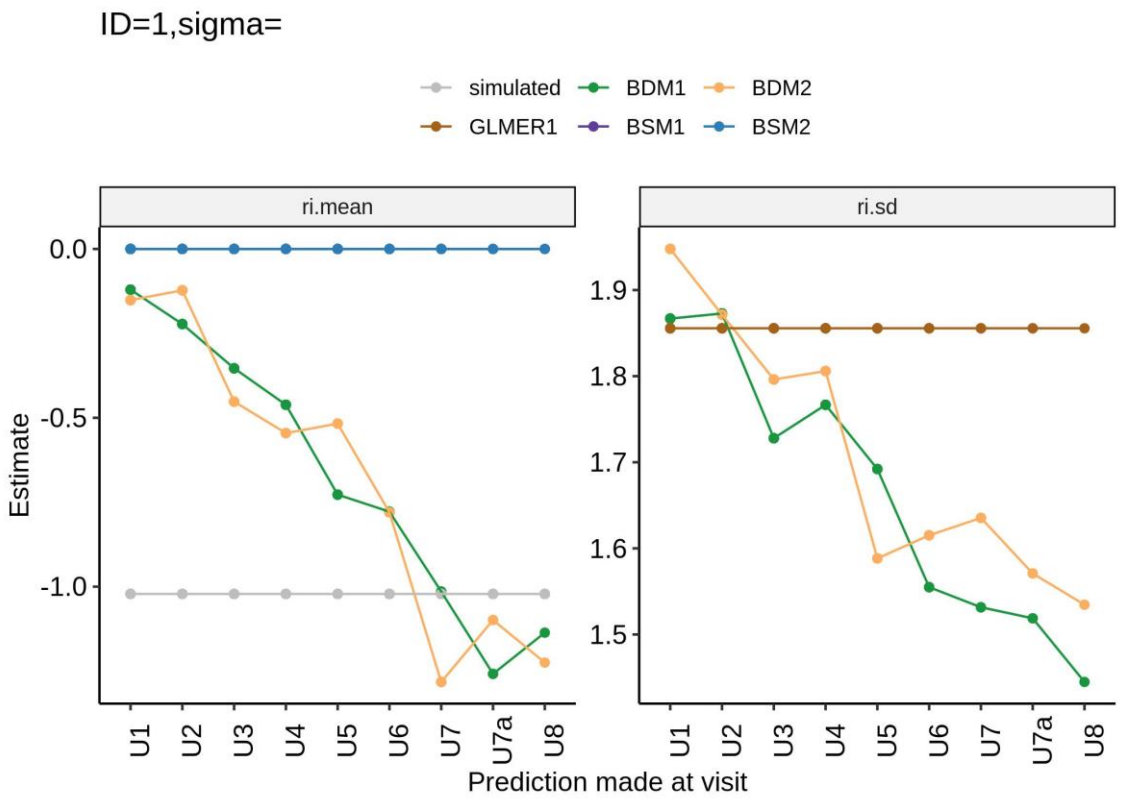


Figure 29. Random intercept estimated by Bayesian models vs. GLMER1 over time in LOOCV with validation ID = 1

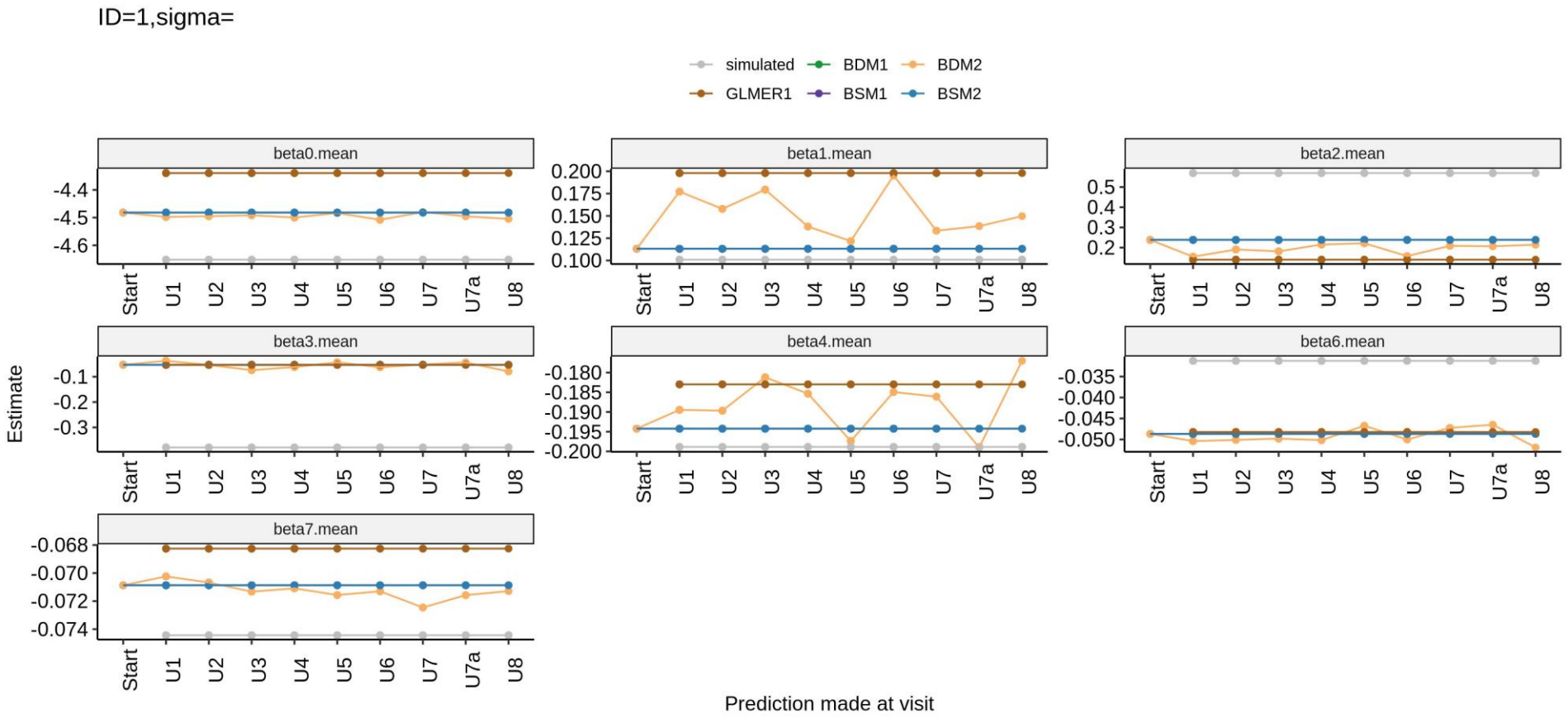


Figure 30. Fixed effects estimated by Bayesian models vs. GLMER1 over time in LOOCV with validation ID = 1

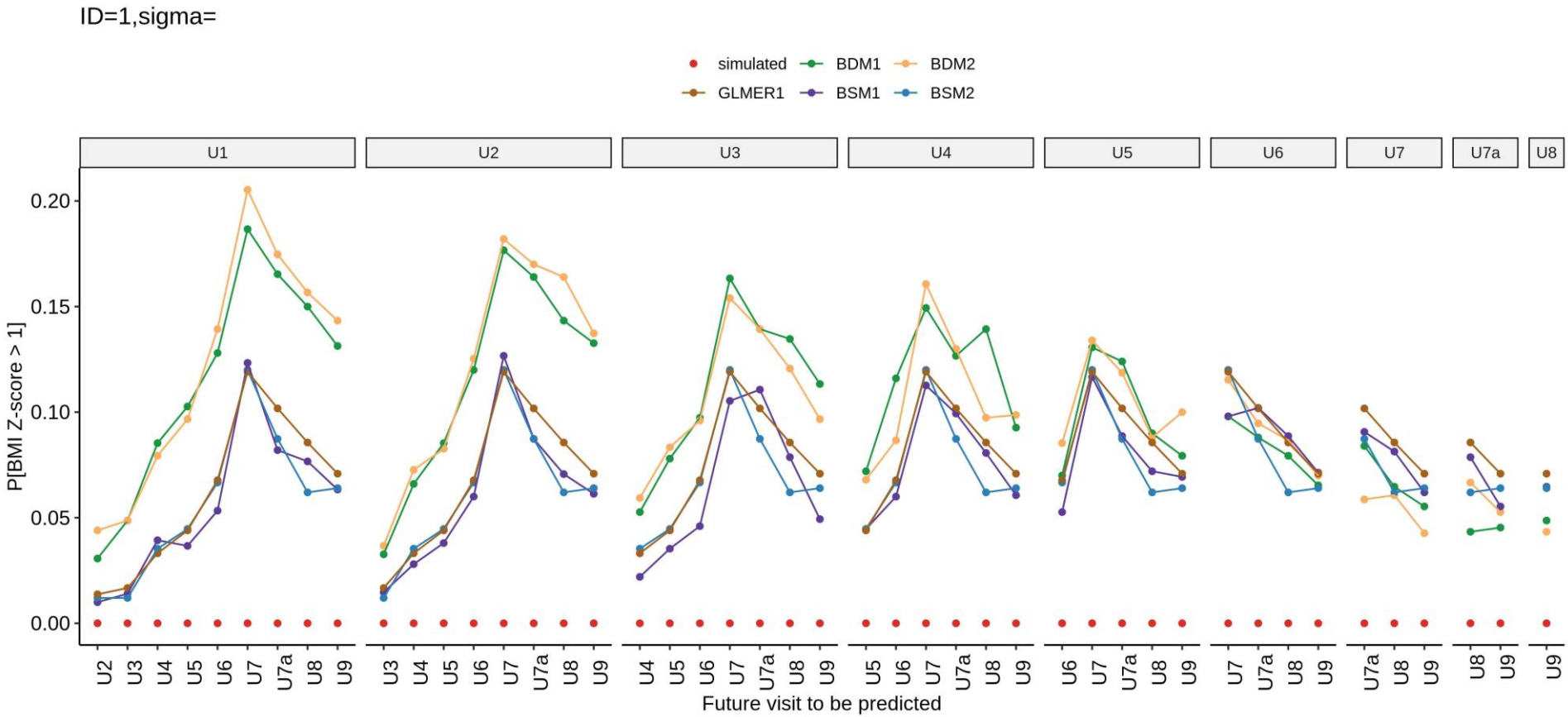


Figure 31. Outcome predicted by Bayesian models vs. GLMER1 over time in LOOCV with validation ID = 1

3.3 Simulation ICC study

3.3.1 Brier Score among Bayesian models

Brier Score among the Bayesian models of different ICC scenarios are shown in Figure 32, Figure 33, and Figure 34. Figure 32 shows the similarity between BSM1 and BSM2 regardless of the amount of available information (whether prediction was made at earlier or later visits) and the ICC that was defined in the simulation design. The same can be observed for BDM1 and BDM2. No remarkable differences between the two curves in each panel can be detected, whether prediction was made at earlier or later visits, whether the ICC is relatively lower or higher.

The comparison between BDM1 and BSM1 (Figure 34) reveals that the prediction performance of BDM1 is comparable with that of BSM1 for lower ICC. With higher ICC, BDM1 overperforms BSM1 and this overperformance seems to increase when more data from the past visits were available for prediction.

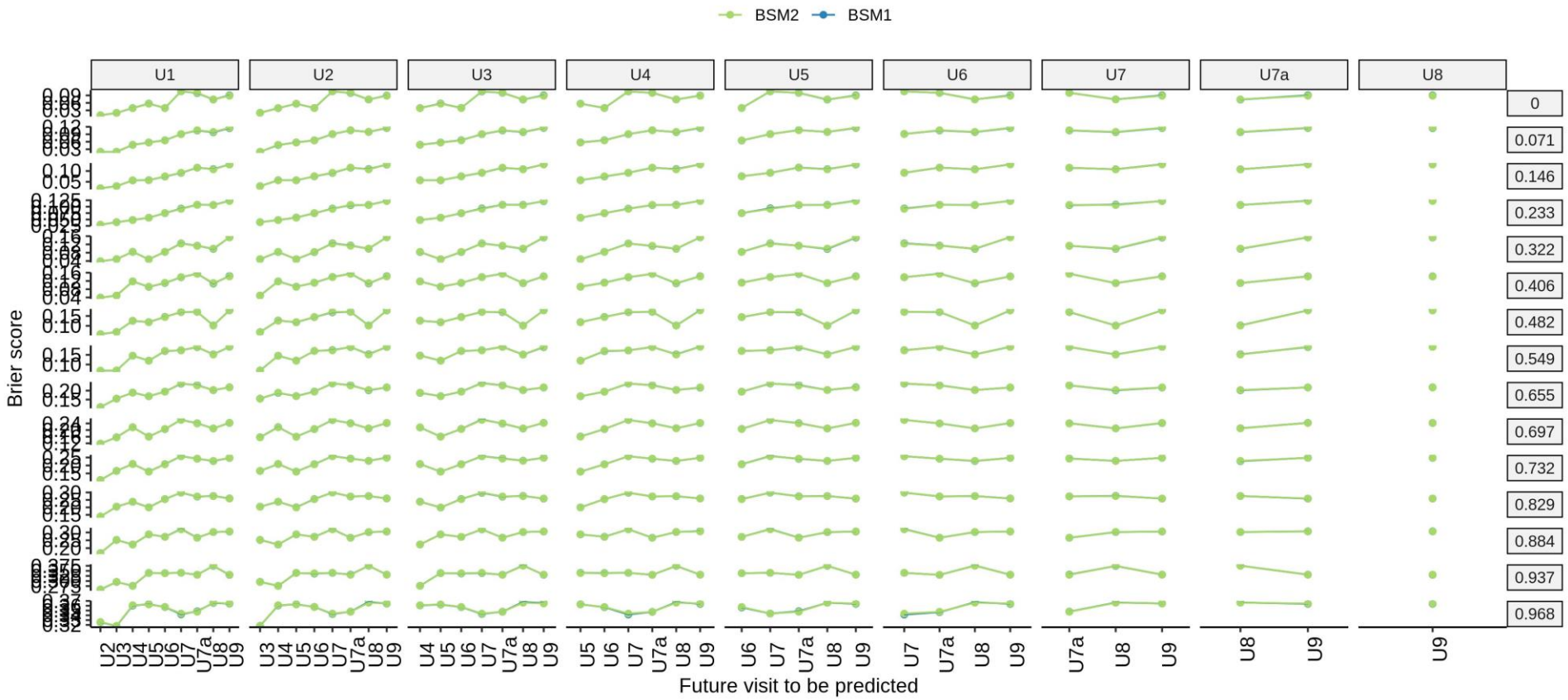


Figure 32. Impact of ICC on difference of performance between BSM1 and BSM2

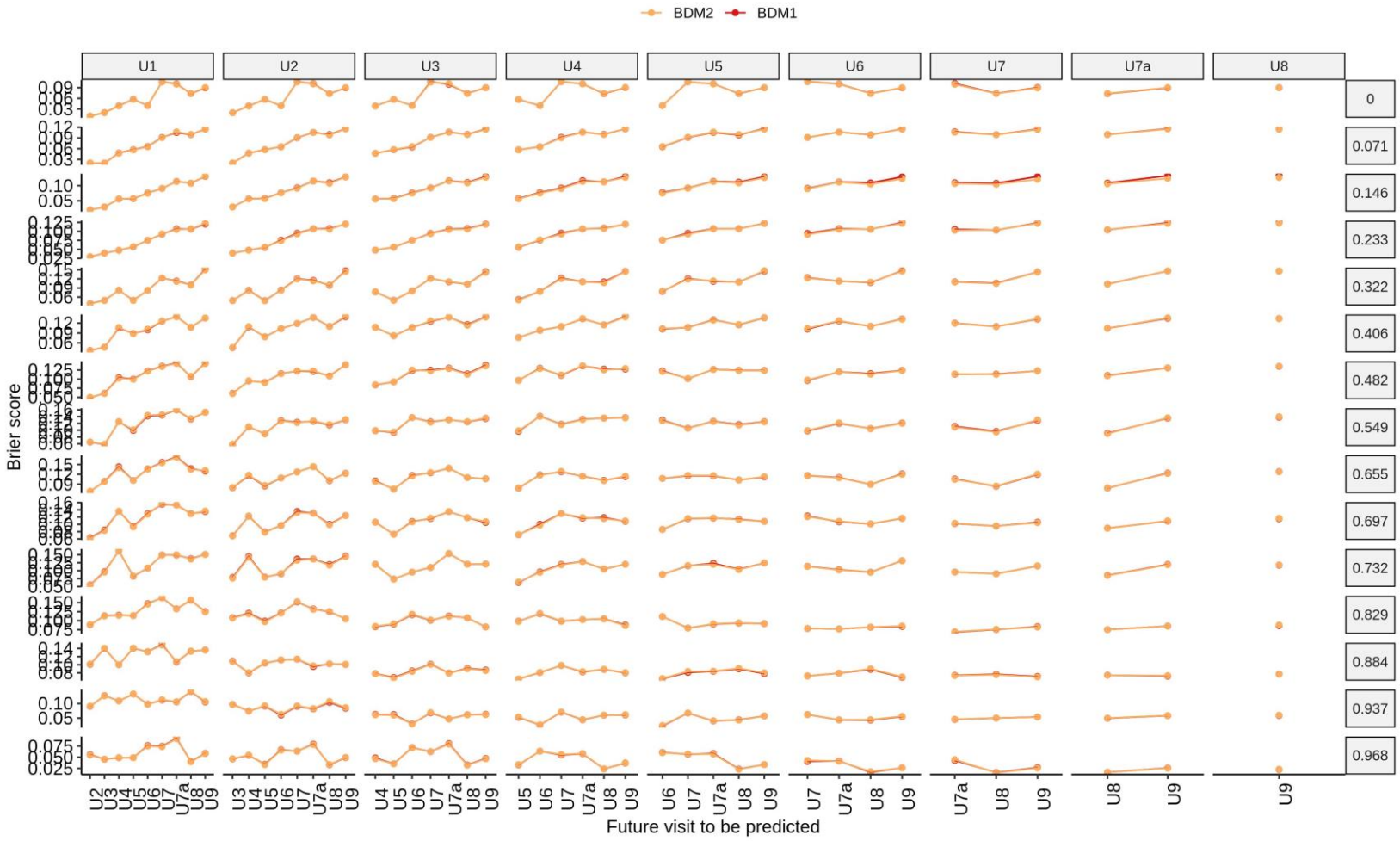


Figure 33. Impact of ICC on difference of performance between BDM1 and BDM2

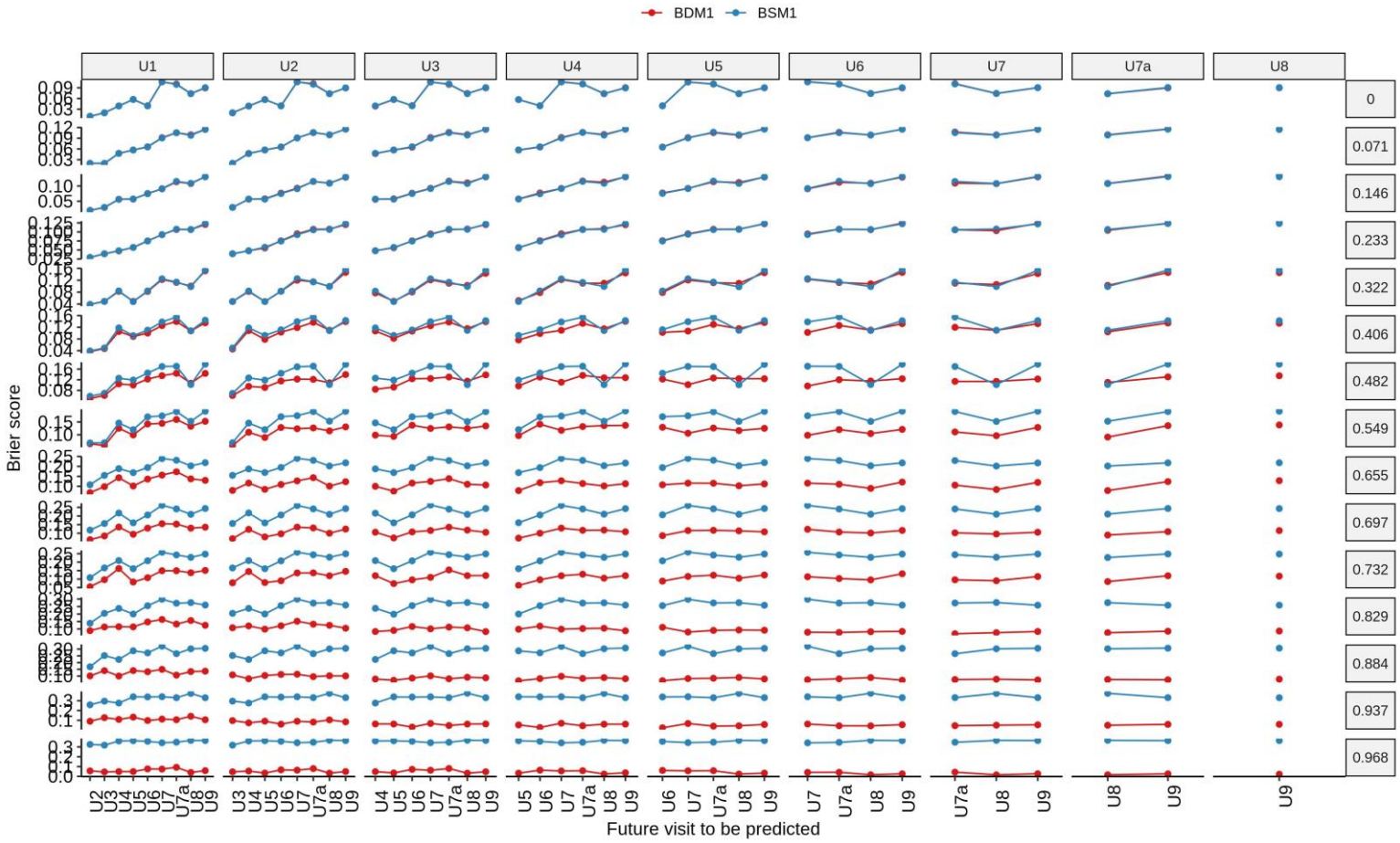


Figure 34. Impact of ICC on difference of performance between BSM1 and BDM1

3.3.2 Brier Score of Bayesian models vs. GLMER1

Brier Score of BSM1 and BDM1, compared with GLMER1 in different ICC scenarios are shown in Figure 35 and Figure 36, respectively. The comparison between BSM2 and BDM2 with GLMER1 is omitted here, same results are expected because the Brier Score of BSM1 is similar to that of BSM2 and the Brier Score of BDM1 is similar to that of BDM2 (see 3.3.1).

For ICCs lower than 0.9, BSM1 and GLMER1 show very similar prediction performance (Figure 35). GLMER1 starts to overperform BSM1 with higher ICCs. The amount of information from past visits used for prediction does not seem to influence these observations.

Figure 36 shows that BDM1 and GLMER1 perform similarly for lower ICCs. BDM1 starts overperforming GLMER1 with ICC larger than 0.4. This overperformance stays consistent with ICC > 0.5 and seems to be more pronounced with more data available from the past visits.

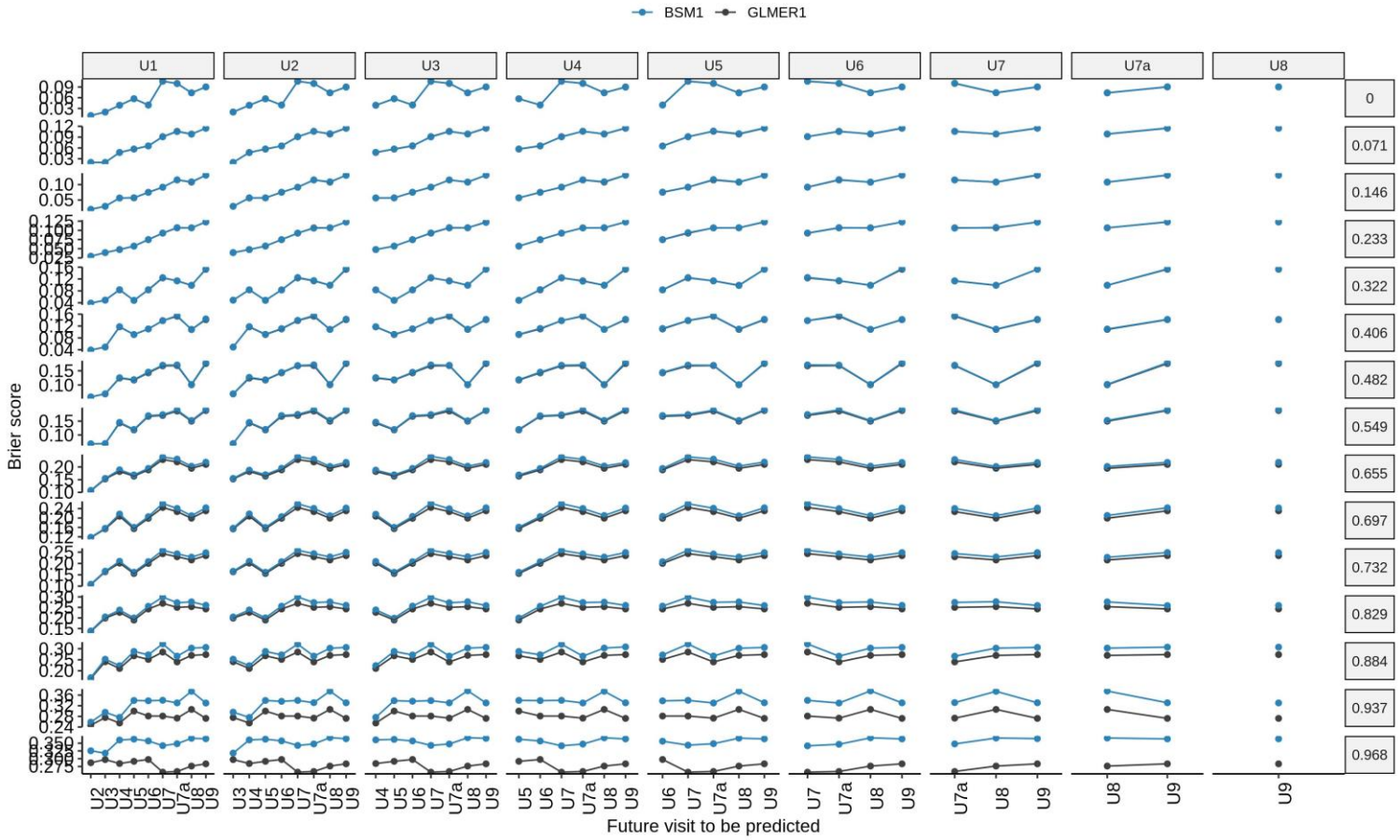


Figure 35. Impact of ICC on difference of Brier Score between BSM1 and GLMER1

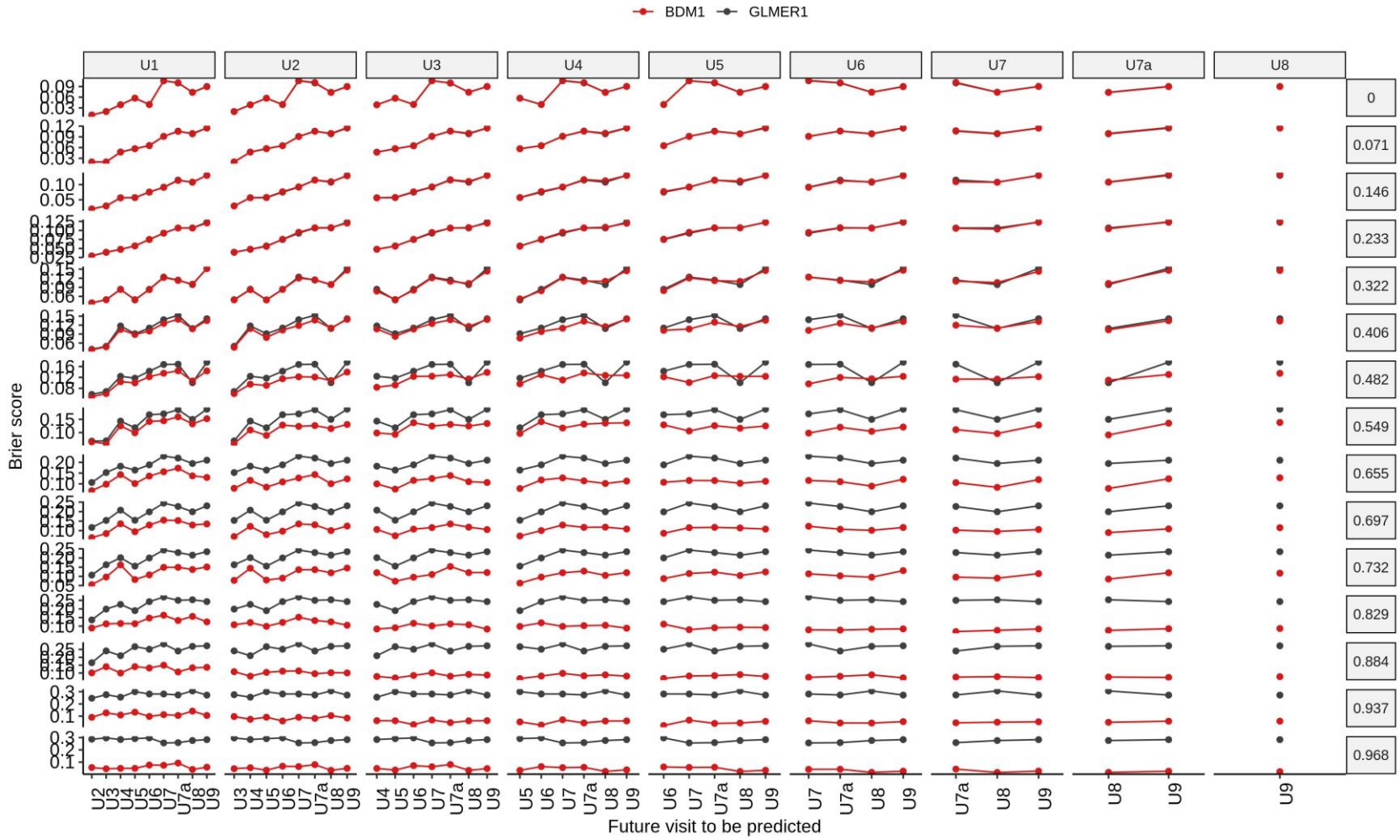


Figure 36. Impact of ICC on difference of Brier Score between BDM1 and GLMER1

3.3.3 Calibration of Bayesian models vs. GLMER1

The following figures show the results of calibration for four Bayesian models and GLMER1 in three different scenarios: scenario $ICC = 0$ in

Figure 37 and Figure 40; scenario $ICC = 0.549$ in Figure 38 and Figure 41; and scenario $ICC = 0.884$ in Figure 39 and Figure 42.

For scenario $ICC = 0$, calibration of the prediction models that predicted future outcome at U2 using data of U1 is depicted in

Figure 37. All models show bad calibration. For other two higher ICC scenarios, BDM1 and BDM2's overperformance can be observed for the prediction of future outcome at U2 using data of U1 (Figure 38 and Figure 39). Using accumulated data from U1 to U8 to predict data at U2, BDM1 and BMD2 show consistent overperformance. This consistency seems to be more pronounced with ICC = 0.884 than with ICC = 0.549.

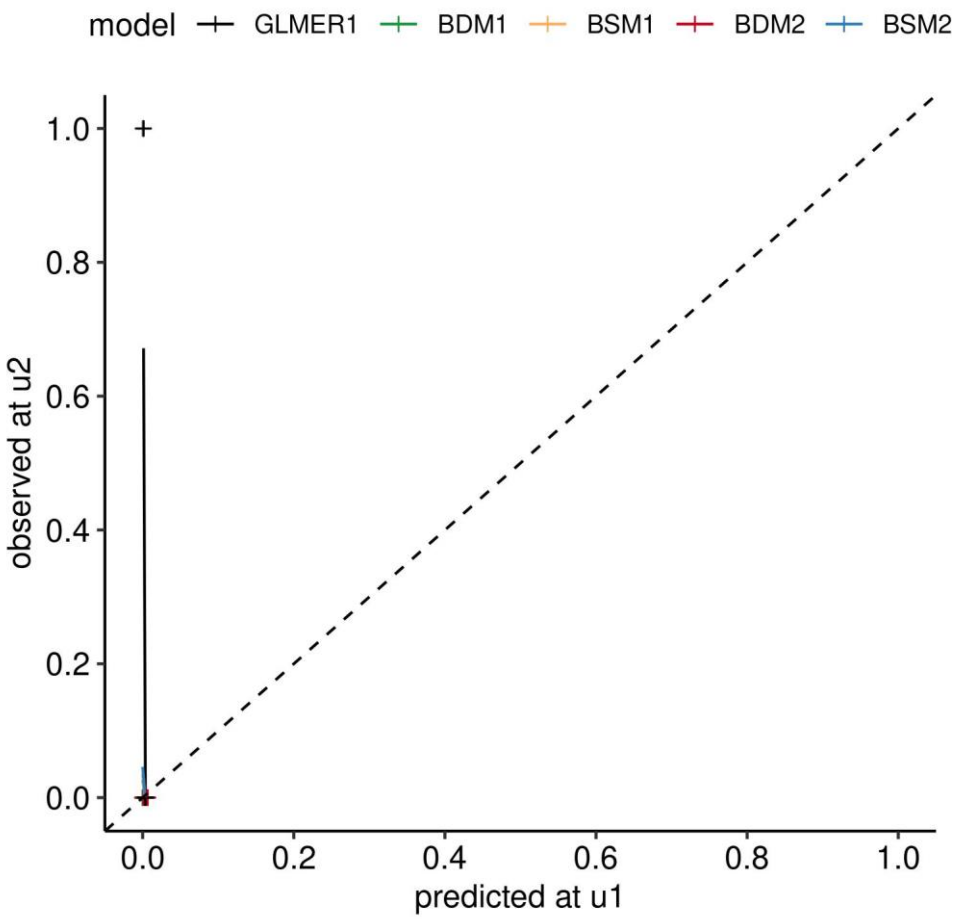


Figure 37. Calibration of Bayesian models vs. GLMER1 predicting U2 outcome using U1 data - $ICC = 0$

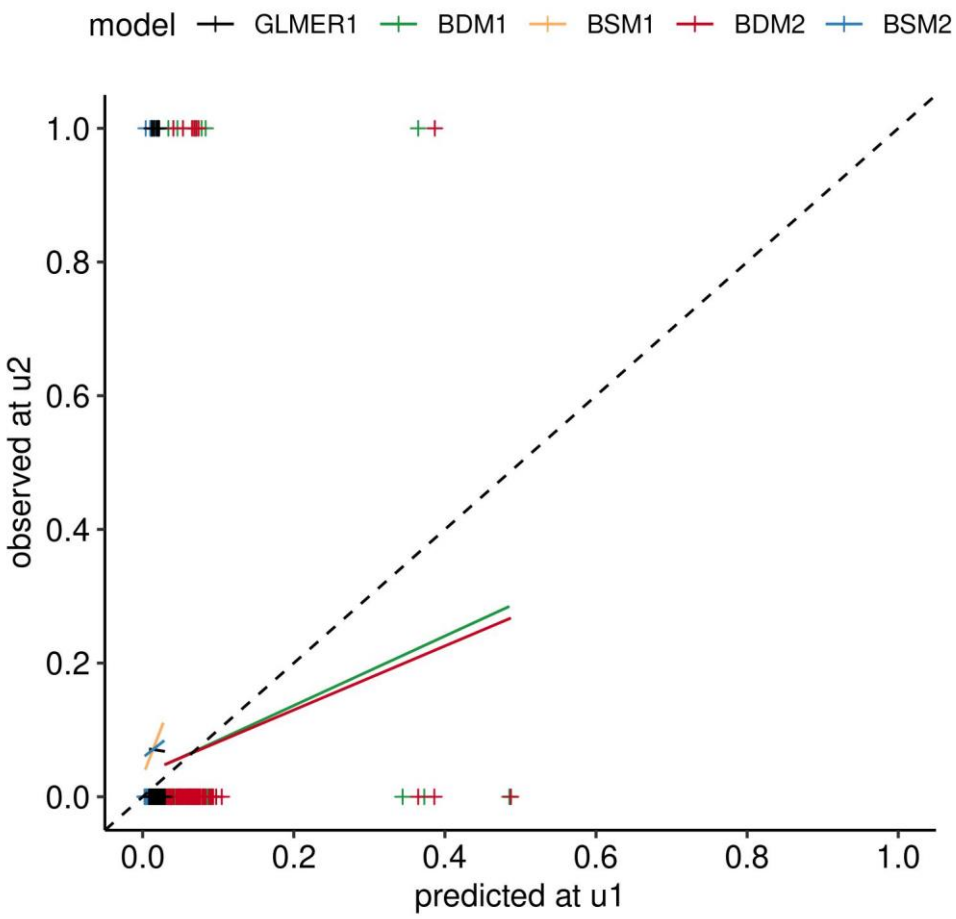


Figure 38. Calibration of Bayesian models vs. GLMER1 predicting U2 outcome using U1 data - $ICC = 0.549$

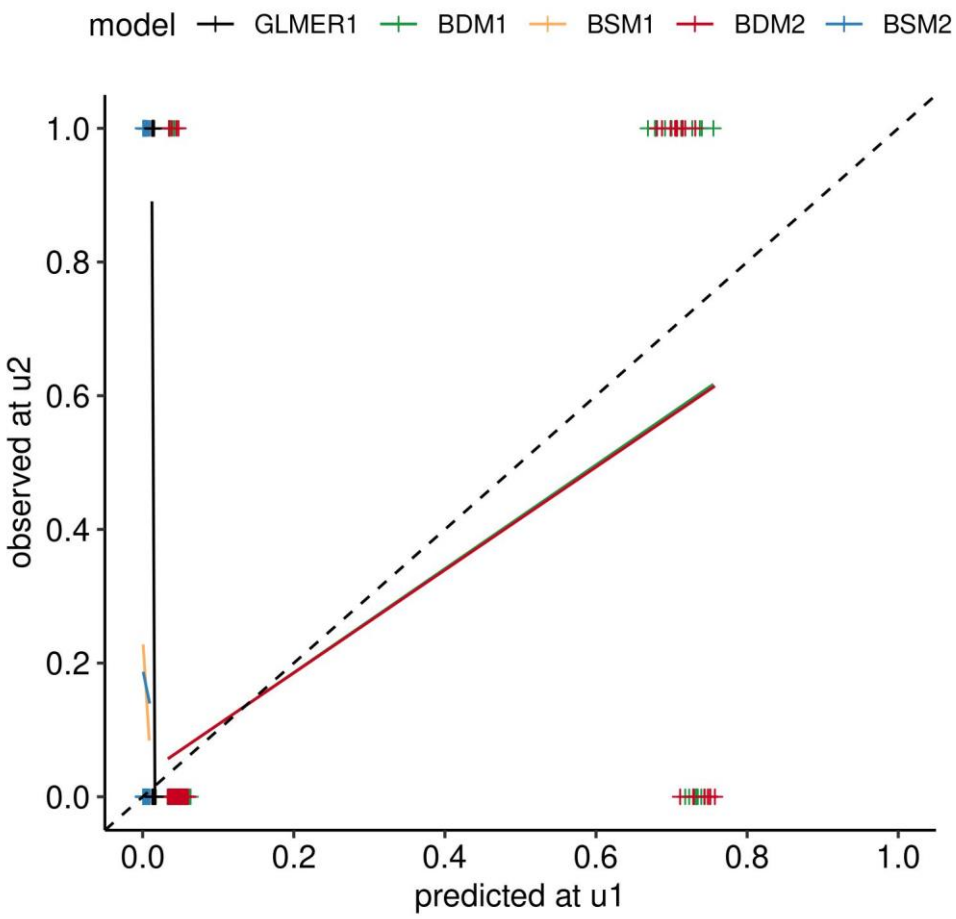


Figure 39. Calibration of Bayesian models vs. GLMER1 predicting U2 outcome using U1 data - $ICC = 0.884$

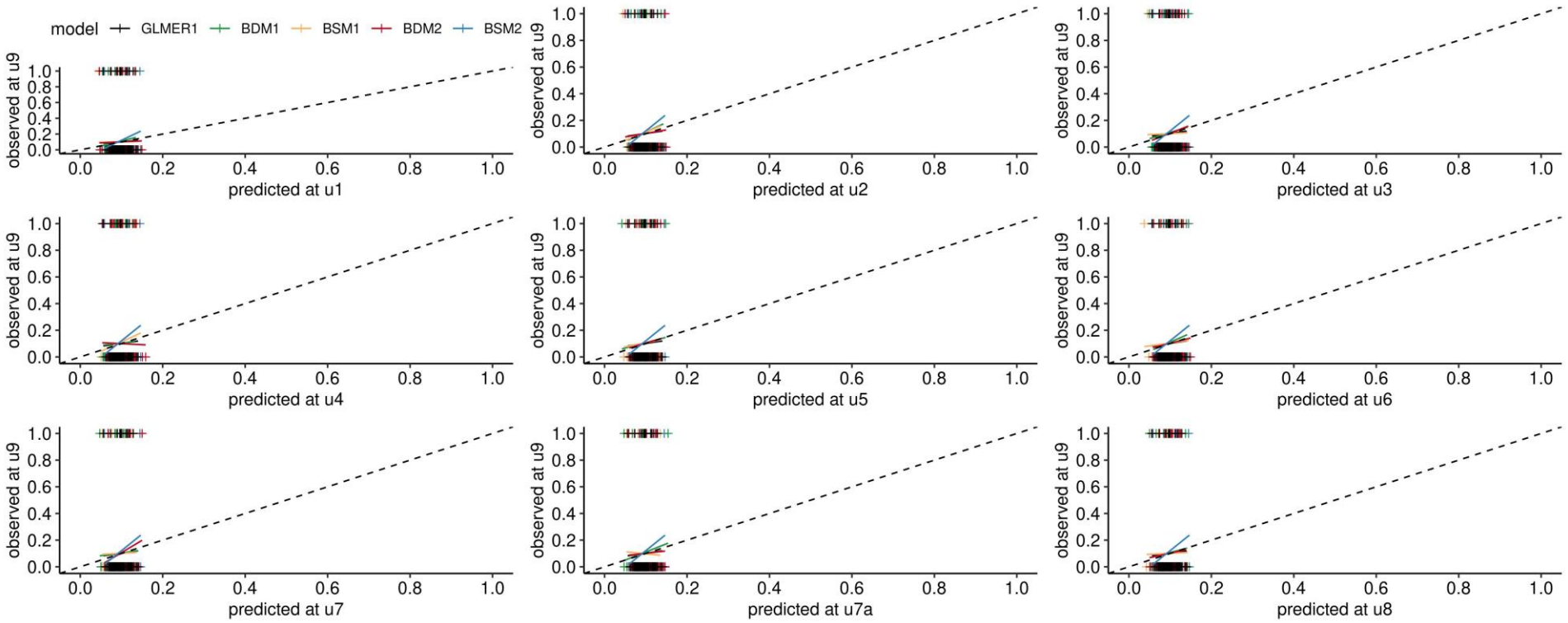


Figure 40. Calibration of Bayesian models vs. GLMER1 predicting U9 outcome using cumulated data from U1 until U8- $ICC = 0$

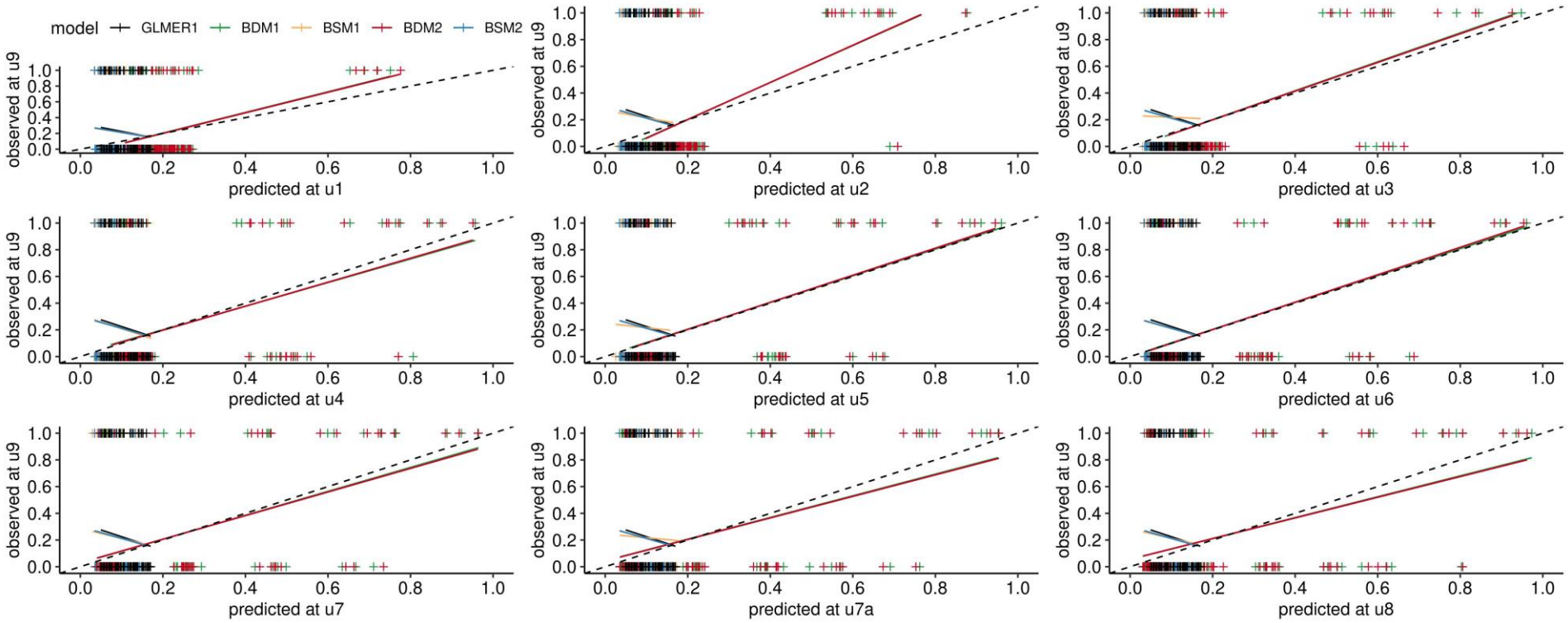


Figure 41. Calibration of Bayesian models vs. GLMER1 predicting U9 outcome using cumulated data from U1 until U8- $ICC = 0.549$

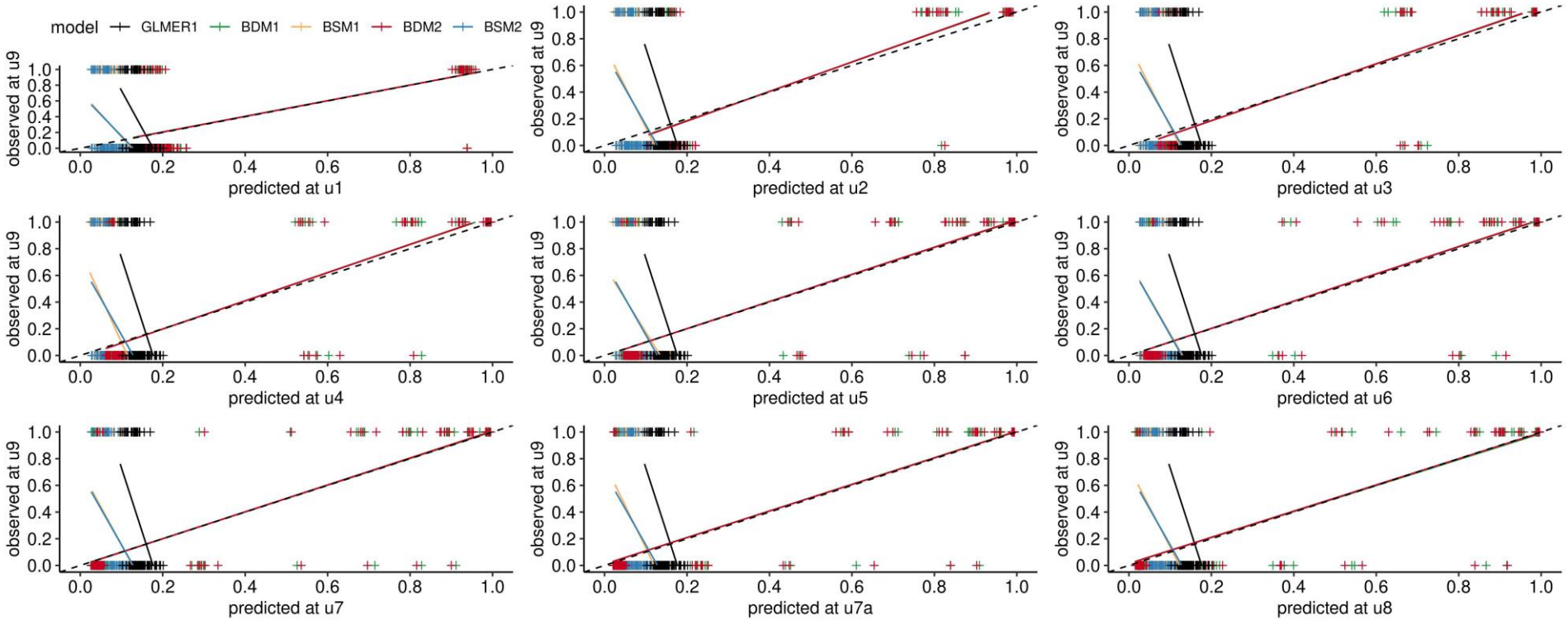


Figure 42. Calibration of Bayesian models vs. GLMER1 predicting U9 outcome using cumulated data from U1 until U8-$ICC = 0.884$

3.3.4 Scaled Brier Score of Bayesian models vs. GLMER1

For ICC =0, four Bayesian models performs similarly with GLMER1, regardless of number of past visits considered to predict future outcome.

For higher ICC, which are 0.549 and 0.884, as respectively shown in Figure 44 and Figure 45, Scaled Brier Score of BDM1 and BDM2 reveal consistent overperformance of these models compared to BSM1, BSM2, and GLMER1 over time. The Scaled Brier Score of BDM1 and BDM2 also show their overperformance compared to a non-informative model, and this overperformance seems to increase with increasing information of past visits used for prediction.

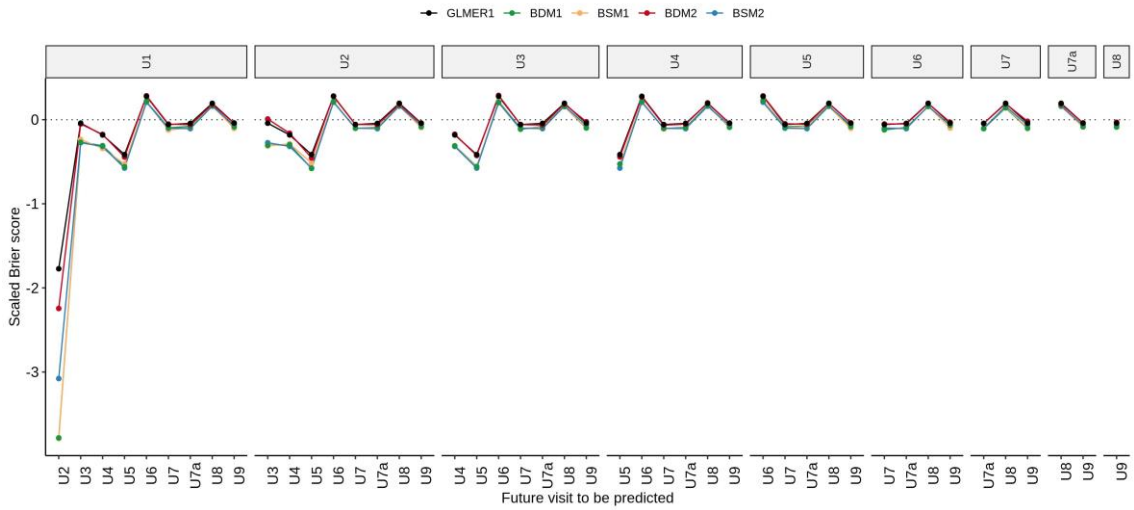


Figure 43. Scaled Brier Score of Bayesian models vs. GLMER1 over time in simulation study - $ICC = 0$

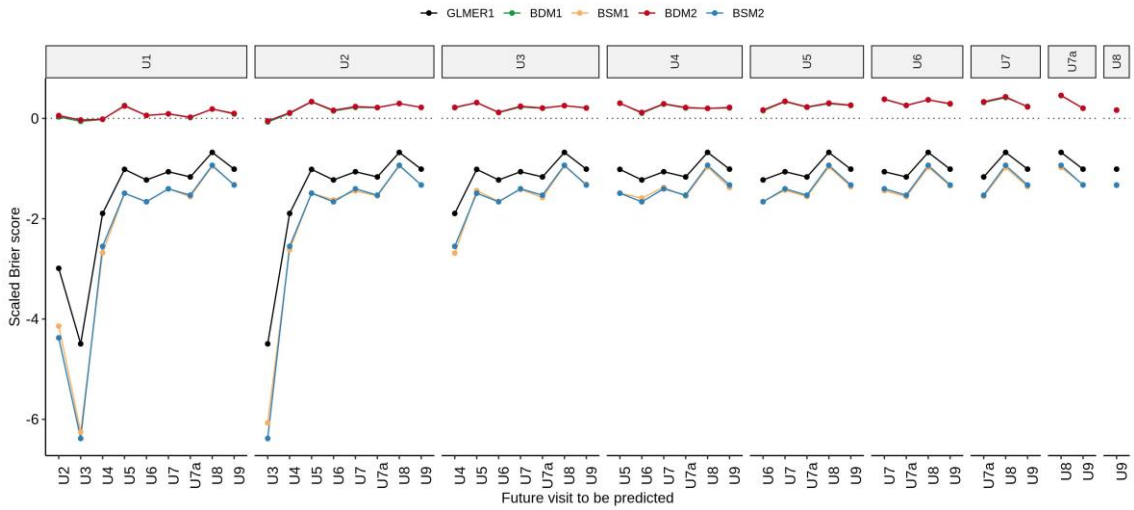


Figure 44. Scaled Brier Score of Bayesian models vs. GLMER1 over time in simulation study - $ICC = 0.549$

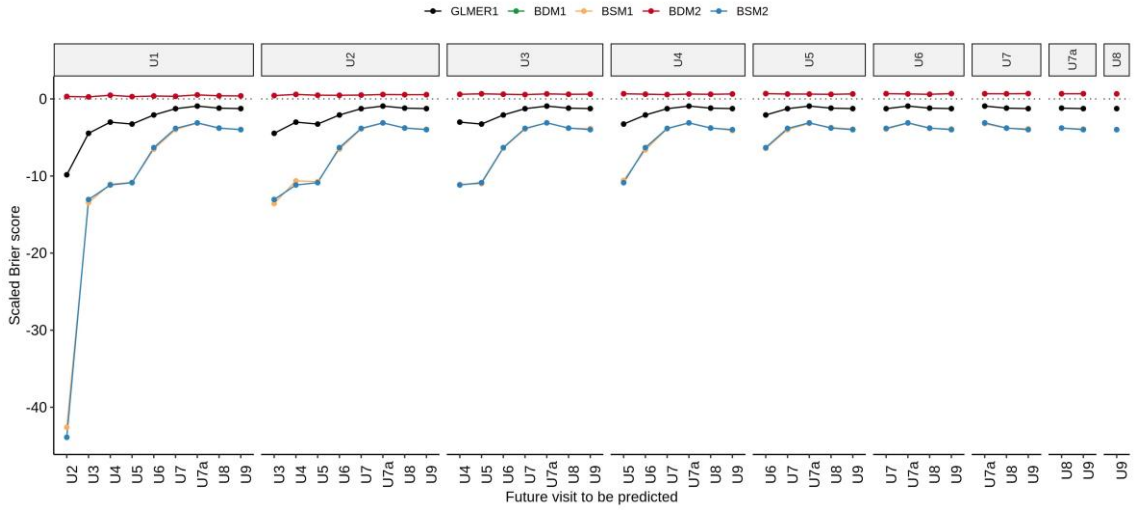


Figure 45. Scaled Brier Score of Bayesian models vs. GLMER1 over time in simulation study - $ICC = 0.884$

3.3.5 Brier Skill Score

3.3.5.1 Brier Skill Score of Bayesian models vs. GLMER1

For ICC = 0, the similarity among models can also be shown with respect to BSS Figure 46.

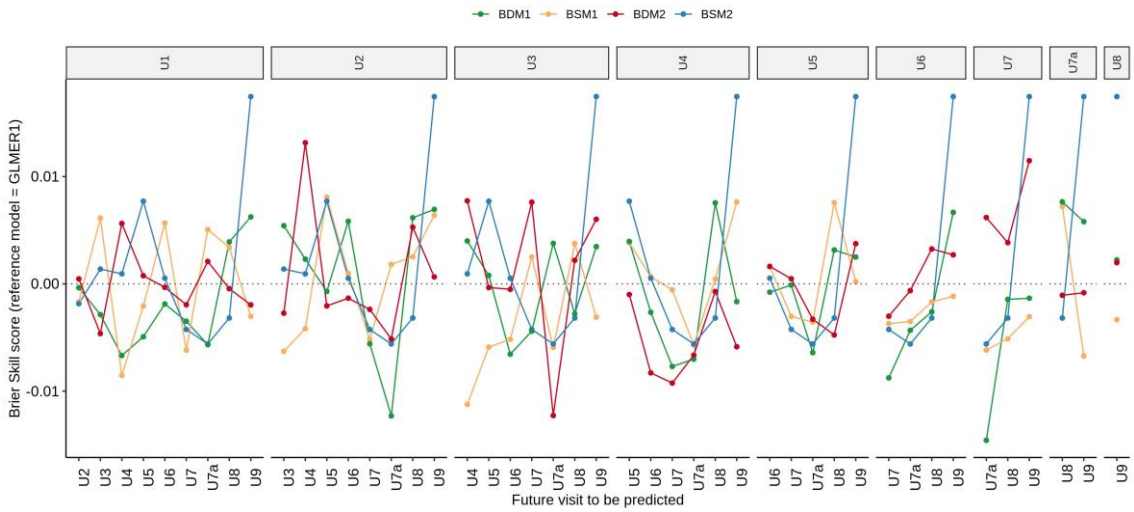


Figure 46. Brier Skill Score of Bayesian models over time in simulation study - $ICC = 0$

For higher ICC, BDM1 and BDM2 clearly overperforms BSM1, BSM2, and GLMER1. BSM1 and BSM2 show poorer performance compared with GLMER1. Both of these observations are more pronounced in scenarios with higher ICC (Figure 47 and Figure 48).

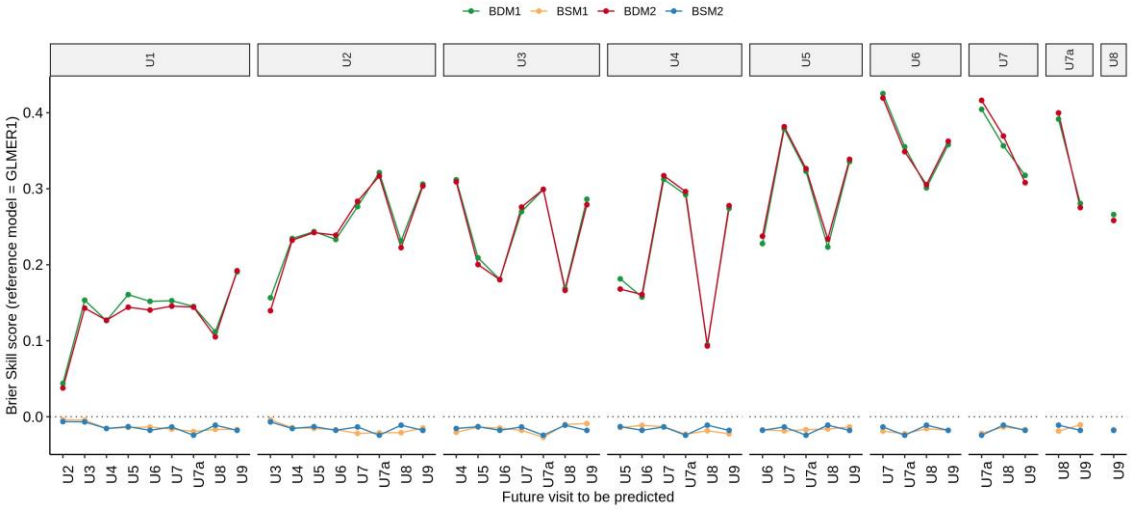


Figure 47. Brier Skill Score of Bayesian models over time in simulation study – $ICC = 0.549$

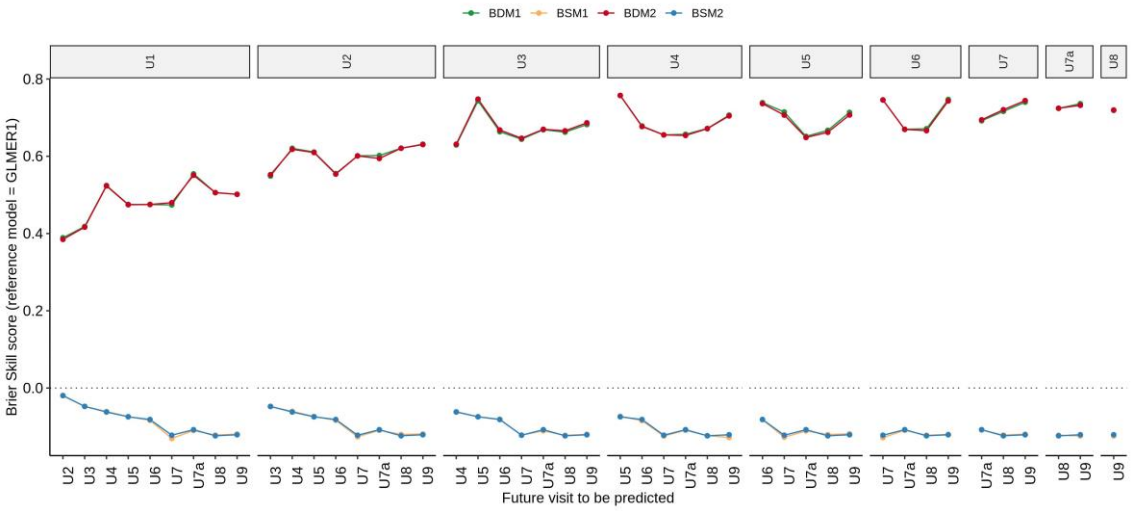


Figure 48. Brier Skill Score of Bayesian models over time in simulation study - $ICC = 0.884$

3.3.5.2 Brier Skill Score with “U1-model” as reference

For ICC = 0, the similarity among models can be shown with respect to Brier Skill Score (Figure 49).

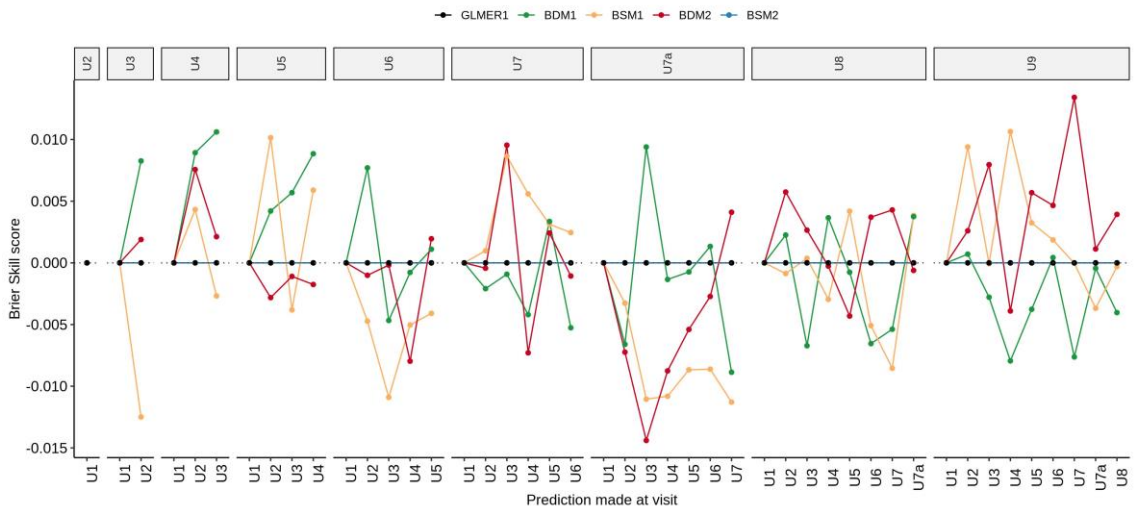


Figure 49. Brier Skill Score of Bayesian models (reference = "U1-model") over time in simulation study - $ICC = 0$

For ICC = 0.549, BDM1 and BDM2 models trained at later visits clearly overperforms those trained at U1. Later models of BSM1 and BSM2 show comparative performance compared with their "U1-models". This observation is more obvious in scenarios where ICC = 0.884 (Figure 50 and Figure 51).

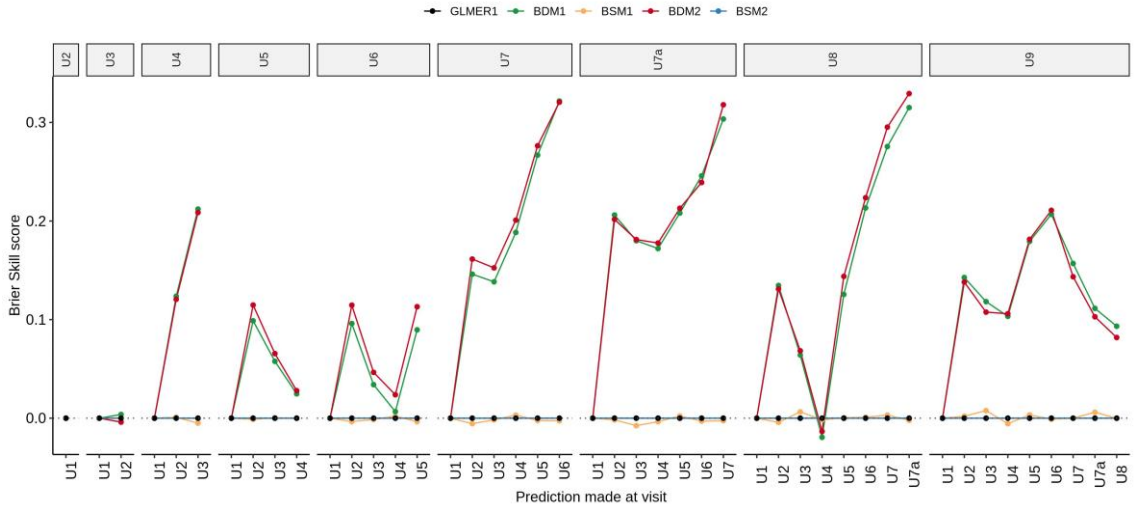


Figure 50. Brier Skill Score of Bayesian models (reference = "U1-model")" over time in simulation study – $ICC = 0.549$

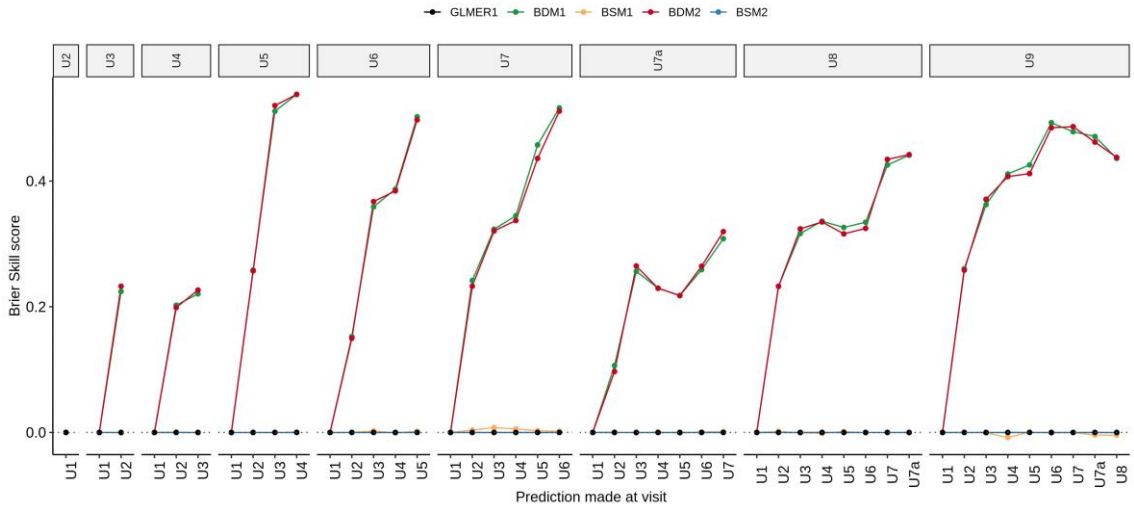


Figure 51. Brier Skill Score of Bayesian models (reference = "U1-model") over time in simulation study - $ICC = 0.884$

3.3.6 Results at individual level - Simulation ICC study

The following figures show the distribution of random intercept estimated by four Bayesian models and GLMER1 in difference ICC scenarios. The distribution of simulated values is depicted in grey. For BSM1, BSM2, and GLMER1, random intercept = 0 was assumed. In all ICC scenarios, these models show the same figures for the random intercepts. In scenario where ICC = 0, BDM1 and BDM2 estimated a τ different from 0. However, the variance of random intercept seems to be close to 0.

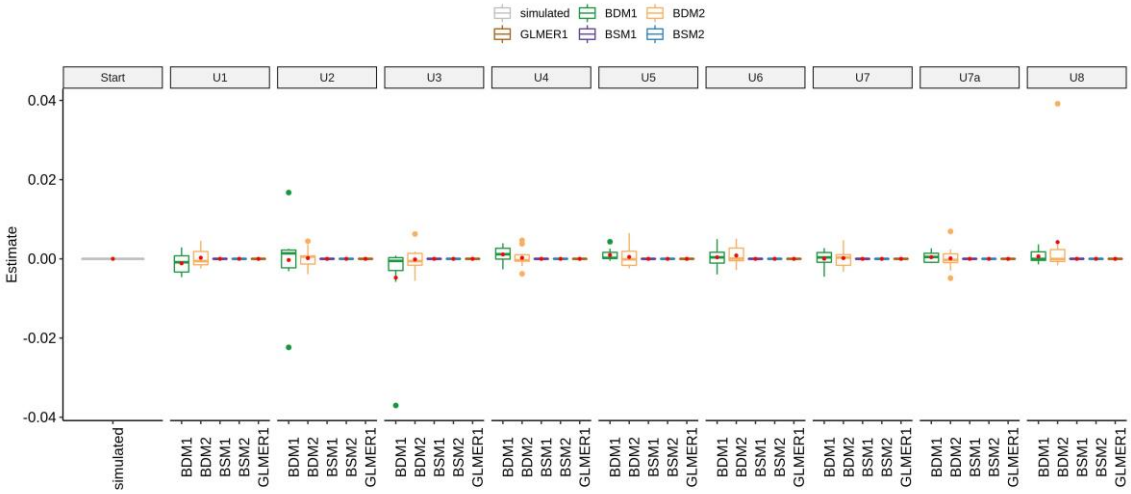


Figure 52. Distributions of random intercepts over time – compared between Bayesian vs. GLMER1 models in simulation study with $ICC = 0$

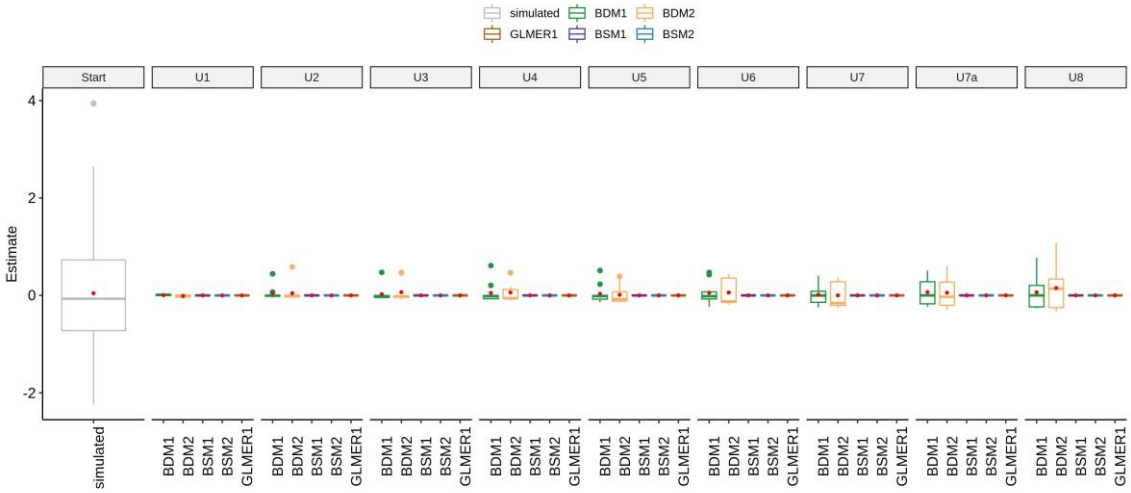


Figure 53. Distributions of random intercepts over time – compared between Bayesian vs. GLMER1 models in simulation study with ICC$= 0.233$

For lower ICC, it took several visits for BDM1 and BDM2 to be able to capture the distribution of the simulated random intercepts (Figure 53 and Figure 54). For higher ICC, this seems to happen right at U1 (Figure 55 and Figure 56)

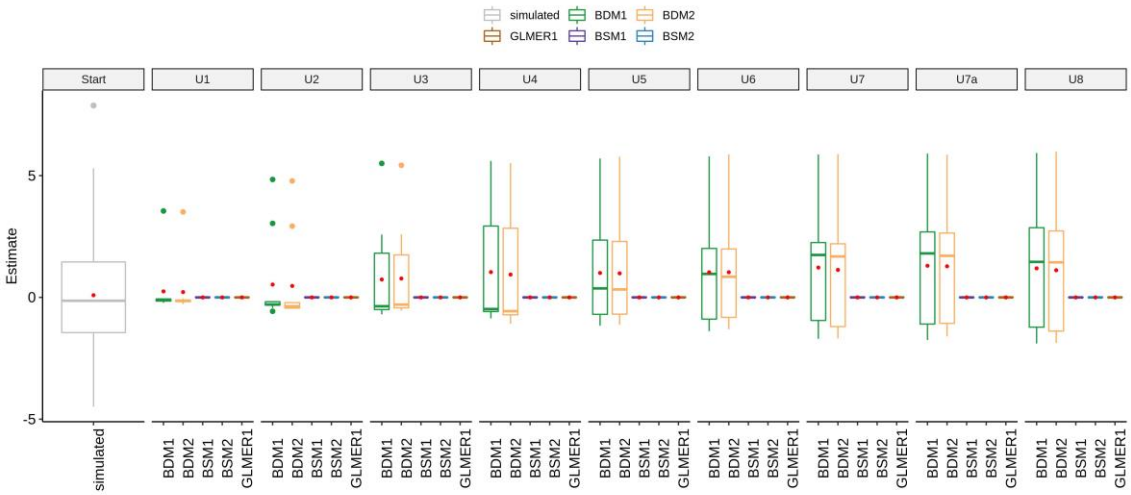


Figure 54. Distributions of random intercepts over time – compared between Bayesian vs. GLMER1 models in simulation study with $ICC = 0.549$

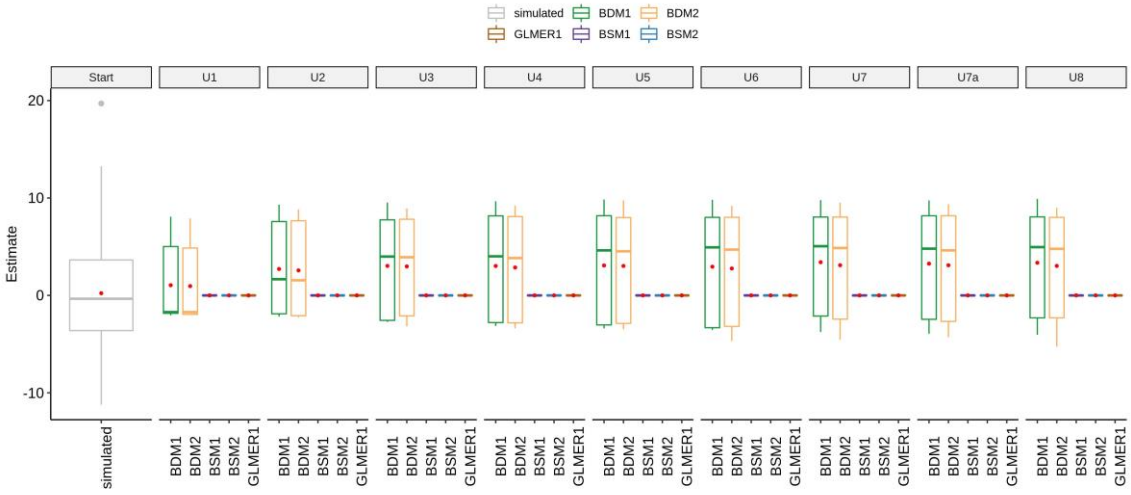


Figure 55. Distributions of random intercepts over time – compared between Bayesian vs. GLMER1 models in simulation study with $ICC = 0.884$

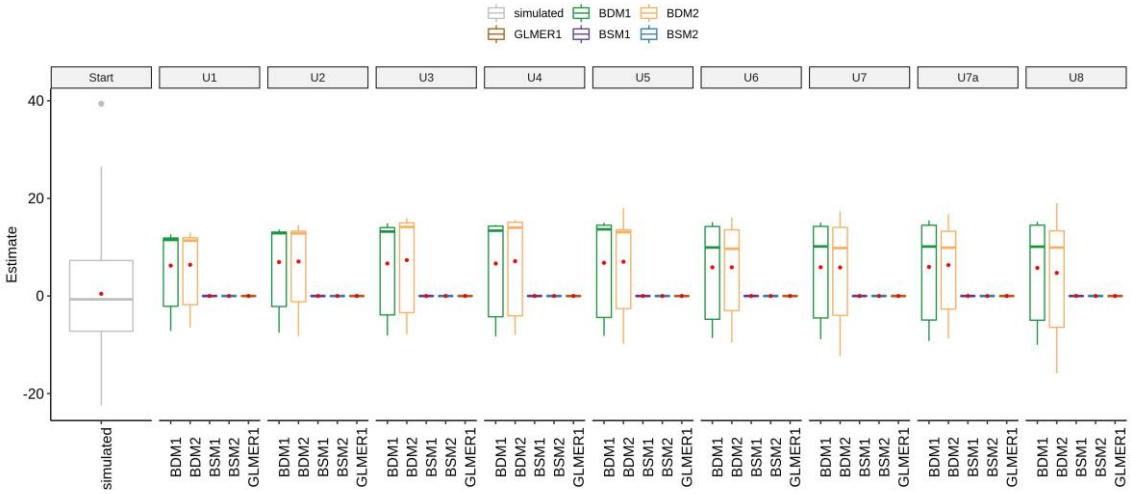


Figure 56. Distributions of random intercepts over time –compared between Bayesian vs. GLMER1 models in simulation study with $ICC = 0.968$

The following figures show results of one individual for three different ICC scenarios. Results of other selected IDs can be found online in the folder `/supplementary_material/3.3.6.`

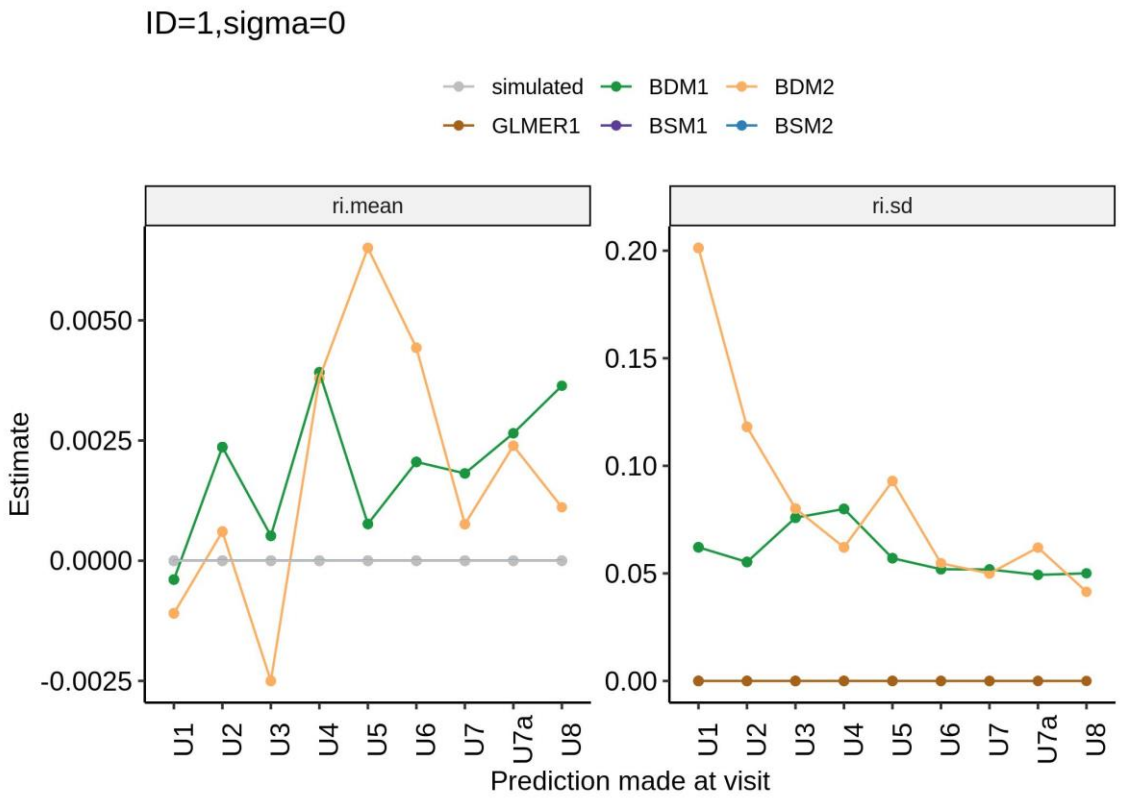


Figure 57. Random intercept estimated by Bayesian vs. GLMER1 models over time in LOOCV with validation ID = 1 in simulation study with $ICC = 0$

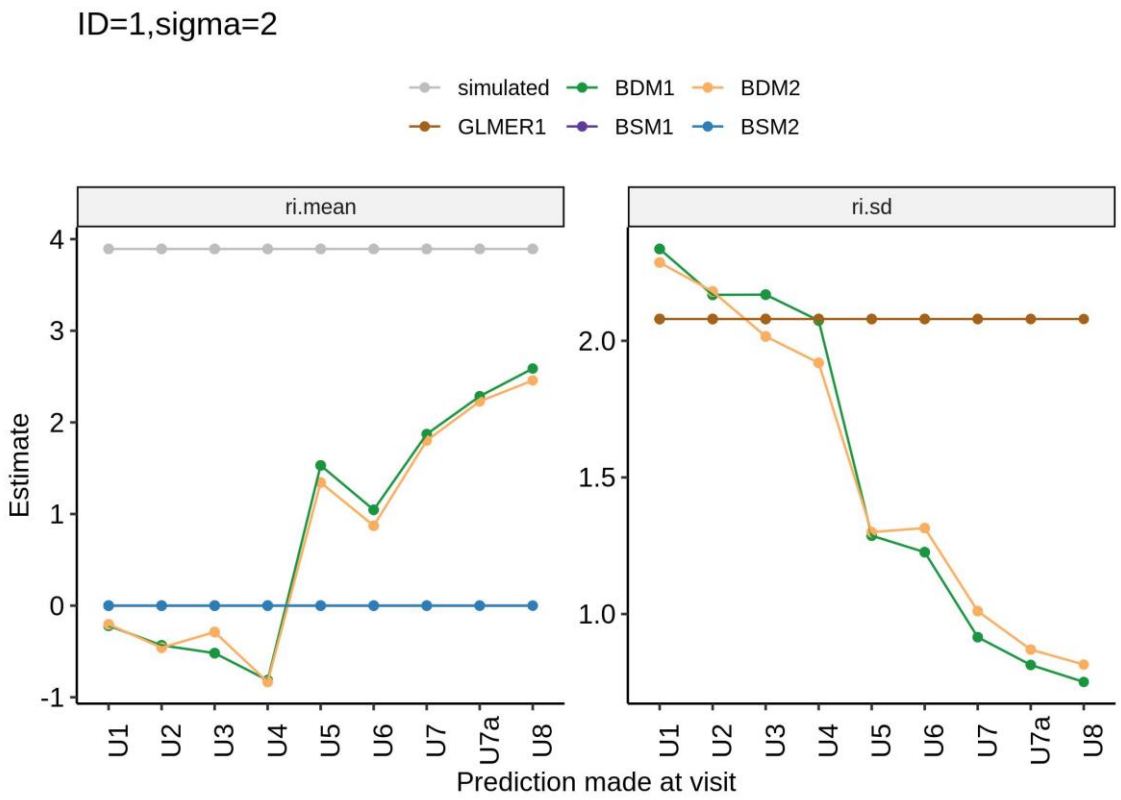


Figure 58. Random intercept estimated by Bayesian vs. GLMER1 models over time in LOOCV with validation ID = 1 in simulation study with $ICC = 0.549$

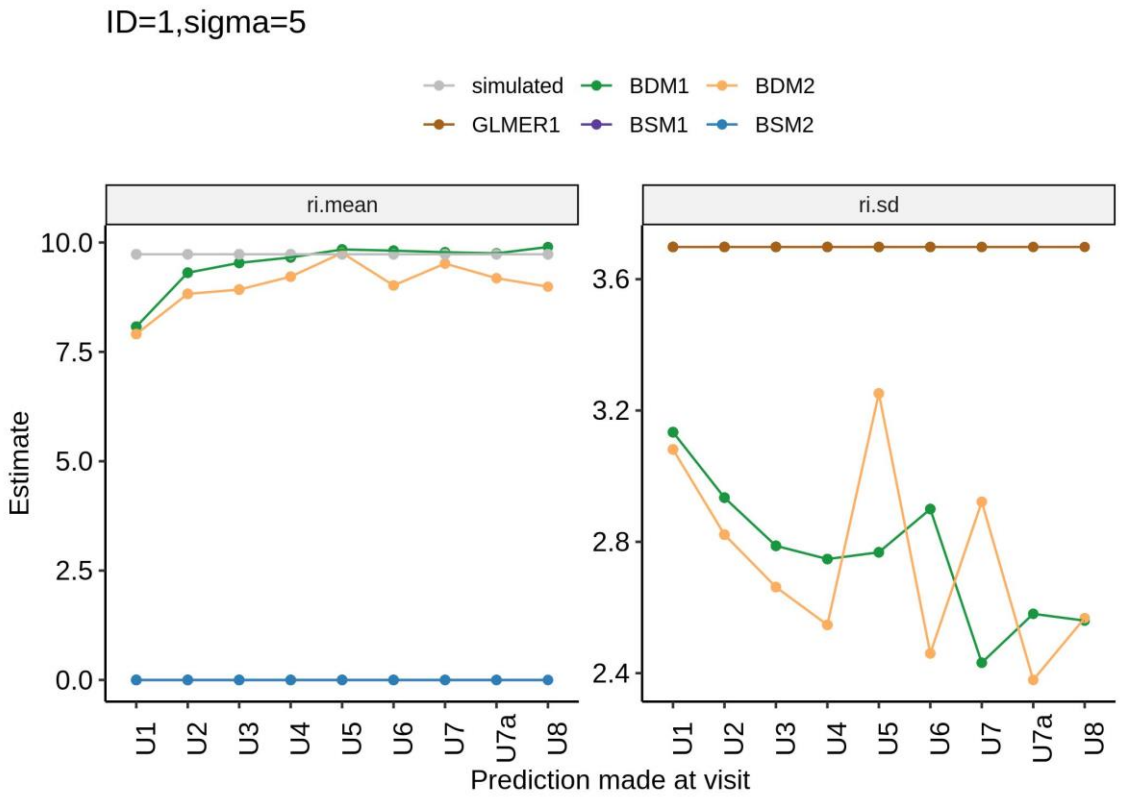


Figure 59. Random intercept estimated by Bayesian vs. GLMER1 models over time in LOOCV with validation ID = 1 in simulation study with $ICC = 0.884$

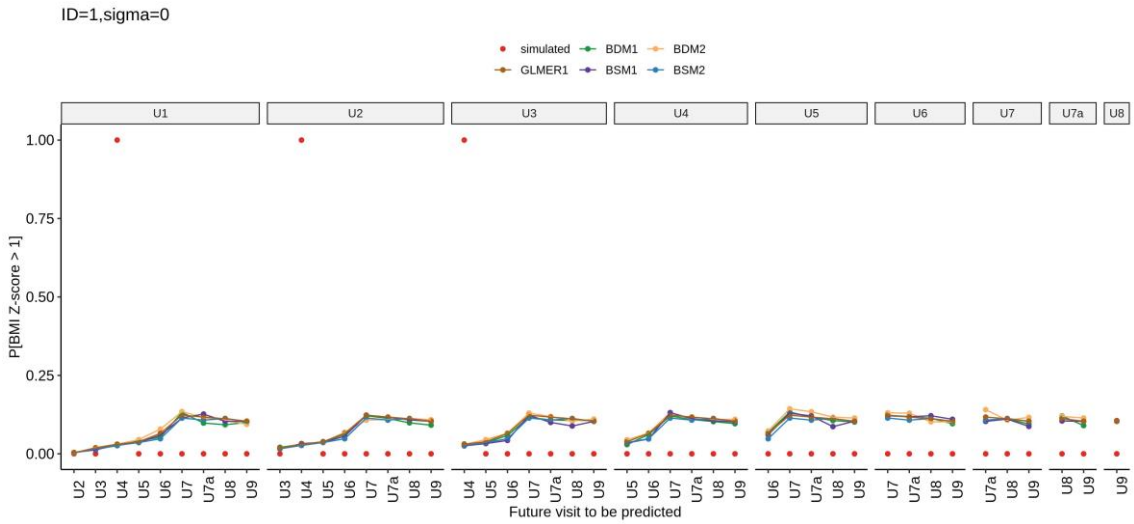


Figure 60. Outcome predicted by Bayesian vs. GLMER1 models over time in LOOCV with validation ID = 1 in simulation study with $ICC = 0$

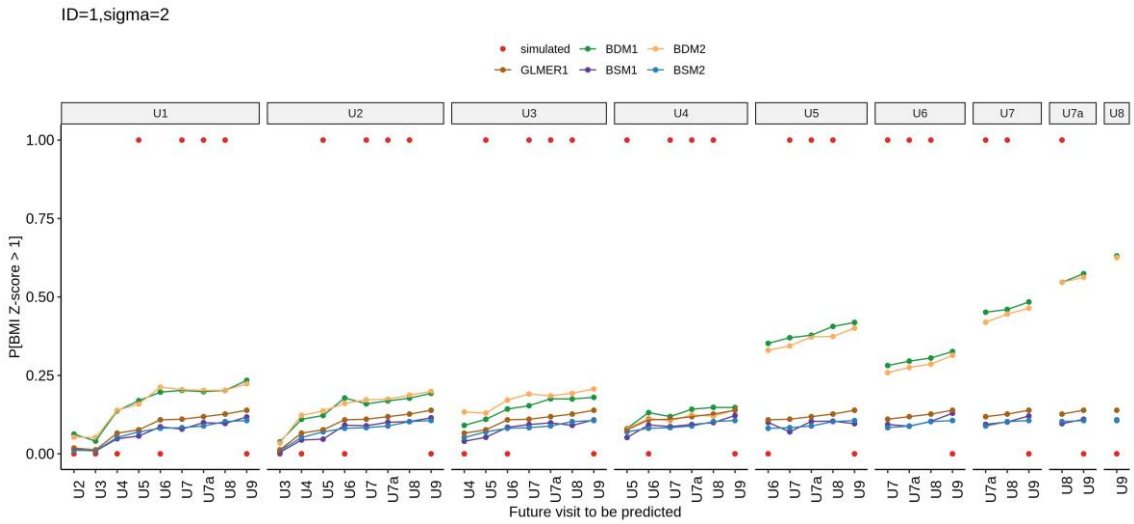


Figure 61. Outcome predicted by Bayesian vs. GLMER1 models over time in LOOCV with validation ID = 1 in simulation study with $ICC = 0.549$

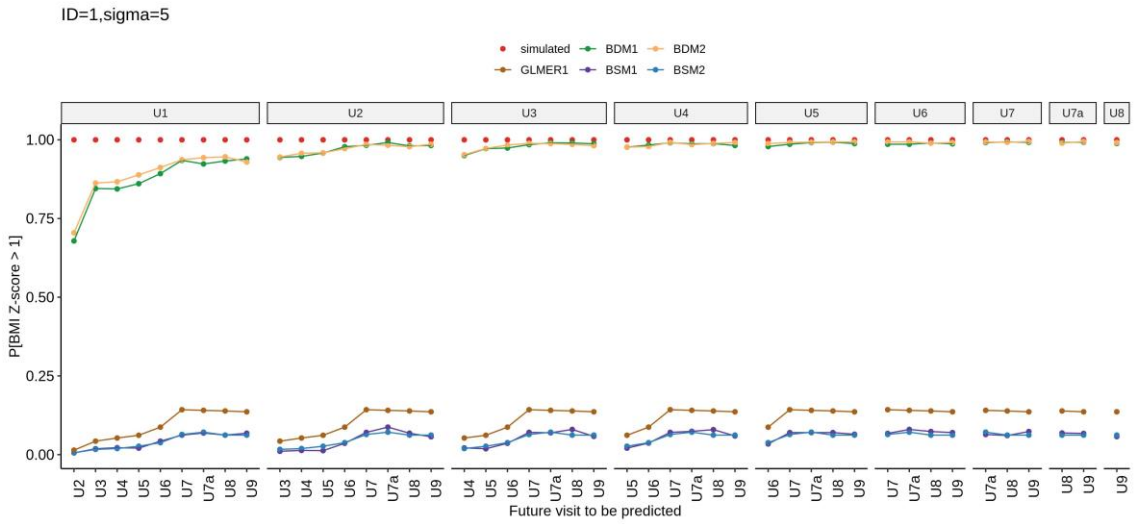


Figure 62. Outcome predicted by Bayesian vs. GLMER1 models over time in LOOCV with validation ID = 1 in simulation study with $ICC = 0.884$

3.3.7 Influence of ICC and time on prediction performance

Table 36 shows the output of the selected LMM.

Table 36. Influence of ICC and time on Brier Score – Output of LMM

term	estimate	std.error	statistic	df	p.value
(Intercept)	0.009	0.004	2.232	479.821	0.026
icc	0.256	0.007	39.038	3344.784	0.000
udata_age_month	0.001	0.000	12.215	3351.057	0.000
modBDM1	0.045	0.006	7.975	3344.785	0.000
modBDM2	0.045	0.006	7.963	3344.785	0.000
modBSM1	-0.010	0.006	-1.819	3344.785	0.069
modBSM2	-0.010	0.006	-1.821	3344.785	0.069
icc:modBDM1	-0.210	0.009	-22.623	3344.785	0.000
icc:modBDM2	-0.210	0.009	-22.636	3344.785	0.000
icc:modBSM1	0.040	0.009	4.274	3344.785	0.000
icc:modBSM2	0.039	0.009	4.262	3344.785	0.000
udata_age_month:modBDM1	0.000	0.000	0.708	3344.785	0.479
udata_age_month:modBDM2	0.000	0.000	0.594	3344.785	0.552
udata_age_month:modBSM1	0.000	0.000	-0.678	3344.785	0.498
udata_age_month:modBSM2	0.000	0.000	-0.683	3344.785	0.494
icc:udata_age_month	-0.001	0.000	-5.627	3344.785	0.000
icc:udata_age_month:modBDM1	-0.001	0.000	-2.858	3344.785	0.004
icc:udata_age_month:modBDM2	-0.001	0.000	-2.749	3344.785	0.006
icc:udata_age_month:modBSM1	0.000	0.000	1.672	3344.785	0.095
icc:udata_age_month:modBSM2	0.000	0.000	1.687	3344.785	0.092
sd__(Intercept)	0.003				
sd__Observation	0.030				

For ICC = 0, directly after birth, on average, GLMER1 model would have a Brier Score of 0.009. At the same time, BDM1 and BDM2 show higher Brier Score, while the difference of BSM1 and

BSM2 compared with GLMER1 is not significant. The prediction performance of GLMER1 decreases with increasing ICC. For ICC = 0.5, a GLMER1 model trained directly after birth would have a Brier Score of 0.009 + 0.256 * 0.5 = 0.137. For the same ICC, BDM1 and BDM2 would have a Brier Score 0.009 + 0.045 + 0.256 * 0.5 - 0.210 * 0.5 = 0.077, while Brier Score of BSM1 and BSM2 would be 0.009 -0.010 + 0.256*0.5+ 0.04 * 0.5 = 0.147. For ICC = 0, prediction performance of GLMER1 for every month after birth would decrease by 0.001 (Brier Score unit). This effect is the same for all models. For ICC >0, the effect of time on Brier Score attenuates with increasing ICC for BDM1 and BDM2. For BSM1 and BSM2, this effect is not significant.

Output of the selected LMM for $MLRPE^{GLMER1}$ is presented in Table 39. Directly after birth, for a scenario where ICC is equal to 0, $MLRPE^{GLMER1}$ of BSM2 is -0.002, indicating that GLMER1 performs better than BSM2. RPE between GLMER1 and BDM1 and BDM2 are $e^{-0.607}$= 0.545 and $e^{-0.846}$ = 0.429, respectively. $MLRPE^{GLMER1}$ of BSM2 does not differ from that of BSM1 significantly. With increasing ICC, MLRPE of BSM2 increases, indicating that BSM2 overperforms GLMER1 for higher ICC. For ICC = 0.5, directly after birth, on average, RPE of GLMER1 vs. BSM2 would be $e^{-0.002 + 1.781 * 0.5}$ = 2.431. For ICC = 0, this effect is the same for all other models. The interaction between ICC and model is not significant. For ICC = 0, RPE of GLMER1 vs. BSM2 for every month after birth would decrease by 0.5%, which means that the overperformance of GLMER1 vs. BSM2 increases over time. For ICC = 0, this effect is the same for all other models. The interaction between time and model is not significant. However, for ICC>0, the overperformance of GLMER1 vs. BDM1 attenuates over time. Taking an example of ICC = 0.5, directly at birth RPE of GLMER1 vs. BDM1 would be $e^{-0.002 + 1.781 * 0.5 - 0.607}$ = 1.325. At month 10 after birth, this RPE would increase to $e^{-0.002 + 1.781 * 0.5 - 0.607 + 0.053 * 0.5 * 10}$ = 1.727. This effect is not significant for BSM1.

Table 37. Influence of ICC and time on MLRPE – Output of LMM

term	estimate	std.error	statistic	df	p.value
(Intercept)	-0.002	0.135	-0.012	84.306	0.991
icc	1.781	0.184	9.668	2651.659	0.000
udata_age_month	-0.005	0.003	-1.587	2658.767	0.113
modBDM1	-0.607	0.159	-3.825	2651.659	0.000
modBDM2	-0.846	0.158	-5.359	2651.658	0.000
modBSM1	-0.009	0.159	-0.060	2651.656	0.952
icc:modBDM1	-0.399	0.260	-1.532	2651.667	0.126
icc:modBDM2	-0.076	0.259	-0.292	2651.660	0.770
icc:modBSM1	0.025	0.260	0.095	2651.656	0.924
udata_age_month:modBDM1	-0.005	0.004	-1.131	2651.658	0.258
udata_age_month:modBDM2	-0.002	0.004	-0.521	2651.658	0.603
udata_age_month:modBSM1	0.000	0.004	0.077	2651.656	0.939
icc:udata_age_month	-0.007	0.005	-1.465	2651.658	0.143
icc:udata_age_month:modBDM1	0.053	0.007	7.357	2651.665	0.000
icc:udata_age_month:modBDM2	0.049	0.007	6.908	2651.663	0.000
icc:udata_age_month:modBSM1	-0.001	0.007	-0.101	2651.656	0.920
sd__(Intercept)	0.211				
sd__Observation	0.835				

3.4 Supplementary analyses

3.4.1 Extended GLMM models

3.4.1.1 Relative prediction error of Bayesian models vs. GLMER4

On average, the relative difference of prediction error between these two models GLMER4 vs. BSM1 is $e^{-0.735}$ = 0.480 at U4. This indicates that GLMER4 predicts better than BSM1 at U4. This difference stays the same regardless of number of previous visits. The proportion of variance that can be explained by different visits is $\frac{0.340^2}{1.455^2+0.340^2+0.129^2} * 100\% = 5.139\%$. Table 38 shows the estimated coefficients of the LMM. Figure 63 shows the distribution of the LRPE between GLMER4 and BSM1 over time.

Table 38. Prediction performance of BSM1 vs. GLMER4 – quantified with LMM

group	term	estimate	std.error	statistic	df	p.value
	(Intercept)	-0.735	0.140	-5.240	5.024	3.31E-03
udata:PseudoID	sd__(Intercept)	1.455				
udata	sd__(Intercept)	0.340				
Residual	sd__Observation	0.129				

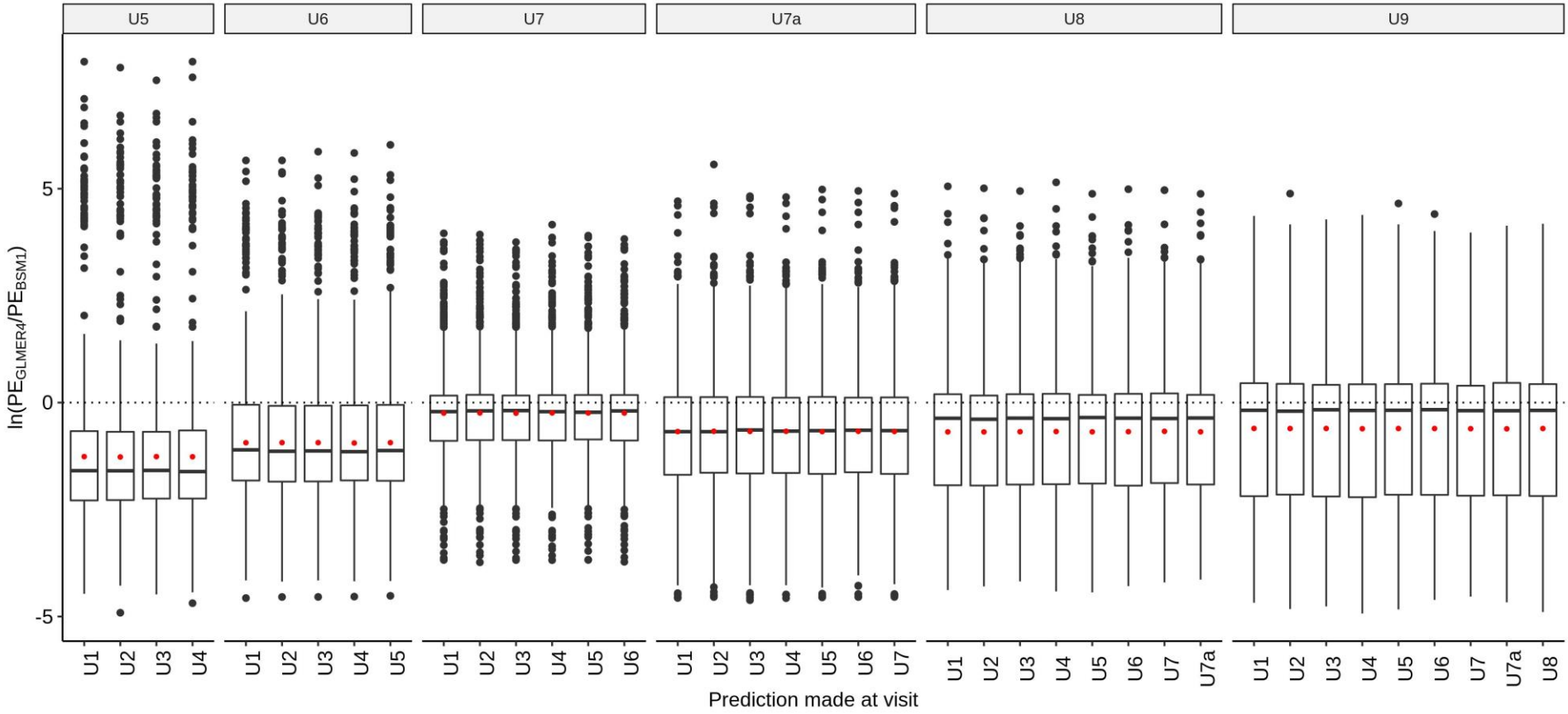


Figure 63. Prediction performance of GLMER4 vs. BSM1 over time in PEACHES study

On average, the relative difference of prediction error between these two models GLMER4 vs. BSM2 is $e^{-0.750}$ = 0.472 at U4. This indicates that GLMER4 predicts better than BSM2 at U4. This difference stays the same regardless of number of previous months. Table 39 shows the estimated coefficients of the LMM. Figure 64 shows the distribution of the LRPE between GLMER4 and BSM2 over time.

Table 39. Prediction performance of BSM2 vs. GLMER4 - quantified with LMM

group	**term**	**estimate**	**std.error**	**statistic**	**df**	**p.value**
	(Intercept)	-0.750	0.037	-20.375	970.223	4.21E-77
PseudoID	sd__(Intercept)	1.044				
Residual	sd__Observation	1.077				

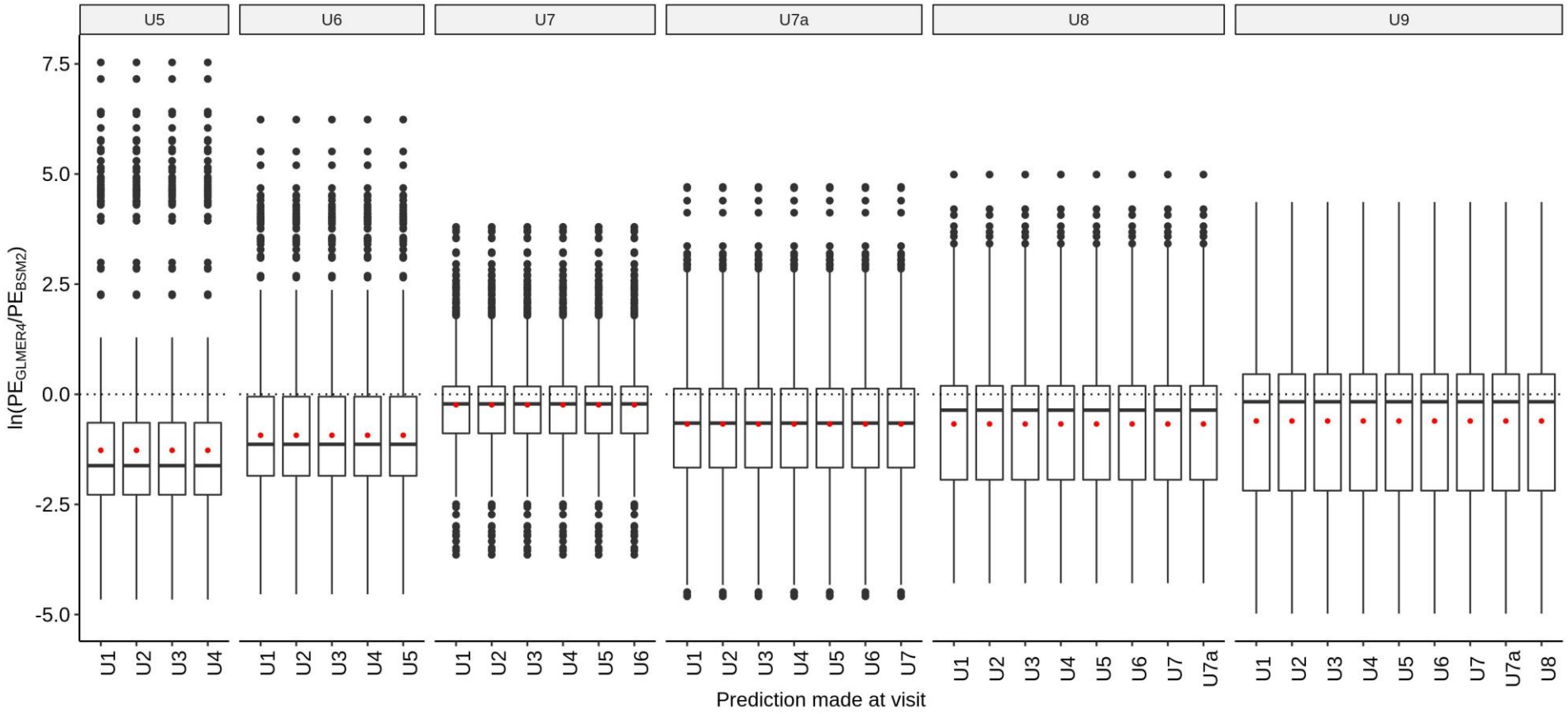


Figure 64. Prediction performance of GLMER4 vs. BSM2 over time in PEACHES study

On average, the relative difference of prediction error between these two models GLMER4 vs. BDM1 is 0.306 at U4. This indicates that GLMER4 predicts better than BDM1 at U4. However, this difference reduces over time in a linear manner, but the reduction attenuates over time due to the non-linearity of the time effect. The proportion of variance that can be explained by different visits is $\frac{0.638^2}{1.339^2+0.638^2+0.537^2} * 100\% = 16.358\%$. Table 40 shows the estimated coefficients of the LMM. Figure 65 shows the distribution of the LRPE between GLMER4 and BDM1 over time.

Table 40. Prediction performance of BDM1 vs. GLMER4 – quantified with LMM

group	term	estimate	std.error	statistic	df	p.value
	(Intercept)	-1.418	0.261	-5.432	5.008	2.85E-03
	months_from_u1	0.092	0.003	35.742	27401.288	1.68E-273
	I(months_from_u1^2)	-5.14E-03	3.00E-04	-17.123	27405.400	2.21E-65
	I(months_from_u1^3)	1.34E-04	1.11E-05	12.105	27408.208	1.20E-33
	I(months_from_u1^4)	-1.19E-06	1.24E-07	-9.591	27410.130	9.45E-22
udata:PseudoID	sd__(Intercept)	1.339				
udata	sd__(Intercept)	0.638				
Residual	sd__Observation	0.537				

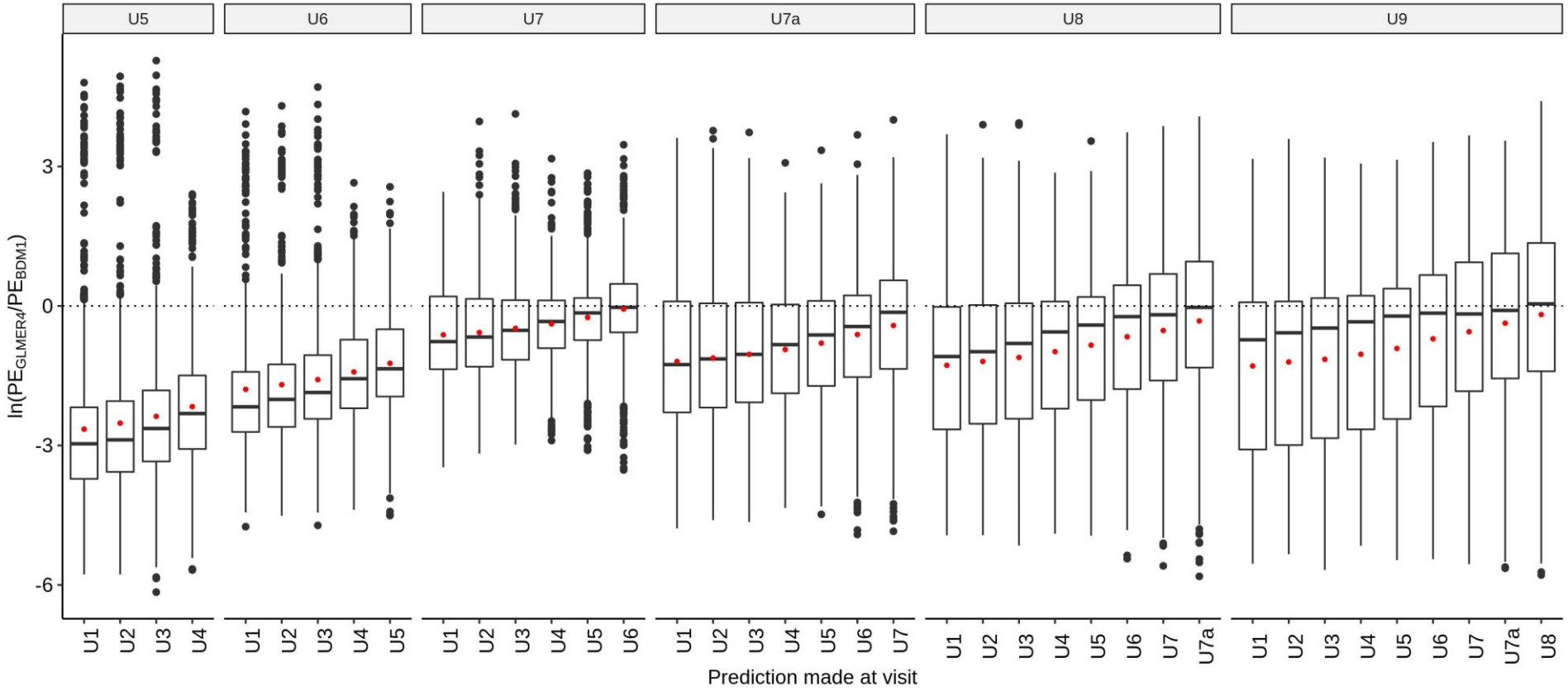


Figure 65. Prediction performance of GLMER4 vs. BDM1 over time in PEACHES study

On average, the relative difference of prediction error between these two models GLMER4 vs. BDM2 is 0.297 at U4. This indicates that GLMER4 predicts better than BDM2 at U4. However, this difference reduces over time in a linear manner but the reduction attenuates over time due to the non-linearity of the time effect. The proportion of variance that can be explained by different visits is $\frac{0.639^2}{1.339^2+0.639^2+0.542^2} * 100\% = 16.366\%$. Table 41 shows the estimated coefficients of the LMM. Figure 66 shows the distribution of the LRPE between GLMER4 and BDM2 over time.

Table 41. Prediction performance of BDM2 vs. GLMER4 – quantified with LMM

group	term	estimate	std.error	statistic	df	p.value
	(Intercept)	-1.422	0.262	-5.436	5.006	2.85E-03
	months_from_u1	0.095	0.003	36.298	27401.459	8.25E-282
	I(months_from_u1^2)	-5.36E-03	3.03E-04	-17.692	27405.646	1.17E-69
	I(months_from_u1^3)	1.42E-04	1.12E-05	12.683	27408.506	9.28E-37
	I(months_from_u1^4)	-1.27E-06	1.25E-07	-10.171	27410.463	2.94E-24
udata:PseudoID	sd__(Intercept)	1.339				
udata	sd__(Intercept)	0.639				
Residual	sd__Observation	0.542				

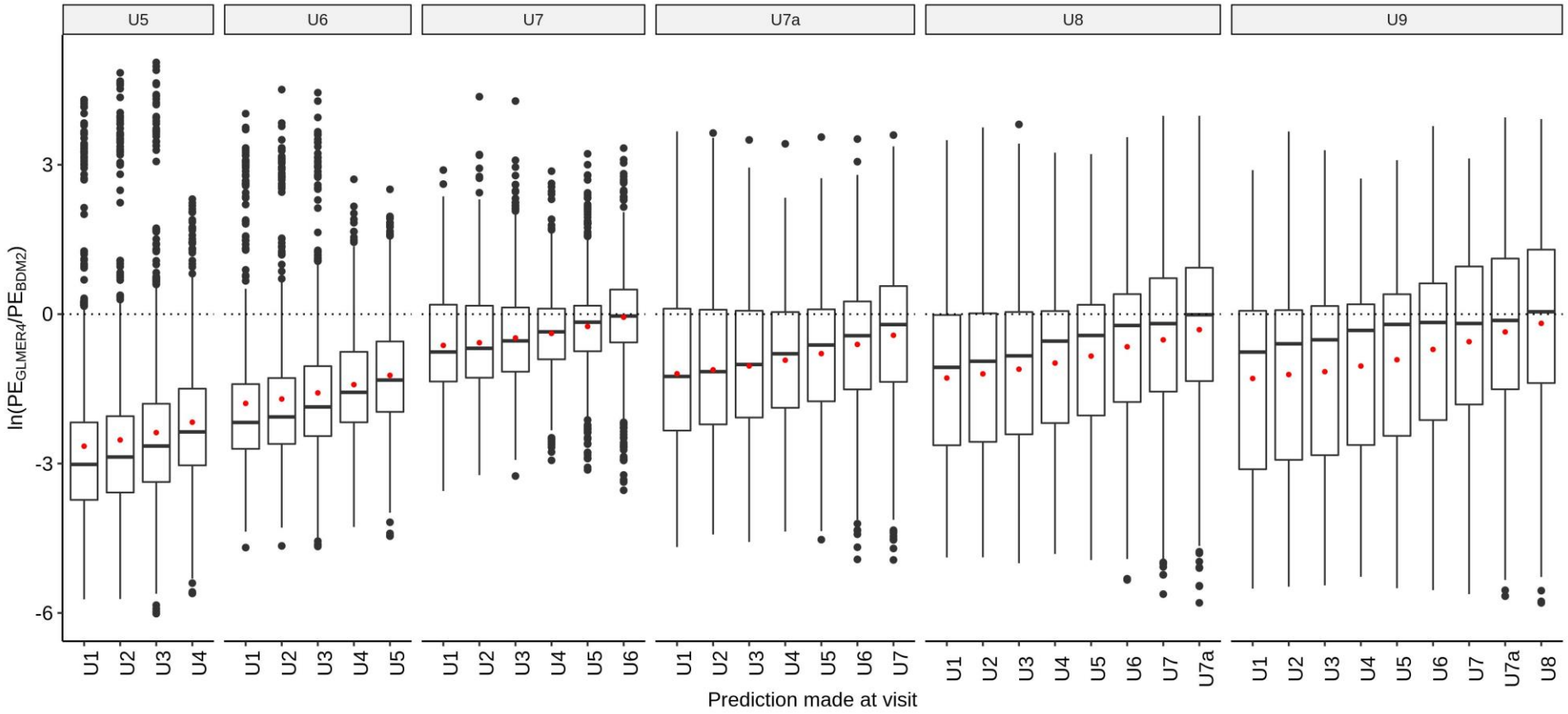


Figure 66. Prediction performance of GLMER4 vs. BDM2 over time in PEACHES study

Overall time effect (see 2.8) of the four Bayesian models is shown in the following table, when incorporating the non-linearity of time effects. RPE of GLMER4 versus BSM1 and BSM2 was shown to be around 1 over time, while RPE of GLMER1 versus BDM1 and BDM2 seems to increase with increasing number of months from U1. These observations do not apply for timepoint 60 months (U9). At this timepoint, GLMER4 performs more poorly than BSM1 and BSM2 and better than BDM1 and BDM2.

Table 42. RPE of GLMER4 vs. Bayesian models in PEACHES study

Number of months from U1	**RPE of GLMER4 vs.**			
	BSM1	**BSM2**	**BDM1**	**BDM2**
3	0.480	0.472	0.306	0.297
6	0.480	0.472	0.360	0.352
12	0.480	0.472	0.430	0.423
24	0.480	0.472	0.493	0.485
36	0.480	0.472	0.605	0.597
48	0.480	0.472	0.725	0.713
60	0.480	0.472	0.431	0.400

3.4.1.2 Brier Score of Bayesian vs. extended GLMM models

Brier Score of the four Bayesian vs. four GLMM models are shown in Figure 67, where all GLMM models are depicted in grey. The curves of four GLMM models (four grey curves) and BSM2 are rigid over time and stay the same regardless of at which visit prediction was made. BSM1 and BSM2 are very close to each other. The same can be observed for BDM1 and BDM2. The prediction performance of these two models improves with increasing number of past visits. From prediction visit U6, the improvement of BDM1 and BDM2 becomes more pronounced. Overall, BSM1 and BSM2 show the worst prediction performance. Figure 68 depicts the four GLMM models distinctively and shows the prediction performance of all eight models from prediction visit U4. Among the GLMM models, GLMER4 (depicted with purple curve) shows the best prediction performance and GLMER1 (depicted with black curve) shows the worst prediction performance.

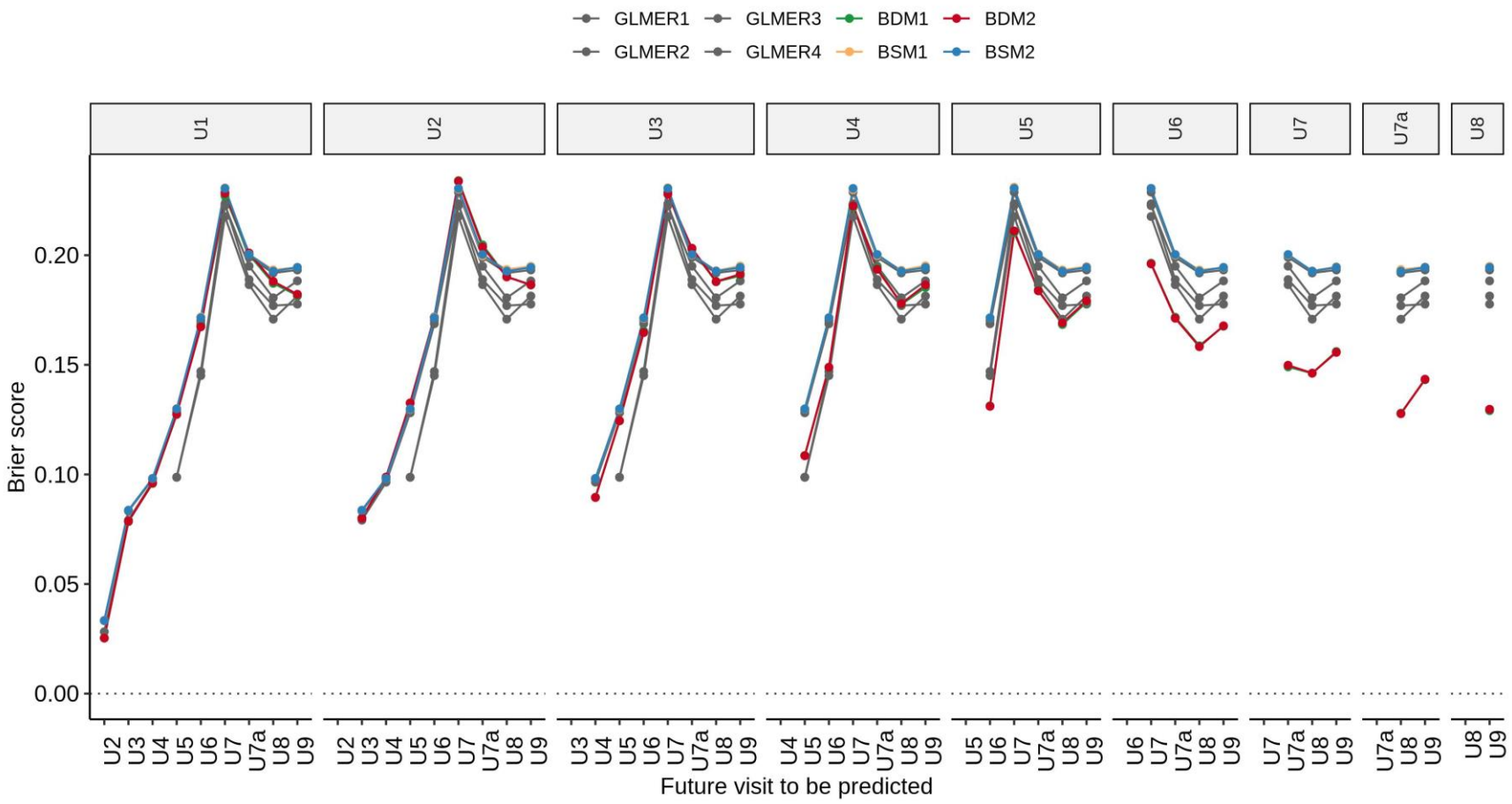


Figure 67. Brier Score of Bayesian vs. all GLMM models in PEACHES study

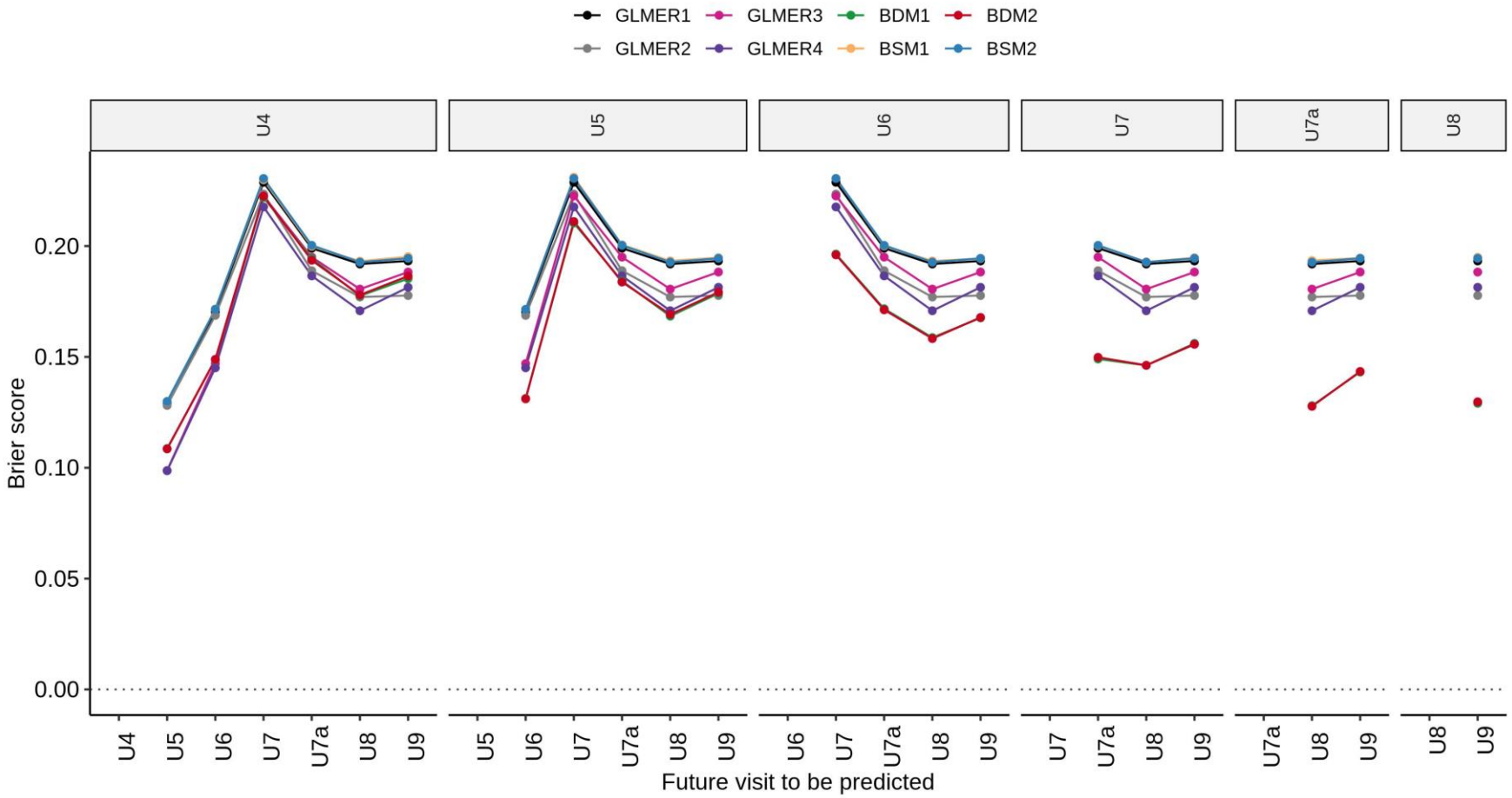


Figure 68. Brier Score of Bayesian vs. all GLMM models, from U4 in PEACHES study

3.4.1.3 Calibration of Bayesian models vs. GLMER4

The calibration of GLMER4 and BSM2 predicting a certain future visit stays the same, regardless at which visit the prediction was made. GLMER4 shows better calibration than the Bayesian models at earlier prediction visits. BDM1 and BDM2 show improvement in calibration with increasing number of past visits and overperforms GLMER4 at the last or second last prediction visit. Calibration of BSM1 seems to change over time. However, this change does not show a consistent trend over time. Results of the calibration of GLMER4 and the Bayesian models are shown in the following table.

Table 43. Calibration of Bayesian models vs. GLMER4

Future visit	Prediction visit	GLMER4 Slope	GLMER4 Intercept	BSM1 Slope	BSM1 Intercept	BSM2 Slope	BSM2 Intercept	BDM1 Slope	BDM1 Intercept	BDM2 Slope	BDM2 Intercept
U5	U1	0.828	0.066	0.248	0.129	0.329	0.124	0.224	0.112	0.218	0.113
U5	U2	0.828	0.066	0.619	0.108	0.329	0.124	0.139	0.124	0.138	0.124
U5	U3	0.828	0.066	0.849	0.096	0.329	0.124	0.433	0.084	0.430	0.085
U5	U4	0.828	0.066	0.523	0.114	0.329	0.124	0.864	0.026	0.867	0.025
U6	U1	0.768	0.103	0.647	0.134	0.847	0.111	0.330	0.136	0.346	0.133
U6	U2	0.768	0.103	0.705	0.127	0.847	0.111	0.265	0.151	0.269	0.150
U6	U3	0.768	0.103	0.904	0.105	0.847	0.111	0.454	0.112	0.467	0.109
U6	U4	0.768	0.103	0.735	0.123	0.847	0.111	0.770	0.046	0.763	0.047
U6	U5	0.768	0.103	0.838	0.112	0.847	0.111	1.005	-0.006	1.005	-0.006
U7	U1	0.665	0.169	0.259	0.265	0.332	0.250	0.176	0.265	0.134	0.277
U7	U2	0.665	0.169	0.376	0.241	0.332	0.250	0.078	0.295	0.084	0.293
U7	U3	0.665	0.169	0.248	0.267	0.332	0.250	0.283	0.235	0.300	0.230
U7	U4	0.665	0.169	0.427	0.230	0.332	0.250	0.427	0.192	0.418	0.194
U7	U5	0.665	0.169	0.236	0.269	0.332	0.250	0.584	0.143	0.575	0.146
U7	U6	0.665	0.169	0.306	0.255	0.332	0.250	0.725	0.102	0.728	0.102
U7a	U1	0.719	0.132	0.594	0.157	0.600	0.156	0.274	0.187	0.248	0.194
U7a	U2	0.719	0.132	0.735	0.133	0.600	0.156	0.201	0.209	0.225	0.202
U7a	U3	0.719	0.132	0.651	0.147	0.600	0.156	0.319	0.176	0.309	0.179
U7a	U4	0.719	0.132	0.716	0.136	0.600	0.156	0.485	0.130	0.497	0.127
U7a	U5	0.719	0.132	0.542	0.167	0.600	0.156	0.620	0.091	0.624	0.090
U7a	U6	0.719	0.132	0.681	0.142	0.600	0.156	0.742	0.058	0.747	0.057
U7a	U7	0.719	0.132	0.643	0.149	0.600	0.156	0.970	-0.009	0.958	-0.006
U8	U1	0.876	0.099	0.515	0.166	0.710	0.136	0.424	0.138	0.392	0.146
U8	U2	0.876	0.099	0.462	0.174	0.710	0.136	0.349	0.161	0.353	0.159
U8	U3	0.876	0.099	0.629	0.149	0.710	0.136	0.438	0.137	0.437	0.138
U8	U4	0.876	0.099	0.585	0.156	0.710	0.136	0.627	0.090	0.620	0.092
U8	U5	0.876	0.099	0.520	0.165	0.710	0.136	0.734	0.062	0.720	0.065
U8	U6	0.876	0.099	0.538	0.163	0.710	0.136	0.818	0.041	0.823	0.040
U8	U7	0.876	0.099	0.691	0.140	0.710	0.136	0.948	0.004	0.944	0.005
U8	U7a	0.876	0.099	0.476	0.172	0.710	0.136	1.117	-0.037	1.116	-0.036
U9	U1	0.671	0.136	0.451	0.181	0.424	0.185	0.529	0.120	0.509	0.125
U9	U2	0.671	0.136	0.257	0.206	0.424	0.185	0.382	0.157	0.376	0.158
U9	U3	0.671	0.136	0.216	0.212	0.424	0.185	0.312	0.171	0.297	0.174
U9	U4	0.671	0.136	0.135	0.222	0.424	0.185	0.448	0.140	0.427	0.145
U9	U5	0.671	0.136	0.309	0.200	0.424	0.185	0.556	0.114	0.547	0.116
U9	U6	0.671	0.136	0.476	0.178	0.424	0.185	0.695	0.084	0.691	0.085
U9	U7	0.671	0.136	0.323	0.198	0.424	0.185	0.825	0.052	0.827	0.052
U9	U7a	0.671	0.136	0.362	0.193	0.424	0.185	0.959	0.024	0.954	0.025
U9	U8	0.671	0.136	0.264	0.205	0.424	0.185	1.085	-0.005	1.074	-0.002

3.4.1.4 Scaled Brier Score of Bayesian models vs. GLMER4

Results of SBS show overall overperformance of BDM1 and BDM2 compared to GLMER4 and BSM1 as well as BSM2. Only at U6 and later visits, BDM1 and BDM2 clearly perform better than a non-informative model.

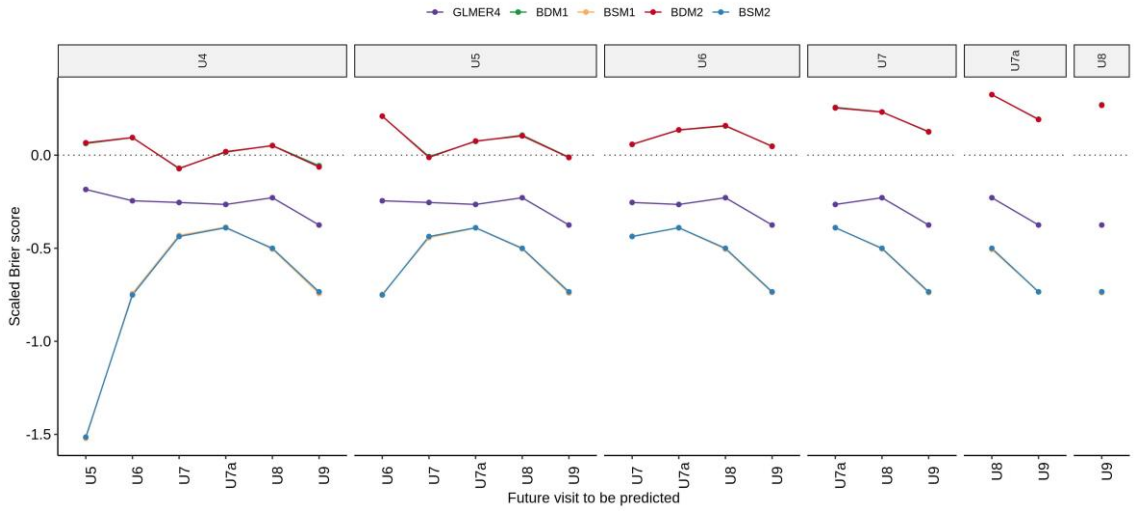


Figure 69. Scaled Brier Score over time in PEACHES study

3.4.1.5 Brier Skill Score of Bayesian models vs. GLMER4

When considering GLMER4 as the reference model, BSS of the Bayesian models show consistent overperformance of BDM1 and BDM2 from prediction visit U5. BSM1 and BSM2 performs consistently worse than GLMER4.

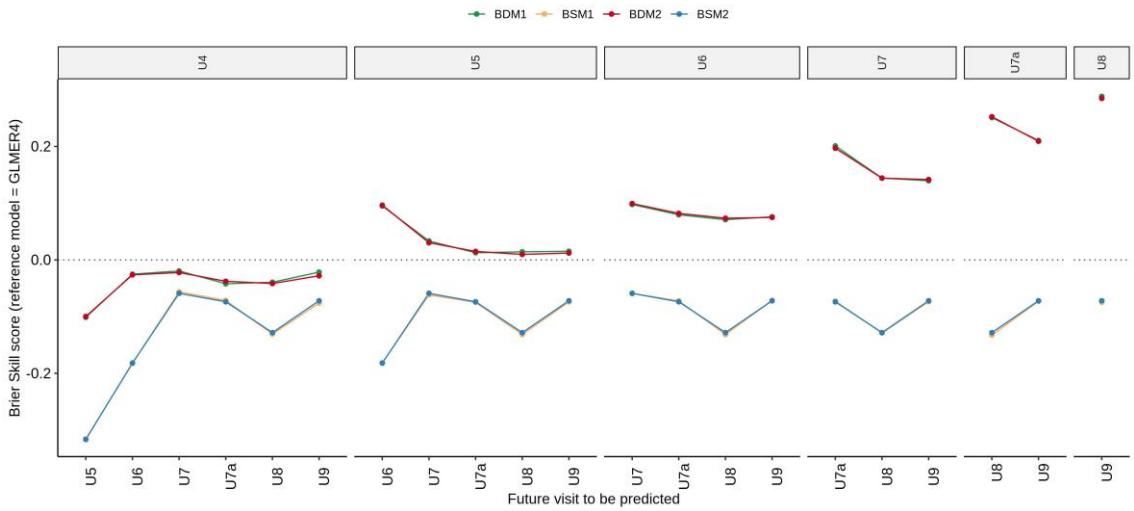


Figure 70. Brier Skill Score vs. GLMER4 over time in PEACHES study

3.4.2 Comparing models using error rate

Using error rate reveals overall similar results of measures related to absolute predicted probability and Brier Score. However, the distinction between models seems to be less pronounced if comparing models using error rate. Also, the small changes of BSM1 over time become indetectable if using error rate. For a certain future visit, BSM1 shows the same error rate regardless of the prediction visit considered. These observations apply for the PEACHES (Table 44, Figure 72 and Figure 71) as well as the simulation main study (Table 45 and Figure 73).

Table 44. Error rate of Bayesian models vs. GLMER1 over time in PEACHES study

Prediction made at timepoint	Future timepoint to be predicted	Error rate				
		GLMER1	BSM1	BSM2	BDM1	BDM2
U1	U2	0.034	0.034	0.034	0.034	0.034
U1	U3	0.088	0.088	0.088	0.088	0.088
U1	U4	0.104	0.104	0.104	0.103	0.103
U1	U5	0.142	0.142	0.142	0.150	0.152
U1	U6	0.205	0.205	0.205	0.232	0.231
U1	U7	0.317	0.317	0.317	0.338	0.338
U1	U7a	0.262	0.262	0.262	0.282	0.283
U1	U8	0.245	0.245	0.245	0.259	0.259
U1	U9	0.240	0.240	0.240	0.245	0.245
U2	U3	0.088	0.088	0.088	0.100	0.100
U2	U4	0.104	0.104	0.104	0.121	0.121
U2	U5	0.142	0.142	0.142	0.162	0.162
U2	U6	0.205	0.205	0.205	0.231	0.228
U2	U7	0.317	0.317	0.317	0.339	0.339
U2	U7a	0.262	0.262	0.262	0.285	0.285
U2	U8	0.245	0.245	0.245	0.259	0.260
U2	U9	0.240	0.240	0.240	0.246	0.246
U3	U4	0.104	0.104	0.104	0.114	0.110
U3	U5	0.142	0.142	0.142	0.162	0.163
U3	U6	0.205	0.205	0.205	0.218	0.220
U3	U7	0.317	0.317	0.317	0.332	0.332
U3	U7a	0.262	0.262	0.262	0.293	0.294
U3	U8	0.245	0.245	0.245	0.264	0.267
U3	U9	0.240	0.240	0.240	0.263	0.265
U4	U5	0.142	0.142	0.142	0.148	0.154
U4	U6	0.205	0.205	0.205	0.204	0.205
U4	U7	0.317	0.317	0.317	0.323	0.325
U4	U7a	0.262	0.262	0.262	0.276	0.280
U4	U8	0.245	0.245	0.245	0.253	0.253
U4	U9	0.240	0.240	0.240	0.251	0.256
U5	U6	0.205	0.205	0.205	0.180	0.183
U5	U7	0.317	0.317	0.317	0.301	0.304
U5	U7a	0.262	0.262	0.262	0.253	0.254
U5	U8	0.245	0.245	0.245	0.232	0.232
U5	U9	0.240	0.240	0.240	0.248	0.246
U6	U7	0.317	0.317	0.317	0.292	0.292
U6	U7a	0.262	0.262	0.262	0.244	0.243
U6	U8	0.245	0.245	0.245	0.222	0.224
U6	U9	0.240	0.240	0.240	0.236	0.241
U7	U7a	0.262	0.262	0.262	0.212	0.222
U7	U8	0.245	0.245	0.245	0.209	0.206
U7	U9	0.240	0.240	0.240	0.214	0.216
U7a	U8	0.245	0.245	0.245	0.182	0.185
U7a	U9	0.240	0.240	0.240	0.206	0.209
U8	U9	0.240	0.240	0.240	0.188	0.188

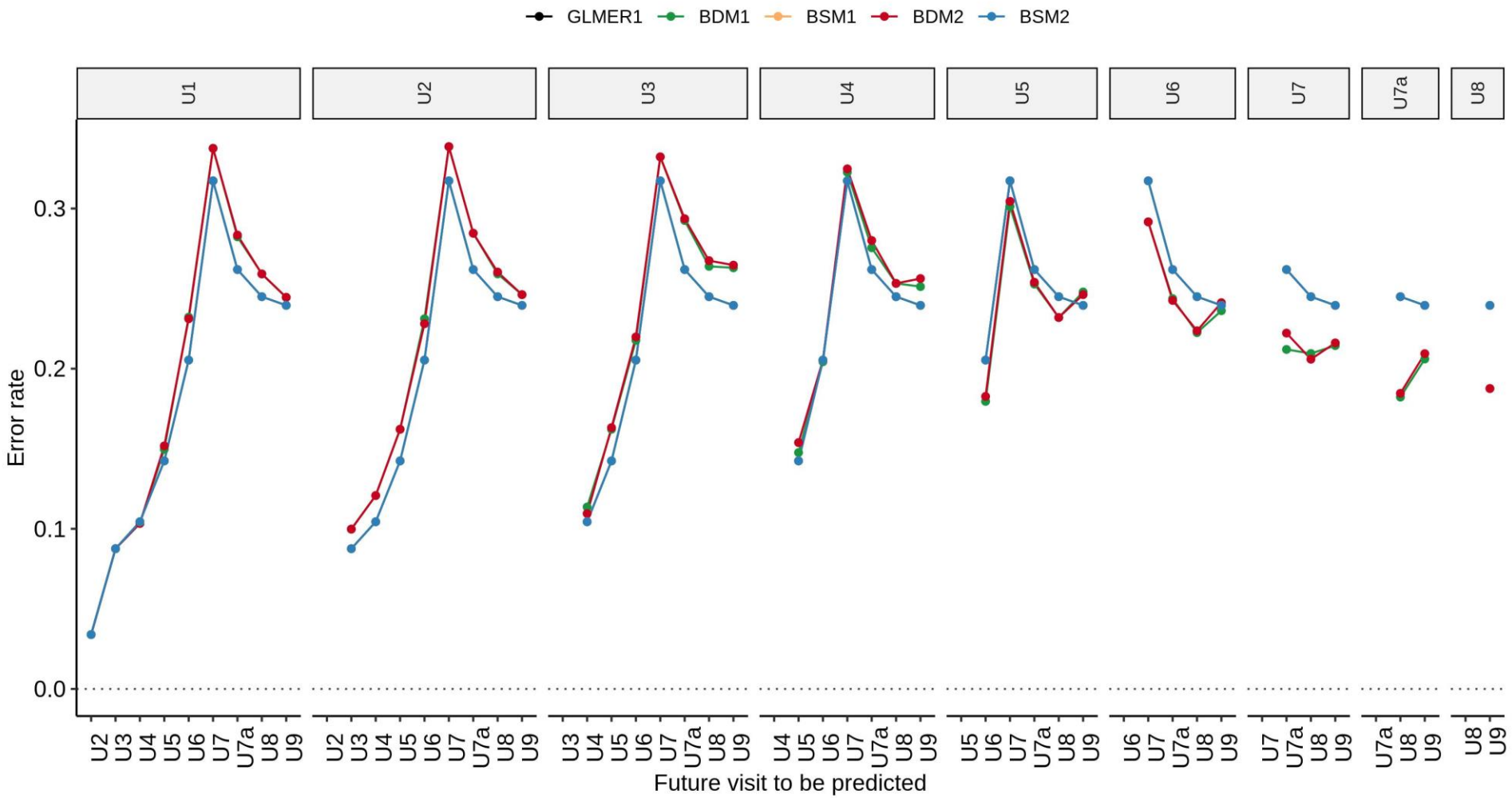


Figure 71. Error rate of Bayesian models vs. GLMER1 over time – PEACHES study

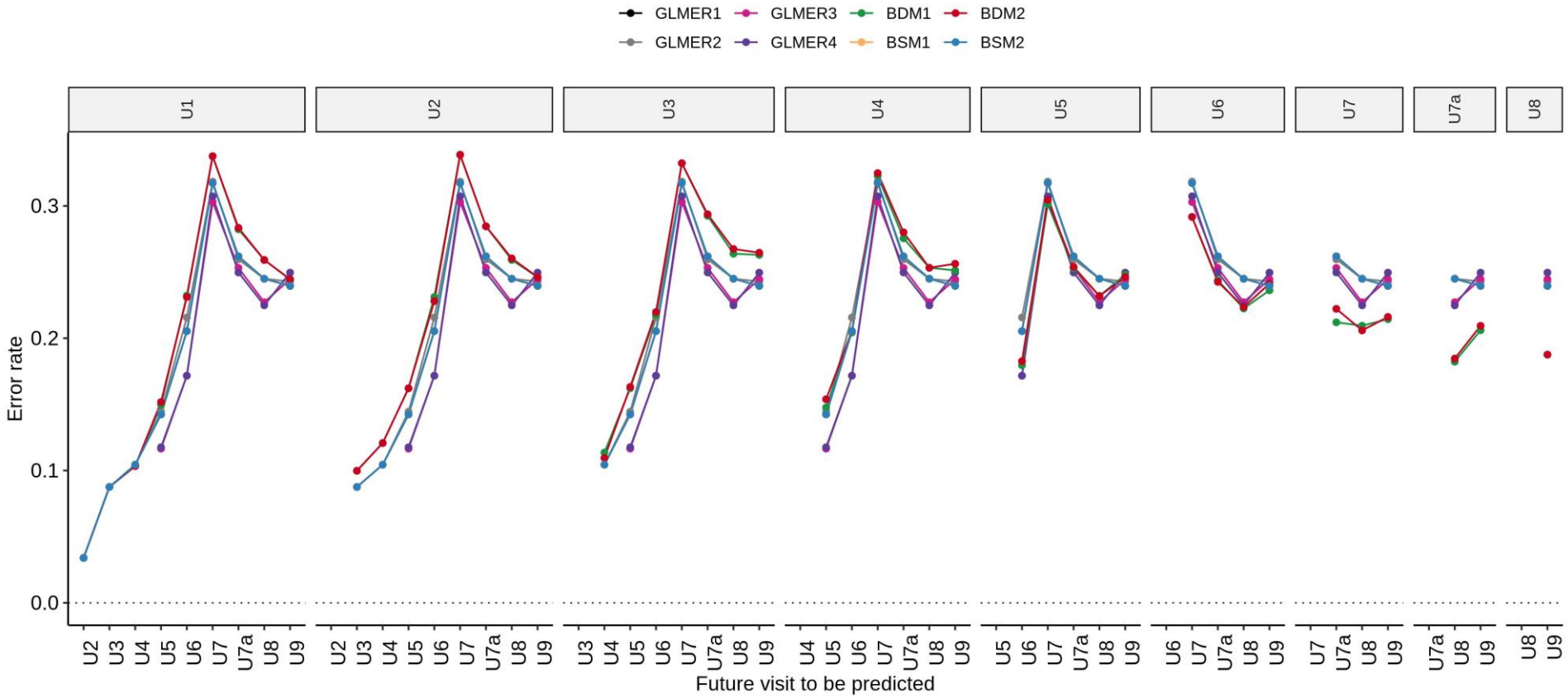


Figure 72. Error rate of Bayesian models vs. GLMM models over time – PEACHES study

Table 45. Error rate of Bayesian models vs. GLMER1 over time in simulation main study

Prediction made at	Future timepoint	Error rate				
		GLMER1	BSM1	BSM2	BDM1	BDM2
U1	U2	0.054	0.054	0.054	0.054	0.054
U1	U3	0.057	0.057	0.057	0.057	0.057
U1	U4	0.107	0.107	0.107	0.107	0.107
U1	U5	0.113	0.113	0.113	0.112	0.115
U1	U6	0.161	0.161	0.161	0.167	0.167
U1	U7	0.222	0.222	0.222	0.204	0.204
U1	U7a	0.199	0.199	0.199	0.191	0.191
U1	U8	0.183	0.183	0.183	0.167	0.167
U1	U9	0.158	0.158	0.158	0.156	0.156
U2	U3	0.057	0.057	0.057	0.061	0.061
U2	U4	0.107	0.107	0.107	0.103	0.103
U2	U5	0.113	0.113	0.113	0.107	0.107
U2	U6	0.161	0.161	0.161	0.164	0.162
U2	U7	0.222	0.222	0.222	0.198	0.198
U2	U7a	0.199	0.199	0.199	0.184	0.185
U2	U8	0.183	0.183	0.183	0.165	0.163
U2	U9	0.158	0.158	0.158	0.152	0.153
U3	U4	0.107	0.107	0.107	0.103	0.104
U3	U5	0.113	0.113	0.113	0.110	0.110
U3	U6	0.161	0.161	0.161	0.150	0.150
U3	U7	0.222	0.222	0.222	0.194	0.190
U3	U7a	0.199	0.199	0.199	0.185	0.180
U3	U8	0.183	0.183	0.183	0.163	0.166
U3	U9	0.158	0.158	0.158	0.143	0.143
U4	U5	0.113	0.113	0.113	0.116	0.114
U4	U6	0.161	0.161	0.161	0.151	0.151
U4	U7	0.222	0.222	0.222	0.194	0.195
U4	U7a	0.199	0.199	0.199	0.179	0.179
U4	U8	0.183	0.183	0.183	0.161	0.161
U4	U9	0.158	0.158	0.158	0.134	0.134
U5	U6	0.161	0.161	0.161	0.145	0.144
U5	U7	0.222	0.222	0.222	0.190	0.190
U5	U7a	0.199	0.199	0.199	0.173	0.173
U5	U8	0.183	0.183	0.183	0.153	0.153
U5	U9	0.158	0.158	0.158	0.138	0.138
U6	U7	0.222	0.222	0.222	0.184	0.184
U6	U7a	0.199	0.199	0.199	0.163	0.169
U6	U8	0.183	0.183	0.183	0.156	0.160
U6	U9	0.158	0.158	0.158	0.131	0.131
U7	U7a	0.199	0.199	0.199	0.163	0.163
U7	U8	0.183	0.183	0.183	0.147	0.147
U7	U9	0.158	0.158	0.158	0.131	0.129
U7a	U8	0.183	0.183	0.183	0.152	0.150
U7a	U9	0.158	0.158	0.158	0.122	0.122
U8	U9	0.158	0.158	0.158	0.125	0.131

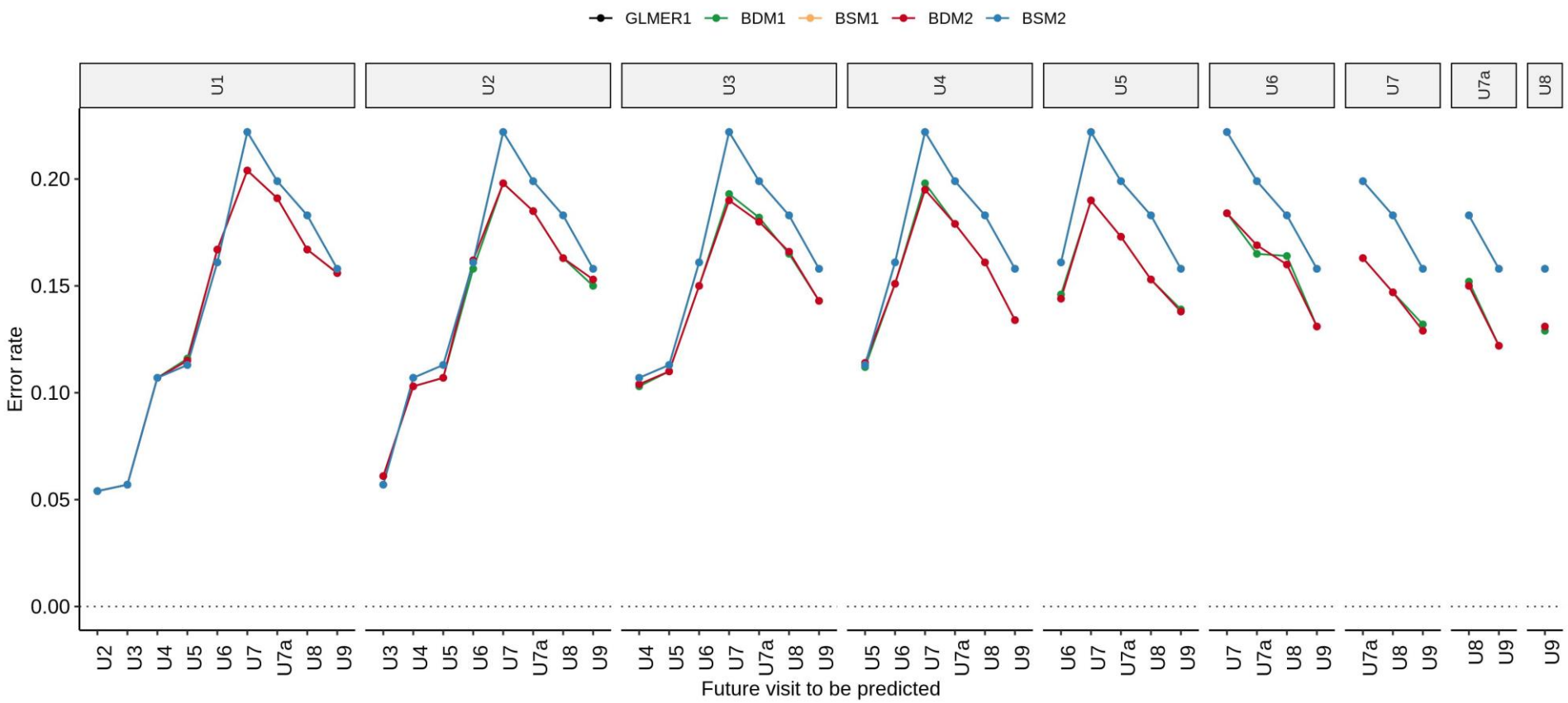


Figure 73. Error rate of Bayesian models vs. GLMER1 over time - simulation main study

3.4.3 Analyses with smaller sample size

For PEACHES and the simulation main study, analyses with sample size of 100 do not differ than those with sample size of 1,000. See Appendix B3 (provided in the folder `supplementary_material/)` for detailed results of analyses with sample size of 100.

3.5 Literature search

Literature search in Google Scholar identified 24 cited and 101 related articles. After removing duplicates, 99 articles were identified. After title screening, 22 articles were considered as relevant to enter screening of abstract and full text, if available. 38 references of Jenkins et al. 2018 (Jenkins et al., 2018) and 37 references of Jenkins et al. 2021 (Jenkins et al., 2021) were also taken into consideration to identify methods dealing with dynamic prediction models in clinical or epidemiological context.

After the first screening (by their titles), 41 articles were selected and assigned into different groups: 1) systematic or non-systematic review about dynamic prediction, 2) dynamic prediction modelling approaches, 3) related methodological aspects, and 4) examples of dynamic prediction models in different clinical contexts. Refer to Appendix C: Results of literature search for the list of the selected articles.

3.6 Summary of the results

In PEACHES study, BSM1 and BSM2 show overall similar prediction errors as GLMER1. Even though these two Bayesian models seem to perform better than GLMER1 at early visits. However, this overperformance is marginal. On the other hand, GLMER1 overperforms BDM1 and BDM2 at early visits. At later visits, BDM1 and BDM2 start overperforming GMLER1 but then performs more poorly at the end of the study time period. Brier Score of BSM1 and BSM2 are comparable with GLMER1's Brier Score, while model updating results in better prediction performance over time. From visit U4, better calibration of BDM1 and BDM2 compared to GLMER1 can be observed. Calibration of BDM1 and BDM2 improves with increasing amount of available information for prediction, while BSM1, BSM2, and GLMER1 show bad calibration, regardless of which prediction visit is considered. Comparing Bayesian models versus GLMER1 using scaled Brier Score and Brier Skill Score revealed overperformance of BDM1 and BDM2 to BSM1, BSM2, and GLMER1. This overperformance can be observed at later visits. BSM1 and BSM2 show similar performance as GLMER1. Brier Skill Score with "U1-model" as reference showed that overperformance to "U1-model" can be observed for BDM1 and BDM2 at later visits. Results of the comparison with GLMER4 show that within the observation period, GLMER4 consistently overperforms Bayesian models at early visits. However, it can be shown that with increasing amount of information, model updating in BDM1 and BDM2 leads to improvement in prediction. Results of the analyses with error rates show similar but less pronounced results. Overall, results of the analyses with sample size of 100 are comparable those with sample size of 1,000.

Results of the simulation main study agreed with those of PEACHES study. The analyses of overall time effect on relative prediction errors in simulation main study showed that GLMER1 performs more poorly than BSM1 and BSM2, regardless of the amount of data used for prediction. The overperformance of BDM1 and BDM2, which was observed with Brier Skill Score and Scaled Brier Score seems to be more pronounced in the simulation settings. Results of the simulation ICC study show that ICC strongly influences the better prediction performance of BDM1 and BDM2 compared with GLMER1. The higher ICC, in other words, the higher the variance of random intercepts is, the more pronounced is the overperformance of BDM1 and BDM2. Results of analyses at individual level in the simulation ICC study show that the more amount of information from past visits, the better BDM1 and BDM2 can capture the overall distribution of the simulated random intercepts.

4. Discussion

This section will discuss the following points: 1) results of PEACHES study, 2) results of simulation main study and the agreement as well as disagreement of PEACHES vs. simulation study, 3) results of simulation ICC study and the role of ICC on usefulness of dynamic approaches, 4) presumptions made in this thesis, 5) well-known challenges of dynamic approaches as well as new insights that this thesis gives, and 6) limitations of this thesis and perspective for future research,

Results of PEACHES study

Different measures of prediction performance, which are Brier Score, calibration, scaled Brier Score, LRPE, and Brier Skill Score provided different perspectives but agreed on the benefit of dynamic models (BDM1 and BDM2) at later visits. While Brier Score, calibration, and scaled Brier Score are absolute measures that stand alone for each model, LRPE and Brier Skill Score consider GLMER1 as reference and serve as relative comparison measures. Because of the static character of BSM2 and GLMER1, it is expected that the absolute prediction performance measures of these models stay constant, regardless of at which visit the prediction was made. It is also expected that BDM1 and BDM2, due to their capability of updating and "adapting" the prediction according to the amount of available information, offer better prediction performance. Both expectations could be observed in this research work. BSM1 owns a semi-dynamic character and provides updates of the outcome according to updated outcome values from previous values. The model specification, however, stays the same over time. It was interesting to see that the flexibility that BSM1 offers is not sufficient to improve prediction performance. At the aggregated level, no profound difference between BDM1 and BDM2 could be observed. This indicates that in this study setting, updating fixed effects does not add any benefit to improve prediction performance of the models. Thus, it can be concluded that updating random effects leads to the improvement of the two dynamic approaches over time.

One of the objectives of this research work was to examine from which time point the overperformance of the dynamic approaches becomes visible. Graphically, Brier Score, calibration, scaled Brier Score, and Brier Skill Score showed that this turning point seems to happen at U4 (about 3 months after birth) or U5 (about 6 months after birth). I believe it could not be observed earlier because of the nature of this study setting, and not necessarily because of the model specifications. In the first living months, weight and height measurements are more prone to measurement errors of different sources (Alsop-Shields and Alexander, 1997). Measurement errors introduced by deviating measuring practices at different medical facilities, with different newborn scales (instruments that are used to measure newborns' weight and height) or by different pediatric nurses also play a role in introducing randomness of measurement. These again are difficult to be incorporated into and explained by regression models. In this way, measurement errors can affect model's prediction performance substantially.

While Brier Score, calibration, scaled Brier Score, and Brier Skill Score describe models' prediction performance at an aggregated level, LMMs with LRPE attempted to quantify the difference at an individual level. Results of the LMM give insights about the turning point of improvement of Bayesian models for a single individual for the same future visit. These results suggest that for

BSM1 and BSM2, there is no such turning point, while for both BDM1 and BDM2 this turning point happens between U7 (24 months after birth) and U7a (36 months after birth).

In the LMMs quantifying the influence of amount of information on LRPE of GLMER1 vs. BSM1, BDM1, and BDM2, random intercepts on the level of prediction visit and on the individual level were allowed in the model specification. It was necessary since predictions for the same visit and of the same individual are naturally correlated. Analysis of variance shows that for BSM1, BDM1, and BDM2, the fitted LMM can describe the association of time and LRPE quite well, while the course of LRPE over time between GLMER1 and BSM2 can rather be explained with randomness. The LMMs estimated the proportion of variances explained by different future visits and these numbers are comparable among BSM1, BDM1, and BDM2. These numbers are 19.37%, 20.85%, and 20.63%, respectively. For BSM2, since prediction was made once at U1, this number is practically 0%. The proportion of variances explained by individuals are 45.60%, 49.08%, 49.05% for BSM1, BDM1, and BDM2, respectively. This number is 9.55% for BSM2.

Regarding results of GLMER1 vs. BSM1 and BSM2: even though the prediction mechanism of BSM2 and GLMER1 is similar, the probability of being at risk of overweight predicted by these two models showed deviations. These observations are expected, because GLMER1 uses the log-odd link function to get the binary outcome, while BSM2 uses Bernoulli distribution to obtain the outcome. Even though BSM1 allows no model updating, it uses the available outcome values until the time of prediction as part of the data and get the predicted outcome. Therefore, BSM1 offers more flexibility than BSM2 and GLMER1, in a way that outcome is updated

Regarding results at individual level of selected children, BDM1 and BDM2 show certain differences. Since fixed effects are not updated over time, if bias is introduced at U1, it will not be adjusted for in BDM1, while BDM2 provides updates of these estimates over time. BDM2 thus shows lower bias with respect to model estimate of the fixed effects. On the other hand, BDM1 seems to adjust the random intercept to adapt to changes in outcome over time. The detailed update process of model components in ten individuals were depicted to examine the effect of updating over time for BDM1 and BDM2. Expected values of random intercepts were shown to develop in different directions but standard error seemed to decrease over time. The updates of outcome over time seemed to directly relate to the updates of random intercepts. No timely trend in the development of the fixed effects were observed. Since only ten individuals were examined, these results cannot be generalized.

The updating effect could also be observed at the population level. While estimate of each random intercept tends to become more precise over time, variance of the overall distribution of RI increases with the increasing amount of available information from previous visits.

Results of simulation main study

All in all, results of the simulation main study show great agreement with the findings in PEACHES study. This simulation study aimed to imitate the study settings but introduce more control over the data. Randomness due to potential measurement errors was introduced into the model in a controlled way. At the aggregated level, Brier Score, calibration, and scaled Brier score agree on the consistent overperformance of BDM1 and BDM2 from prediction visit U1 (at birth). Interestingly, the results at individual level revealed that within the simulated time frame (birth until about five years after birth), no turning point could be observed, while results of PEACHES study suggest that BDM1 and BDM2 overperform GLMER1 at U7a and U8. Despite this deviation, I believe results of both studies agree in one point: since LRPE describes relative prediction performance

at individual level, it is inadequate to describe the overall performance of a model. For this reason, it fails to detect the overperformance of BDM1 and BDM2 at the aggregated level.

Results at individual level at time point U9 differ from what has been discussed in the previous. Specifically, while in PEACHES and simulation main study, from U1 to U8, LRPE seemed to decrease, the timely trend of LRPE for BDM1 and BDM2 was not observed for U9. Because of the substantial proportion of missing data at U9, which is about 40% (Table 2), results at this timepoint need be interpreted with care.

Results of simulation ICC study

Among Bayesian models, regardless of ICC, the prediction performance of BSM1 vs. BSM2 as well as the prediction performance of BDM1 vs. BDM2 are comparable. This implies that rather updating random intercepts contributes on the improvement of prediction performance. Using all available outcome from previous visits to update the future outcome is not sufficient to improve prediction performance. Updating model's fixed effects over time introduces extra flexibility into the model but might not be necessary in this study setting.

Comparing Bayesian models with GLMER1 in different ICC scenarios revealed a significant influence of ICC on the difference of models' prediction performance. With increasing ICC, the overperformance of the Bayesian models is more distinctive. The overperformance of Bayesian approaches start showing at ICC about 0.5. With the same study settings, dynamic approaches might only be relevant for populations with ICC greater than 0.5. For populations with ICC smaller than 0.5, traditional approaches might be more suitable. This points out that the clinical relevance of the overperformance needs always be considered when assessing the usefulness of dynamic versus traditional prediction approaches.

This research work also aimed to answer the question to what extent ICC would influence the overperformance of dynamic approaches, if any observed. Results of the ICC study revealed that ICC as well as its interaction with amount of available information (measured with number of previous months) significantly associate with the improvement of BDM1 and BDM2 compared with GLMER1, while the amount of available information alone does not show any significant effect.

Description of the updating process at individual level revealed that with higher ICC, the updating effect described above for PEACHES study is visible at earlier visits. As expected, for ICC = 0, which means the random intercepts are homogenously equal to 0, such updating effect was not observed, either at population or at individual level.

Presumptions

It was assumed that it is common practice in Germany for parents to comply with the recommended time frame of the well-child visits. Therefore, in most of the analyses, discrete time variables presented as well-child visits were considered. This simplification in presenting the results might fail to visualize the actual time distance between visits. However, it has an advantage of making the interpretation of the aggregated results more clearly.

In this thesis, it was assumed that time effects that underlie the shape of the growth curve within their first five living years in PEACHES study population are "true" ones and need not be modified over time. The necessity for refitting the time model was not addressed. However, in the literature, the tendency of favoring small model updates over complete model revision was observed (Davis et al., 2019a, Houwelingen and Thorogood, 1995, Janssen et al., 2008, Steyerberg et al., 2004,

Vergouwe et al., 2017). Therefore, I believe the assumption made is acceptable and helps gain focus on the main objective of this thesis.

In three extended GLMM models, the following exposure variables were additionally considered: mother's pre-conceptional obesity status, child's largeness for gestational age, child's BMI Z-score at U1, and child's risk of overweight at U4. The choice of these variables was based on informal discussion with the study team. Including these variables mainly aims to introduce complexity and interaction into the model, rather than investigate the underlying medical mechanism of developing risk of overweight in children. Whether other exposure variables should be included is not within the scope of this thesis and needs be examined separately.

Well-known challenges of dynamic approaches and new insights

Published works identified from the literature review all agree on advantages of dynamic prediction models to improve future prediction. Dynamic model update offers possibilities to overcome initial sampling bias (Finkelman et al., 2016, Janssen et al., 2008) and improving model calibration over time (Finkelman et al., 2016), to tackle the issue of calibration drift and was shown to be a good alternative of static approaches (Siregar et al., 2016) such as periodic recalibration (Hickey, 2013). While extensive model revision suffers from overfitting, thus harms prediction performance (Steyerberg et al., 2004), dynamic model update provides small corrections over time that improve predictive performance but keeps the variance-bias trade-off in good balance (Steyerberg et al., 2004, Vergouwe et al., 2017).

Lenert et al. addressed the following issue as nature of prediction models, that the more effective a prediction model is, the faster it became useless (Lenert et al., 2019). Especially in interventional settings, where patients receive treatment over time, and if this treatment improves patients' outcome, it will naturally result in degrading of the prognostic model. The authors suggested to incorporate treatment in the model and also proposed that model updating is the most effective way to tackle this issue (Lenert et al., 2019). Results of the comparison between Bayesian models versus GLMER4 in this thesis agree with this proposal in a way that they showed the potential ability of dynamic approaches to adapt the models according to individuals' longitudinal outcome alone without incorporating other possible confounders or influencing factors. In a setting of complex intervention, where treatment effect over time cannot be modeled easily, dynamic approaches might be able to offer good solution to improve prediction performance.

This thesis gave new insights about factors that could influence the usefulness of dynamic approaches in a study setting with repeated measurements. The focused literature review showed that current literature has not dealt with the same question yet. In another context, a (simulated) clustered data setting, results from the study by Finkelman et al. could show that dynamic approaches are superior to static approaches with regard to prediction accuracy and more robust in situation of misspecified relationship between outcome and cluster size (Finkelman et al., 2016). The authors suggested that best results can be obtained when high-quality data is used (Finkelman et al., 2016). Another conclusion of this study indicated the improved prediction accuracy by using dynamic models with increasing values of variance of random intercepts (Finkelman et al., 2016), which agrees with the findings of this thesis. The authors stated that using the random intercept could account for the inter-cluster variability and therefore offered better prediction accuracy than static models and the dynamic linear models.

Dynamic approaches might need time to be improved and adapted to achieve acceptable grade of overperformance compared with traditional approaches. For this reason, they are probably

rather appropriate for long-term settings like chronic diseases or cohorts with long follow-up period. It needs also be mentioned that the use of dynamic approaches is related to data collection over multiple follow-up visits, which can increase effort and cost of conducting studies.

Generally, it is challenging to implement dynamic models as handy and portable tools for health care professionals and to make it easy for health care professionals as well as patients to comprehend and communicate these models with one another. However, I agree with Hickey et al. that with the current trend of digitalization in health care, the fast development of information technologies, as well as health care professionals' and patients' increasing contacts with digital tools in their work and daily routine, this challenge can be tackled in a near future (Hickey et al., 2013, Jenkins et al., 2021).

There is also a trade-off between complexity and prediction precision that needs be considered. It is therefore important to understand what factors influence the usefulness of the dynamic approaches, in order to make a decision for or against the use of such approaches as an alternative to traditional prediction approaches.

Due to these challenges, there is a necessity to examine the usefulness of dynamic approaches in the preparing stage of research. With simulation studies, different scenarios of unknown parameters can be considered and simulated. The usefulness of dynamic approaches and its influencing factors can then be investigated, in order to make a judgement whether it is worth applying a dynamic instead of traditional approach. Cost-utility analyses can also be performed to evaluate the use of dynamic approaches. Such analyses can, for instance, aim to answer the question how much monetary unit it would cost additionally to lower a prediction error by a certain percentage, which is clinically relevant for the settings. In prevention studies this objective could certainly be interesting.

In a medical point of view, this research work also pointed out potential breakpoints in children's growth curve from birth until five years of age that might be worth being further investigated and validated in future research.

Limitations and perspective for future research

The thesis did not aim to examine the medical mechanism that underlies the growth curve of PEACHES children over five years of observation but rather compare different Bayesian approaches that describe these curves with traditional regression models. For this reason, the results of this research work cannot provide any inference on the epidemiology of children's overweight. Exposure variables in extended models were chosen in an informal way. Therefore, interpretation of these models' output needs be done with care.

In the foreseeable future, analyses can be performed to gain a deeper look into the medical mechanisms that underlie child's development of risk for overweight by incorporating other mother's child's markers. Maternal glycated hemoglobin at delivery has been shown to associate with child's large-gestational-age status (Ensenauer et al., 2015). After a model is trained and tested, validation study can be done with the validation data set that was set aside in this research work but not yet used. The ultimate goal would be to provide health care professionals a prediction tool, which offers good prediction performance and straight forward to communicate.

Future research can focus on the question whether adding more complexity (covariates) into a dynamic model would bring essential improvement in prediction performance. Time model for the shape of the growth curve can be updated over time. It will be then possible that at each prediction visit, new time effects will be added into the model. Model updating will be not limited with updating

random intercepts or fixed effects but also updating its structure. The best model can be trained selected with training dataset, then tested with cross-validation and compared with other approaches. It can also be investigated in which settings more complex dynamic approaches are needed and, where using the data itself for model updating would be sufficient.

In this thesis, models updated after a single individual comes into the study and after each well-child visit of the same child. The benefit of the dynamic model might be more visible if updates done after more than one individual comes into the study. For instance, Davis et al. suggested model retraining if the number of individuals in the new data set is great enough but recommended small model corrections if the new data set is small (Davis et al., 2019a). The effect of sample size of new data set on the usefulness of dynamic approaches in similar study settings as PEACHES study needs be examined in future research.

This thesis aims to compare different approaches with regard to prediction performance. To do so, different numeric as well as graphical measures were applied. Score-based measures like (scaled) Brier score, Brier skill score or calibration intercepts or slopes give an insight into the quantitative difference of prediction performance of the different models, while graphical measures like calibration plots offer an overall visual comparison of these differences over all observation timepoints. The relative prediction error offers a better insight when comparing two specific models with each other. This measure also gives the approximate timepoint, where a specific model shows its overperformance compared to the other. Each of this applied measure offers insights from a different angle. In this thesis, no investigation was made to determine the superior measure of all.

Usefulness of dynamic approaches can be examined in a broader context, such as biomarker studies. In such a context, not only time and patient's outcome are considered in the model, but also time-varying biomarkers and/or treatment schemes.

Diagnosis and parameter tuning for Bayesian models as well as the handling of missing data in PEACHES study were not within the scope of this research work and is worth being studied in future research. Cohort and follow-up studies are probably study settings, where dynamic approaches can be implemented and made use. Missing data due to loss-to-follow-up in such settings is a well-known issue. It is also worth to examine to what extent which mechanisms of missing data influences the usefulness of dynamic approaches.

All in all, dynamic prediction methods, despite the well-known challenges they bring with them, are shown to have the potential to offer advantages over traditional prediction methods. Some of the challenges can be tackled with the development of information technologies. However, it is necessary to carefully evaluate the usefulness of dynamic approaches, considering feasible and established alternative approaches. The evaluation also needs be done closely within the context of the study assumptions and settings.

References

ALSOP-SHIELDS, L. & ALEXANDER, H. 1997. A study of errors that can occur when weighing infants. *J Adv Nurs,* 25**,** 587-94.

ANDRINOPOULOU, E.-R., RIZOPOULOS, D., TAKKENBERG, J. J. & LESAFFRE, E. 2017. Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statistical methods in medical research,* 26**,** 1787-1801.

BUNDESZENTRALE FÜR GESUNDHEITLICHE AUFKLÄRUNG. 2020. *U1 bis U9 – zehn Chancen für Ihr Kind* [Online]. Available: https://www.kindergesundheit-info.de/themen/entwicklung/frueherkennung-u1-u9-und-j1/untersuchungen-u1-bis-u9/ [Accessed].

DAVIS, S. E., GREEVY JR, R. A., FONNESBECK, C., LASKO, T. A., WALSH, C. G. & MATHENY, M. E. 2019a. A nonparametric updating method to correct clinical prediction model drift. *Journal of the American Medical Informatics Association,* 26**,** 1448-1457.

DAVIS, S. E., GREEVY, R. A., LASKO, T. A., WALSH, C. G. & MATHENY, M. E. Comparison of Prediction Model Performance Updating Protocols: Using a Data-Driven Testing Procedure to Guide Updating. AMIA Annual Symposium Proceedings, 2019b. American Medical Informatics Association, 1002.

DE ONIS M, L. T. 2010. Defining obesity risk status in the general childhood population: which cut-offs should we use? *In:* WORLD HEALTH ORGANIZATION (ed.). International journal of pediatric obesity.

ENSENAUER, R., BRANDLHUBER, L., BURGMANN, M., SOBOTZKI, C., ZWAFINK, C., ANZILL, S., HOLDT, L., TEUPSER, D., HASBARGEN, U., NETZ, H., ROSCHER, A. A. & VON KRIES, R. 2015. Obese Nondiabetic Pregnancies and High Maternal Glycated Hemoglobin at Delivery as an Indicator of Offspring and Maternal Postpartum Risks: The Prospective PEACHES Mother-Child Cohort. *Clin Chem,* 61**,** 1381-90.

FINKELMAN, B. S., FRENCH, B. & KIMMEL, S. E. 2016. The prediction accuracy of dynamic mixed-effects models in clustered data. *BioData Min,* 9.

GENDERS, T. S. S. 2011. A clinical prediction rule for the diagnosis of coronary artery disease: validation, updating and extension. *Eur Heart J,* 32.

GLEN, S. 2016. *"Brier Score: Definition, Examples" From StatisticsHowTo.com: Elementary Statistics for the rest of us!* [Online]. Available: https://www.statisticshowto.com/brier-score/ [Accessed].

GOMES, D., VON KRIES, R., DELIUS, M., MANSMANN, U., NAST, M., STUBERT, M., LANGHAMMER, L., HAAS, N. A., NETZ, H., OBERMEIER, V., KUHLE, S., HOLDT, L. M., TEUPSER, D., HASBARGEN, U., ROSCHER, A. A. & ENSENAUER, R. 2018. Late-pregnancy dysglycemia in obese pregnancies after negative testing for gestational diabetes and risk of future childhood overweight: An interim analysis from a longitudinal mother-child cohort study - Supporting information - Study protocol. *PLoS Med,* 15**,** e1002681.

HAFKAMP-DE GROEN, E. 2013. Predicting asthma in preschool children with asthma-like symptoms: validating and updating the PIAMA risk score. *J Allergy Clin Immunol,* 132.

HEYARD, R., TIMSIT, J. F., ESSAIED, W., HELD, L. & CONSORTIUM, C.-M. 2019. Dynamic clinical prediction models for discrete time-to-event data with competing risks-A case study on the OUTCOMEREA database. *Biom J,* 61**,** 514-534.

HICKEY, G. L. 2013. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. *Eur J Cardiothoracic Surg,* 43.

HICKEY, G. L., GRANT, S. W., CAIADO, C., KENDALL, S., DUNNING, J., POULLIS, M., BUCHAN, I. & BRIDGEWATER, B. 2013. Dynamic prediction modeling approaches for cardiac surgery. *Circulation: Cardiovascular Quality and Outcomes,* 6**,** 649-658.

HIPPISLEY-COX, J., COUPLAND, C., ROBSON, J. & BRINDLE, P. 2011. Derivation validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. *Bmj,* 342.

HOUWELINGEN, H. C. & THOROGOOD, J. 1995. Construction, validation and updating of a prognostic model for kidney graft survival. *Stat Med,* 14.

JANSSEN, K. J. M., MOONS, K. G. M., KALKMAN, C. J., GROBBEE, D. E. & VERGOUWE, Y. 2008. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol,* 61.

JENKINS, D. A., MARTIN, G. P., SPERRIN, M., RILEY, R. D., DEBRAY, T. P., COLLINS, G. S. & PEEK, N. 2021. Continual updating and monitoring of clinical prediction models: time for dynamic prediction systems? *Diagnostic and Prognostic Research,* 5**,** 1-7.

JENKINS, D. A., SPERRIN, M., MARTIN, G. P. & PEEK, N. 2018. Dynamic models to predict health outcomes: current status and methodological challenges. *Diagnostic and prognostic research,* 2**,** 1-9.

KEOGH, R. H., SEAMAN, S. R., BARRETT, J. K., TAYLOR-ROBINSON, D. & SZCZESNIAK, R. 2019. Dynamic Prediction of Survival in Cystic Fibrosis: A Landmarking Analysis Using UK Patient Registry Data. *Epidemiology,* 30**,** 29-37.

KUZNETSOVA, A., BROCKHOFF, P. B. & CHRISTENSEN, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *2017,* 82**,** 26.

LENERT, M. C., MATHENY, M. E. & WALSH, C. G. 2019. Prognostic models will be victims of their own success, unless…. *Journal of the American Medical Informatics Association,* 26**,** 1645-1650.

LI, L., LUO, S., HU, B. & GREENE, T. 2017. Dynamic Prediction of Renal Failure Using Longitudinal Biomarkers in a Cohort Study of Chronic Kidney Disease. *Stat Biosci,* 9**,** 357-378.

MRC BIOSTATISTICS UNIT 2003. WinBUGS User Manual. 1.4 ed.

NAKAGAWA, S., JOHNSON, P. C. D. & SCHIELZETH, H. 2017. The coefficient of determination R(2) and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J R Soc Interface,* 14.

PLATE, J. D., VAN DE LEUR, R. R., LEENEN, L. P., HIETBRINK, F., PEELEN, L. M. & EIJKEMANS, M. 2019. Incorporating repeated measurements into prediction models in the critical care setting: a framework, systematic review and meta-analysis. *BMC medical research methodology,* 19**,** 1-11.

POSCH, F., RIEDL, J., REITTER, E. M., CROWTHER, M. J., GRILZ, E., QUEHENBERGER, P., JILMA, B., PABINGER, I. & AY, C. 2020. Dynamic assessment of venous thromboembolism risk in patients with cancer by longitudinal D-Dimer analysis: A prospective study. *J Thromb Haemost,* 18**,** 1348-1356.

RASMUSSEN, K. M., CATALANO, P. M. & YAKTINE, A. L. 2009. New guidelines for weight gain during pregnancy: what obstetrician/gynecologists should know. *Curr Opin Obstet Gynecol,* 21**,** 521-6.

RIZOPOULOS, D., MOLENBERGHS, G. & LESAFFRE, E. 2017. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biom J,* 59**,** 1261-1276.

SIREGAR, S., NIEBOER, D., VERGOUWE, Y., VERSTEEGH, M. I., NOYEZ, L., VONK, A. B., STEYERBERG, E. W. & TAKKENBERG, J. J. 2016. Improved prediction by dynamic modeling: an exploratory study in the Adult Cardiac Surgery database of the Netherlands Association for Cardio-Thoracic Surgery. *Circulation: Cardiovascular Quality and Outcomes,* 9**,** 171-181.

STEYERBERG, E. W. 2009. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*, Springer.

STEYERBERG, E. W., BORSBOOM, G. J. J. M., HOUWELINGEN, H. C., EIJKEMANS, M. J. C. & HABBEMA, J. D. F. 2004. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med,* 23.

STEYERBERG, E. W., VICKERS, A. J., COOK, N. R., GERDS, T., GONEN, M., OBUCHOWSKI, N., PENCINA, M. J. & KATTAN, M. W. 2010. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology,* 21**,** 128-38.

SU, T.-L., JAKI, T., HICKEY, G. L., BUCHAN, I. & SPERRIN, M. 2018. A review of statistical updating methods for clinical prediction models. *Statistical methods in medical research,* 27**,** 185-197.

SURESH, K., TAYLOR, J. M. G., SPRATT, D. E., DAIGNAULT, S. & TSODIKOV, A. 2017. Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model. *Biom J,* 59**,** 1277-1300.

VERGOUWE, Y., NIEBOER, D., OOSTENBRINK, R., DEBRAY, T. P., MURRAY, G. D., KATTAN, M. W., KOFFIJBERG, H., MOONS, K. G. & STEYERBERG, E. W. 2017. A closed testing procedure to select an appropriate method for updating prediction models. *Statistics in medicine,* 36**,** 4529-4539.

VOIGT, M., ROCHOW, N., SCHNEIDER, K. T., HAGENAH, H. P., SCHOLZ, R., HESSE, V., WITTWER-BACKOFEN, U., STRAUBE, S. & OLBERTZ, D. 2014. [New percentile values for the anthropometric dimensions of singleton neonates: analysis of perinatal survey data of 2007-2011 from all 16 states of Germany]. *Z Geburtshilfe Neonatol,* 218**,** 210-7.

WIKIPEDIA CONTRIBUTORS. *Brier score* [Online]. Wikipedia, The Free Encyclopedia. Available: https://en.wikipedia.org/w/index.php?title=Brier_score&oldid=1026611845 [Accessed 8 June 2021 15:27 UTC 2021].

WORLD CLIMATE RESEARCH PROGRAMME. 2009. *Forecast Verification - Issues, Methods and FAQ* [Online]. Available: https://www.cawcr.gov.au/projects/verification/verif_web_page.html#BSS [Accessed 8 June 2021].

WORLD HEALTH ORGANIZATION 2006. WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development.

WUJCIAK-JENS, K. G. A. J. 2020. simstudy: Illuminating research methods through data generation. *Journal of Open Source Software,* 5**,** 2763.

Appendix A: Technical information

```
R version 4.0.4 (2021-02-15)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Debian GNU/Linux bullseye/sid

Matrix products: default
BLAS:   /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.9.0
LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.9.0

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
 [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
LC_PAPER=en_US.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] latex2exp_0.5.0      gridExtra_2.3        ggrepel_0.9.1
foreach_1.5.1        forcats_0.5.1
 [6] stringr_1.4.0        purrr_0.3.4          readr_1.4.0
tidyr_1.1.2          tibble_3.0.6
[11] tidyverse_1.3.0      mcmcplots_0.4.3      coda_0.19-4
rstudioapi_0.13      R2OpenBUGS_3.2-3.2.1
[16] simstudy_0.2.1       DT_0.17              shiny_1.5.0
ggpubr_0.4.0         dplyr_1.0.4
[21] plyr_1.8.6           ggplot2_3.3.3        openxlsx_4.2.3
broom.mixed_0.2.6    lmerTest_3.1-3
[26] lme4_1.1-26          Matrix_1.3-2

loaded via a namespace (and not attached):
 [1] minqa_1.2.4         colorspace_2.0-0    ggsignif_0.6.0
ellipsis_0.3.1      rio_0.5.16          rprojroot_2.0.2
 [7] fs_1.5.0            farver_2.0.3        fansi_0.4.2
lubridate_1.7.9.2   xml2_1.3.2          codetools_0.2-18
[13] splines_4.0.4       pkgload_1.1.0       jsonlite_1.7.2
nloptr_1.2.2.2      broom_0.7.4         dbplyr_2.1.0
```

```
[19] rjags_4-10          sfsmisc_1.1-8       compiler_4.0.4
httr_1.4.2          backports_1.2.1     assertthat_0.2.1
[25] fastmap_1.1.0       cli_2.3.0           later_1.1.0.1
htmltools_0.5.1.1   tools_4.0.4         gtable_0.3.0
[31] glue_1.4.2          reshape2_1.4.4      Rcpp_1.0.6
carData_3.0-4       cellranger_1.1.0    vctrs_0.3.6
[37] nlme_3.1-152        iterators_1.0.13    testthat_3.0.1
rvest_0.3.6         mime_0.9            lifecycle_0.2.0
[43] statmod_1.4.35      rstatix_0.6.0       MASS_7.3-53.1
scales_1.1.1        hms_1.0.0           promises_1.1.1
[49] parallel_4.0.4      TMB_1.7.18          curl_4.3
stringi_1.5.3       desc_1.2.0          boot_1.3-27
[55] zip_2.1.1           rlang_0.4.10        pkgconfig_2.0.3
lattice_0.20-41     htmlwidgets_1.5.3   labeling_0.4.2
[61] tidyselect_1.1.0    magrittr_2.0.1      R6_2.5.0
generics_0.1.0      DBI_1.1.1           pillar_1.4.7
[67] haven_2.3.1         foreign_0.8-81      withr_2.4.1
abind_1.4-5         modelr_0.1.8        crayon_1.4.0
[73] car_3.0-10          denstrip_1.5.4      utf8_1.1.4
grid_4.0.4          readxl_1.3.1        data.table_1.13.6
[79] reprex_1.0.0        digest_0.6.27       xtable_1.8-4
httpuv_1.5.5        numDeriv_2016.8-1.1 munsell_0.5.0
```

Appendix B: Supplementary results

Appendix B1. Selected time effects that associate with BMI-Z score

The following model output describes the chosen model after backward selection for PEACHES data set.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: zbmi ~ months_from_u1 + months_from_u2 + months_from_u3 +
months_from_u4 +      months_from_u6 + months_from_u7 + (1 | PseudoID)
   Data: antab

REML criterion at convergence: 36980.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-7.2371 -0.5793  0.0039  0.5854  7.6531

Random effects:
 Groups   Name        Variance Std.Dev.
 PseudoID (Intercept) 0.4364   0.6606
 Residual             0.6472   0.8045
Number of obs: 14112, groups:  PseudoID, 1557

Fixed effects:
                 Estimate Std. Error         df t value Pr(>|t|)
(Intercept)    -4.329e-01  2.479e-02  4.355e+03 -17.461  < 2e-16 ***
months_from_u1 -2.068e+00  1.974e-01  1.255e+04 -10.473  < 2e-16 ***
months_from_u2  2.458e+00  2.080e-01  1.243e+04  11.817  < 2e-16 ***
months_from_u3 -4.088e-01  3.133e-02  1.409e+04 -13.045  < 2e-16 ***
months_from_u4  7.475e-02  1.231e-02  1.370e+04   6.072 1.29e-09 ***
months_from_u6 -2.618e-02  4.735e-03  1.294e+04  -5.528 3.30e-08 ***
months_from_u7 -3.486e-02  2.615e-03  1.273e+04 -13.331  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) mnt__1 mnt__2 mnt__3 mnt__4 mnt__6
mnths_frm_1 -0.464
mnths_frm_2  0.426 -0.994
mnths_frm_3  0.095  0.279 -0.379
mnths_frm_4  0.010  0.045  0.005 -0.645
mnths_frm_6 -0.002  0.006 -0.001  0.073 -0.515
mnths_frm_7  0.000 -0.004  0.003 -0.008  0.151 -0.734
```

Interpretation of the model output:

At U1, on average, a child in the PEACHES cohort had a BMI-Z score of -0.433. Every day after U1, the BMI-Z score of this average child reduces by 2.068 * 30.4375 = 0.068 unit. Every month after U2, his/her BMI-Z score increases by -2.068 + 2.458 = 0.39 unit. Every month after U3, his/her BMI-Z score decreases by -2.068 + 2.458 – 0.4088= 0.019 unit. Every month after U4, his/her BMI-Z score increases by -2.068 + 2.458 - 0.4088 + 7.475e-02 = 0.056 unit. Every month after U6, his/her BMI-Z score increases by -2.068 + 2.458 - 0.4088 + 7.475e-02 - 2.618e-02 = 0.030 unit. Every month after U7, his/her BMI-Z score decreases by -2.068 + 2.458 - 0.4088 + 7.475e-02 - 2.618e-02 -3.486e-02 = 0.005 unit.

Taking the simplified time frame of the visits as 0 days (U1), 3 days (U2), 30 days (U3), 3 months (U4), 6 months (U5), 12 months (U6), 24 months (U7), 36 months (U7a), 48 months (U8), and 60 months (U9), the estimated BMI-Zscore of an average child would be: -0.43 at U1, -0.433 -0.068*3 days = -0.637 at U2, -0.637 + 0.39/30.4375*(30-3) days = -0.291 at U3, -0.291 - 0.019*2 months = -0.329 at U4, -0.329 + 0.056*3 months = -0.161 at U5, -0.329 + 0.056*9 months = 0.175 at U6, 0.175 + 0.030* 12 months = 0.535 at U7, 0.535 – 0.005*12 months = 0.475 at U7a, 0.535 – 0.005*24 months = 0.415 at U8, and 0.535 – 0.005*36 months = 0.355 at U9. This interpretation applies assuming that the random intercept of this child is 0.

Appendix B2. Time effects used in the simulation design

The following model output was obtained by fitting a model with the training dataset using the selected time effects above (see Appendix B1):

```
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: risk ~ months_from_u1 + months_from_u2 + months_from_u3 +
months_from_u4 +
    months_from_u6 + months_from_u7 + (1 | PseudoID)
   Data: antab.train
Control: glmerControl(tolPwrss = 0.1)

     AIC      BIC   logLik deviance df.resid
   590.4    628.9   -287.2    574.4      901

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8455 -0.2658 -0.1771 -0.0722  6.0236

Random effects:
 Groups   Name        Variance Std.Dev.
 PseudoID (Intercept) 3.551    1.884
Number of obs: 909, groups:  PseudoID, 100

Fixed effects:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)    -4.65186    0.51158  -9.093   <2e-16 ***
months_from_u1  0.10092    2.94906   0.034   0.9727
months_from_u2  0.56928    3.06273   0.186   0.8525
months_from_u3 -0.37908    0.54987  -0.689   0.4906
months_from_u4 -0.19887    0.21051  -0.945   0.3448
months_from_u6 -0.03126    0.07224  -0.433   0.6652
months_from_u7 -0.07443    0.03589  -2.074   0.0381 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) mnt__1 mnt__2 mnt__3 mnt__4 mnt__6
mnths_frm_1 -0.385
mnths_frm_2  0.306 -0.990
mnths_frm_3  0.336  0.120 -0.255
mnths_frm_4  0.046  0.069 -0.016 -0.604
mnths_frm_6  0.047  0.019 -0.016  0.096 -0.515
mnths_frm_7  0.018 -0.003  0.007 -0.039  0.161 -0.708
optimizer (Nelder_Mead) convergence code: 0 (OK)
Model is nearly unidentifiable: large eigenvalue ratio
 - Rescale variables?
```

Appendix B3. Analyses with sample size of 100

Appendix B3 (saved as "`supplementary_material/AppendixB3.pdf`") contains results of the supplementary analyses for PEACHES and simulation main study with sample size of 100. Supplementary material can be accessed by using the link provided in Appendix D.

Appendix C: Results of literature search

The following publications were identified as relevant and entered the full-text analysis. The publications are listed in alphabetical order, regarding first author's name.

ALTMAN, D. G. & ROYSTON, P. 2000. What do we mean by validating a prognistic model? *Stat Med,* 19.

ANDRINOPOULOU, E.-R., RIZOPOULOS, D., TAKKENBERG, J. J. & LESAFFRE, E. 2017. Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statistical methods in medical research,* 26**,** 1787-1801.

BOOTH, S. R., R. D.; ENSOR, J.; LAMBERT, P. C. & RUTHERFORD, M. J. 2020. Temporal recalibration for improving prognostic model development and risk predictions in settings where survival is improving over time. . *Int. J. Epidemiol. ,* 1.

BULL, L. M., LUNT, M., MARTIN, G. P., HYRICH, K. & SERGEANT, J. C. 2020. Harnessing repeated measurements of predictor variables for clinical risk prediction: a review of existing methods. *Diagnostic and prognostic research,* 4**,** 1-16.

DAMEN JAAG, E. A. 2016. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*.

DAVIS, S. E., GREEVY JR, R. A., FONNESBECK, C., LASKO, T. A., WALSH, C. G. & MATHENY, M. E. 2019. A nonparametric updating method to correct clinical prediction model drift. *Journal of the American Medical Informatics Association,* 26**,** 1448-1457.

DAVIS, S. E., GREEVY JR, R. A., LASKO, T. A., WALSH, C. G. & MATHENY, M. E. 2020. Detection of calibration drift in clinical prediction models to inform model updating. *Journal of Biomedical Informatics,* 112**,** 103611.

DAVIS, S. E., GREEVY, R. A., LASKO, T. A., WALSH, C. G. & MATHENY, M. E. Comparison of Prediction Model Performance Updating Protocols: Using a Data-Driven Testing Procedure to Guide Updating. AMIA Annual Symposium Proceedings, 2019. American Medical Informatics Association, 1002.

DEBRAY, T. P. A., KOFFIJBERG, H., VERGOUWE, Y., MOONS, K. G. M. & STEYERBERG, E. W. 2012. Aggregating published prediction models with individual participant data: a comparison of different approaches. *Stat. Med.,* 31.

FAN, J. & ZHANG, W. 2008. Statistical methods with varying coefficient models. *Stat Interface,* 1.

FINKELMAN, B. S., FRENCH, B. & KIMMEL, S. E. 2016. The prediction accuracy of dynamic mixed-effects models in clustered data. *BioData Min,* 9.

GENDERS, T. S. S. 2011. A clinical prediction rule for the diagnosis of coronary artery disease: validation, updating and extension. *Eur Heart J,* 32.

GOLDSTEIN, B. A., POMANN, G. M., WINKELMAYER, W. C. & PENCINA, M. J. 2017. A comparison of risk prediction methods using repeated observations: an application to electronic health records for hemodialysis. *Statistics in medicine,* 36**,** 2750-2763.

HAFKAMP-DE GROEN, E. 2013. Predicting asthma in preschool children with asthma-like symptoms: validating and updating the PIAMA risk score. *J Allergy Clin Immunol,* 132.

HALABI, S. 2014. Updated prognostic model for predicting overall survival in first-line chemotherapy for patients with metastatic castration-resistant prostate cancer. *J. Clin. Oncol.,* 32.

HICKEY, G. L. 2013. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. *Eur J Cardiothoracic Surg,* 43.

HICKEY, G. L., GRANT, S. W., CAIADO, C., KENDALL, S., DUNNING, J., POULLIS, M., BUCHAN, I. & BRIDGEWATER, B. 2013. Dynamic prediction modeling approaches for cardiac surgery. *Circulation: Cardiovascular Quality and Outcomes,* 6**,** 649-658.

HIPPISLEY-COX, J., COUPLAND, C., ROBSON, J. & BRINDLE, P. 2011. Derivation validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. *Bmj,* 342.

HOOVER, D. R., RICE, J. A., WU, C. O. & YANG, L. P. 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika,* 85.

HOUWELINGEN, H. C. & THOROGOOD, J. 1995. Construction, validation and updating of a prognostic model for kidney graft survival. *Stat Med,* 14.

JANSSEN, K. J. M., MOONS, K. G. M., KALKMAN, C. J., GROBBEE, D. E. & VERGOUWE, Y. 2008. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol,* 61.

JENKINS, D. A., MARTIN, G. P., SPERRIN, M., RILEY, R. D., DEBRAY, T. P., COLLINS, G. S. & PEEK, N. 2021. Continual updating and monitoring of clinical prediction models: time for dynamic prediction systems? *Diagnostic and Prognostic Research,* 5**,** 1-7.

JENKINS, D. A., SPERRIN, M., MARTIN, G. P. & PEEK, N. 2018. Dynamic models to predict health outcomes: current status and methodological challenges. *Diagnostic and prognostic research,* 2**,** 1-9.

LENERT, M. C., MATHENY, M. E. & WALSH, C. G. 2019. Prognostic models will be victims of their own success, unless…. *Journal of the American Medical Informatics Association,* 26**,** 1645-1650.

LUIJKEN, K., WYNANTS, L., VAN SMEDEN, M., VAN CALSTER, B., STEYERBERG, E. W., GROENWOLD, R. H., TIMMERMAN, D., BOURNE, T. & UKAEGBU, C. 2020. Changing predictor measurement procedures affected the performance of prediction models in clinical examples. *Journal of clinical epidemiology,* 119**,** 7-18.

MARTIN, G. P., MAMAS, M. A., PEEK, N., BUCHAN, I. & SPERRIN, M. 2018. A multiple-model generalisation of updating clinical prediction models. *Stat. Med.,* 37.

MCCORMICK, T. H., RAFTERY, A. & MADIGAN, D. 2018. *dma: dynamic model averaging*.

MCCORMICK, T. H., RAFTERY, A. E., MADIGAN, D. & BURD, R. S. 2012. Dynamic logistic regression and dynamic model averaging for binary classification. *Biometrics,* 68.

MEMORIAL SLOAN KETTERING CENTER 2020. Dynamic Prostate Cancer Nomogram: Coefficients.

MOONS, K. G. M. 2012. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart,* 98.

PAJOUHESHNIA, R., VAN SMEDEN, M., PEELEN, L. & GROENWOLD, R. 2019. How variation in predictor measurement affects the discriminative ability and transportability of a prediction model. *Journal of clinical epidemiology,* 105**,** 136-141.

PATE, A., EMSLEY, R., ASHCROFT, D. M., BROWN, B. & VAN STAA, T. 2019. The uncertainty with using risk prediction models for individual decision making: an exemplar cohort study examining the prediction of cardiovascular disease in English primary care. *BMC medicine,* 17**,** 1-16.

PLATE, J. D., VAN DE LEUR, R. R., LEENEN, L. P., HIETBRINK, F., PEELEN, L. M. & EIJKEMANS, M. 2019. Incorporating repeated measurements into prediction models in the critical care setting: a framework, systematic review and meta-analysis. *BMC medical research methodology,* 19**,** 1-11.

POSCH, F., RIEDL, J., REITTER, E. M., CROWTHER, M. J., GRILZ, E., QUEHENBERGER, P., JILMA, B., PABINGER, I. & AY, C. 2020. Dynamic assessment of venous thromboembolism risk in patients with cancer by longitudinal D-Dimer analysis: A prospective study. *J Thromb Haemost,* 18**,** 1348-1356.

RAFTERY, A. E. & ETTLER, P. 2010. Online prediction under model uncertainty via dynamic model averaging : application to a cold rolling mill. *Technometrics,* 52.

SIREGAR, S., NIEBOER, D., VERGOUWE, Y., VERSTEEGH, M. I., NOYEZ, L., VONK, A. B., STEYERBERG, E. W. & TAKKENBERG, J. J. 2016. Improved prediction by dynamic modeling: an exploratory study in the Adult Cardiac Surgery database of the Netherlands Association for Cardio-Thoracic Surgery. *Circulation: Cardiovascular Quality and Outcomes,* 9**,** 171-181.

STEYERBERG, E. W., BORSBOOM, G. J. J. M., HOUWELINGEN, H. C., EIJKEMANS, M. J. C. & HABBEMA, J. D. F. 2004. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med,* 23.

SU, T.-L., JAKI, T., HICKEY, G. L., BUCHAN, I. & SPERRIN, M. 2018. A review of statistical updating methods for clinical prediction models. *Statistical methods in medical research,* 27**,** 185-197.

TOLL, D. B., JANSSEN, K. J. M., VERGOUWE, Y. & MOONS, K. G. M. 2008. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol,* 61.

VAN ES, N. 2020. Dynamic prediction modeling for cancer-associated venous thromboembolism. *Journal of Thrombosis and Haemostasis,* 18**,** 1276.

VERGOUWE, Y., NIEBOER, D., OOSTENBRINK, R., DEBRAY, T. P., MURRAY, G. D., KATTAN, M. W., KOFFIJBERG, H., MOONS, K. G. & STEYERBERG, E. W. 2017. A closed testing procedure to select an appropriate method for updating prediction models. *Statistics in medicine,* 36**,** 4529-4539.

Appendix D: Reproducibility of the results

R codes to reproduce the results are provided under this link:
https://drive.google.com/drive/folders/1hNrE-X3UBT9JlyU6PyJrWkQUA6WxePVf?usp=sharing

Supplementary material as well as full texts of the references can also be found under the link above.

Acknowledgements

I would like to thank Prof. Ulrich Mansmann for his untiring support over the years. His broad knowledge in mathematics, statistics, and his work ethics at the highest standard have been inspriring and supporting me in various research projects that we have been working in together the last seven years and motivating me to finish this research thesis. Furthermore, I want to thank Prof. Anne-Laure Boulesteix and Prof. Eva Hoster for their helpful advices, suggestions, and their openness towards my ideas and questions.

I want to dedicate my sincere thanks to my colleagues Dr. Ursula Berger and Dr. Christine Adrion for their continuous encouragement and support. Thank you for bringing culture and authencity into my work life.

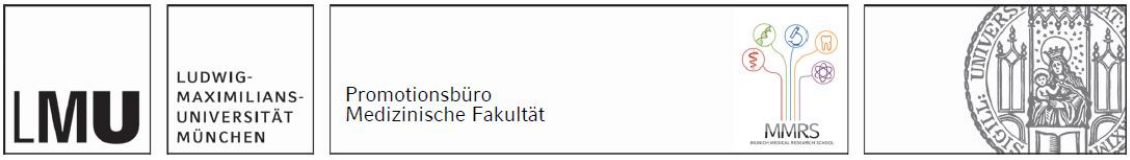
I want to thank Delphina Gomes and Prof. Ensenauer for providing me the PEACHES data and their cooperation in processing the data.

I am grateful for the instructions of Magda Radermacher, Monika Darchinger and Annette Hartmann from the PhD office throughout my PhD program. I want to thank Nikolaus von Bomhard for the technical support, Anja Friedrichs, Frau Pelagia Pajonk, and Shirley von Stuckrad for the organizational support. Without this seamless support, I would have not been able to focus on the actual research work.

My dearest thanks are dedicated to my parents, An, Manuel & Marie, and especially Beate, who have been there for me and sharing with me the ups and downs of this journey.

Finally, I want to give my special thanks to Prof. Jörg Hasford. I had the privilege to become his mentee in 2010. Over the years, he had been wholeheartedly teaching me about work ethics and life. He never lost faith and hope in me. This thesis, I dedicate wholly to him, with all my respect and thankfulness.

Affidavit



Affidavit

Le, Dung Lien
_____
Surname, first name

Gräfelfingerstr. 54
_____
Street

81375 München, Germany
_____
Zip code, town, country

I hereby declare, that the submitted thesis entitled:

The usefulness of dynamic approaches in predicting risk of overweight in children within the PEACHES cohort

…….

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

Munich, 01.04.2023 Lien Dung Le

_____ _____

place, date Signature doctoral candidate

Confirmation of congruency

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN | Promotionsbüro Medizinische Fakultät | MMRS

Confirmation of congruency between printed and electronic version of the doctoral thesis

Le, Lien Dung

Surname, first name

Gräfelfingerstr. 54

Street

81375 München, Germany

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

The usefulness of dynamic approaches in predicting risk of overweight in children within the PEACHES cohort

……

is congruent with the printed version both in content and format.

Munich, 01.04.2023

place, date

Lien Dung Le

Signature doctoral candidate

List of publications

GLIMM, A.-M., SPRENGER, L. I., HAUGEN, I. K., MANSMANN, U., HERMANN, S., HÄUPL, T., HOFF, P., BURMESTER, G.-R., BACKHAUS, M., LE, L. & OHRNDORF, S. 2019. Fluorescence optical imaging for treatment monitoring in patients with early and active rheumatoid arthritis in a 1-year follow-up period. *Arthritis Research & Therapy,* 21**,** 209.

LE, L. D., MANSMANN, U. R., JUNG, A., KIRCHNER, T., SCHÄFER, R., NEUREITER, D., HOLCH, J. W., HEINEMANN, V. & STINTZING, S. 2017. Is the primary tumor location (PTL) associated with differential gene expression profiles in patients with metastatic colorectal cancer (mCRC)? Analysis of the FIRE1-trial. *Journal of Clinical Oncology,* 35**,** 598-598.

NGUYEN, N., THALHAMMER, R., BEUTNER, K., SAAL, S., SERVATY, R., KLINGSHIRN, H., ICKS, A., FREYBERG, K., VOMHOF, M., MANSMANN, U., LE, L., MULLER, M. & MEYER, G. 2019. Effectiveness of a complex intervention to improve participation and activities in nursing home residents with joint contractures (JointConEval): study protocol of a multicentre cluster-randomised controlled trial [DRKS-ID:DRKS00015185]. *Trials,* 20**,** 305.

PHILIPP SEWERIN, A. M.-L., CHRISTOPH SCHLEICH, FLORIAN FICHTE, MARKUS EICHNER,, RUBEN SENGEWEIN, LIEN LE, HANS-JÖRG WITTSACK, MATTHIAS SCHNEIDER AND BENEDIKT OSTENDORF 2017. The Value of Dynamic Contrast-Enhanced MRI and Delayed Gadolinium Enhanced MRI of the Cartilage in Patients with Early Rheumatoid Arthritis: Leads Local Hyperperfusion to Cartilage Loss? *2017 ACR/ARHP Annual Meeting.*

RADEMACHER, J., TIETZ, L., LE, L., HARTL, A., RUDWALEIT, M., SIEPER, J., MANSMANN, U. & PODDUBNYY, D. 2018. FRI0154 Added value of biomarkers compared to routine clinical parameters for the prediction of radiographic spinal progression in axial spondyloarthritis. *Annals of the Rheumatic Diseases,* 77**,** 620-621.

RADEMACHER, J., TIETZ, L. M., LE, L., HARTL, A., HERMANN, K. A., SIEPER, J., MANSMANN, U., RUDWALEIT, M. & PODDUBNYY, D. 2019. Added value of biomarkers compared with clinical parameters for the prediction of radiographic spinal progression in axial spondyloarthritis. *Rheumatology (Oxford),* 58**,** 1556-1564.

RADEMACHER, J., TIETZ, L. M., LE, L., HARTL, A., HERMANN, K. A., SIEPER, J., MANSMANN, U., RUDWALEIT, M. & PODDUBNYY, D. 2019. Added value of biomarkers compared with clinical parameters for the prediction of radiographic spinal progression in axial spondyloarthritis. *Rheumatology (Oxford),* 58**,** 1556-1564.

SCHUNK, M., BERGER, U., LE, L., REHFUESS, E. A., SCHWARZKOPF, L., STREITWIESER, S., MÜLLER, T., HOFMANN, M., HOLLE, R., HUBER, R. M., MANSMANN, U. & BAUSEWEIN, C. 2020. Randomisiert-kontrollierte Studie zur Evaluation der Atemnot-Ambulanz München (BreathEase): Rekrutierung und Beschreibung der Studienteilnehmer [164]. *Zeitschrift für Palliativmedizin,* 21**,** P44.

SCHUNK, M., LE, L., SYUNYAEVA, Z., HABERLAND, B., TÄNZLER, S., MANSMANN, U., NASTASSJA, B., SCHWARZKOPF, L., SEIDL, H., STREITWIESER, S., HOFMANN, M., MÜLLER, T., WEIß, T., MORAWIETZ, P., REHFUESS, E. A., HUBER, R. M., BERGER, U. & BAUSEWEIN, C. 2020. Effectiveness of a breathlessness service for patients suffering from breathlessness in advanced disease: pragmatic fast-track randomized controlled trial. *European Respiratory Journal,* 56**,** 3820.

SCHUNK, M., BERGER, U., LE, L., REHFUESS, E., SCHWARZKOPF, L., STREITWIESER, S., MULLER, T., HOFMANN, M., HOLLE, R., HUBER, R. M., MANSMANN, U. & BAUSEWEIN, C. 2021. BreathEase: rationale, design and recruitment of a randomised trial and embedded mixed-methods study of a multiprofessional breathlessness service in early palliative care. *ERJ Open Res,* 7.

SCHUNK, M., LE, L., SYUNYAEVA, Z., HABERLAND, B., TANZLER, S., MANSMANN, U., SCHWARZKOPF, L., SEIDL, H., STREITWIESER, S., HOFMANN, M., MULLER, T., WEISS, T., MORAWIETZ, P., REHFUESS, E. A., HUBER, R. M., BERGER, U. & BAUSEWEIN, C. 2021. Effectiveness of a specialised breathlessness service for patients with advanced disease in Germany: a pragmatic fast track randomised controlled trial (BreathEase). *Eur Respir J*.

SCHUNK, M., LE, L., SYUNYAEVA, Z., HABERLAND, B., TÄNZLER, S., MANSMANN, U., SCHWARZKOPF, L., SEIDL, H., STREITWIESER, S., HOFMANN, M., MÜLLER, T., WEIß, T., MORAWIETZ, P., REHFUESS, E. A., HUBER, R. M., BERGER, U. & BAUSEWEIN, C. 2020. Behandlung in der Atemnot-Ambulanz führt zu besserem Umgang mit chronisch refraktärer Atemnot bei Patienten mit fortgeschrittenen Erkrankungen: Ergebnisse der randomisiert-kontrollierten Studie BreathEase [160]. *Zeitschrift für Palliativmedizin,* 21**,** V7: 15:00–15:15 Uhr.

SCHUNK, M., STREITWIESER, S., HABERLAND, B., EGLI, M., LE, L., HOFMANN, M., MÜLLER, T., BERGER, U., MANSMANN, U., REHFUESS, E., SEIDL, H., HOLLE, R., HUBER, R. M. & BAUSEWEIN, C. 2016.

Erfahrungen mit der Patientenrekrutierung und der Durchführung einer randomisierten kontrollierten komplexen Interventionsstudie bei palliativ erkrankten Patienten am Beispiel der BreathEase-Studie. *Zeitschrift für Palliativmedizin,* 17**,** P86.

SEIDL, H., SCHUNK, M., LE, L., SYUNYAEVA, Z., STREITWIESER, S., BERGER, U., MANSMANN, U., SZENTES, B. L., BAUSEWEIN, C. & SCHWARZKOPF, L. 2022. Cost-Effectiveness of a Specialized Breathlessness Service Versus Usual Care for Patients With Advanced Diseases. *Value Health*.

SEWERIN, P., LE, L., VORDENBAUMEN, S., SCHLEICH, C., SENGEWEIN, R., BRINKS, R., PONGRATZ, G., BLECK, E., LESCH, J., MANSMANN, U., SCHNEIDER, M. & OSTENDORF, B. 2018. Rheumatoid Arthritis Magnetic Resonance Imaging Score Predicts Therapy Response: Results of the German ArthroMark Cohort. *J Rheumatol,* 45**,** 753-759.

SEWERIN, P., LE, L., VORDENBAUMEN, S., SCHLEICH, C., SENGEWEIN, R., BRINKS, R., PONGRATZ, G., BLECK, E., LESCH, J., MANSMANN, U., SCHNEIDER, M. & OSTENDORF, B. 2018. Rheumatoid Arthritis Magnetic Resonance Imaging Score Predicts Therapy Response: Results of the German ArthroMark Cohort. *J Rheumatol,* 45**,** 753-759.

SHEN, Y. M., LE, L. D., WILSON, R. & MANSMANN, U. 2017. Graphical Presentation of Patient-Treatment Interaction Elucidated by Continuous Biomarkers. Current Practice and Scope for Improvement. *Methods Inf Med,* 56**,** 13-27.

ZHANG, M., SAAD, C., LE, L., HALFTER, K., BAUER, B., MANSMANN, U. R. & LI, J. 2018. Computational modeling of methionine cycle-based metabolism and DNA methylation and the implications for anticancer drug response prediction. *Oncotarget,* 9**,** 22546-22558.