Joint modeling versus discriminant analysis for dynamic prediction of a binary outcome based on longitudinal data: a simulation study

Rana Dandis¹, Joanna IntHout¹, Kit Roes¹, Steven Teerenstra¹

¹ Radboud university medical center, Research Institute for Health Sciences, Department for Health Evidence, Section Biostatistics, Geert Grooteplein Zuid 21, 6525 EZ Nijmegen, the Netherlands

Correspondence:

Rana Dandis, Radboud university medical center, Research Institute for Health Sciences, Department for Health Evidence, Section Biostatistics, Geert Grooteplein Zuid 21, 6525 EZ Nijmegen, the Netherlands

Email: Rana.Dandis@Radboudumc.nl

Abstract

Background: In literature, much emphasis has been placed on new statistical methods for dynamic risk prediction using longitudinal biomarker information. However, few studies have compared their performance for predicting a long-term binary outcome. In this paper we perform a simulation study to compare the dynamic predictive performance of two commonly used methods namely: longitudinal discriminant analysis (LoDA) and joint modeling (JM-Bin) for longitudinal and binary data.

Methods: Motivated by a real-world dataset, we simulate different scenarios in which we changed the event rate (i.e. the percentage of patients) and the subject-specific variability of the biomarker to assess their influence on the predictive performance of the two approaches. More specifically, we allow the variability to be different between the subjects based on their outcome, i.e., the between- and/or within- subjects variability is larger in patients compared to healthy subjects. Time-dependent predictive measures (mean squared error of prediction, area under ROC curve) are used to evaluate the dynamic predictive performance.

Results: Results show that LoDA produces more accurate predictions than JM-Bin in most of the simulation scenarios. In general, increasing the biomarker's between-subjects variability reduces the predictive accuracy of both approaches to the same extent. The increase in the number of events does not influence the prediction accuracy of both methods.

Conclusions: The predictive performance of LoDA is especially better than JM-Bin when the biomarker's (within- and/or between-subjects) variability is different between the outcome groups.

Keywords: Joint model, Longitudinal discriminant analysis, Dynamic prediction, longitudinal data, dichotomous outcome.

1 Background

In many medical applications, longitudinal biomarker data can serve as predictor for a future clinical outcome, for example the occurrence of a disease. In some applications the interest is to use at each moment in time all collected information of a patient to obtain (up to date) predictions of the risk of that disease, and to revise these predictions whenever new information is available. For example, the prostate-specific antigen (PSA) levels are used to provide dynamic predictions of the future development of prostate cancer (1), and the measurement of blood pressure at antenatal appointments is used to identify risk of pre-eclampsia in pregnant women.

In the past, only the most recent information (obtained at the most recent follow up visit) was considered in predicting patients' risk of a clinical outcome. All previously gathered information was not considered, which could be an inefficient use of data. For example, in cases where the change in patients' biomarker values over time is more informative in predicting their risk than simply the most recent value of the biomarker, or in situations where the variability of the biomarker values within the same subject can be predictive for the outcome of interest.

Recently, more flexible statistical methods have been developed for prediction using a biomarker's longitudinal history, and for subsequently (prediction-based) classifications. Some of these methods involve joint modeling strategies where estimation and prediction are based on specification of a joint density of the future binary outcome and the vector of repeated biomarker values. The estimation of the model's parameters is done by maximizing the joint likelihood function. Depending on the factorization of the resulting joint likelihood one can fit a joint model for the longitudinal biomarker and binary outcome (JM-Bin) as a special case of

a shared random effects models (SREM) or perform a longitudinal discriminant analysis (LoDA) for two groups as a special case of the pattern mixture models (PMM).

In a previously published paper we presented a tutorial for using different SREM approaches to obtain dynamic prediction for a binary outcome (2). The presented SREM approaches consist of two sub-models. The first sub-model is a linear mixed model fitted to the longitudinal biomarker data, while the second is a binary logistic regression for the clinical outcome where the predicted random effects from the first model are used as covariates. The two sub-models are either linked in separate steps through a two-stage approach or by using the joint modeling (JM-Bin) approach where simultaneous estimation of parameters for both the longitudinal biomarker and the binary outcome model is done. The joint modeling approach is a novel statistical tool that not only can be used to predict a binary outcome but also it can accommodate categorical, count, continuous and survival outcomes when using the suitable regression model. We recommended applying joint modeling approaches above two stage methods in applications where the biomarker data are measured with large error or/and when the variability in the biomarker data between the study subjects is large.

The LoDA framework models the longitudinal biomarker distribution using a mixed effects model separately for patients and healthy subjects in available, historic data. Then, for a new subject, it estimates the probability of the biomarker measurements given the disease status and obtains prediction using Bayes' theorem. Sajobi et al (2010) provided a literature review on discriminant procedures for univariate and multivariate repeated measures data, focusing on linear mixed-effects models (3). Komárek et al (2010) developed a discriminant analysis approach where they used a multivariate linear mixed model with a normal mixture in the random effects distribution (4). Later, Komárek et al (2014) developed an R package to apply the multivariate longitudinal discriminant analysis approach (5). Hughes et al (2016) applied

longitudinal discriminant analysis to provide dynamic classification of patients to a risk group based on multiple longitudinal markers of different types (6).

Both JM-Bin and LoDA can be used to provide updated predictions. However, there is a lack of studies that compare their predictive performance. Therefore, in this study we aim to compare the dynamic predictive performance of these two approaches and to explore the situations where one of them outperforms the other. We will focus on the situation where we have one longitudinal biomarker and one binary outcome. We will investigate the influence of changes in the event rate (i.e. the percentage of patients) and the subject- specific variability of the biomarker on the predictive performance of the two approaches. We will do that using various simulation scenarios informed by a real-world dataset.

The rest of this paper is organized as follows: section 2 briefly reviews the general settings of the JM-Bin and LoDA frameworks. In section 3 we describe the simulation scenarios. In section 4 we present the results of our simulation study. Section 5 ends this paper with discussion and recommendations.

2 Methods

2.1 Statistical frameworks

Suppose that we have longitudinal measurements of a biomarker and a binary outcome for N subjects. Let D_i , $i = 1, \dots, N$, denote the binary indicator of the i th subject who either belongs to the patient group ($D_i = 1$) or to the healthy subject group ($D_i = 0$). Assume that the outcome for subject i , D_i , is determined only after last follow up time, T_i . Let the first M subjects be patients and the last $(N - M)$ subjects healthy, and $y_i(t_{ij})$ represent the longitudinal biomarker measurement for subject i measured at the corresponding time point t_{ij} , $j = 1, \dots, T_i$. For simplicity of presentation, for both modeling approaches we assume a linear evolution of the biomarker over time.

2.1.1 Joint model for longitudinal data and binary outcome (JM-Bin)

The joint model consists of two submodels: a mixed effects model for the longitudinal biomarker over time, and a logistic regression model for the binary outcome. The submodel for the linear trajectories of the longitudinal biomarker can be written as follows,

$$y_i(t_{ij}) = \beta_0 + b_{0i} + (\beta_1 + b_{1i})t_{ij} + \epsilon_{ij}, \quad (1)$$

where $y_i(t_{ij})$ is the biomarker measurements for the subject i measured at time point t_{ij} , β_0 and β_1 are unknown fixed effect parameters, b_{0i} and b_{1i} are an unknown subject-specific random intercept and slope respectively which are assumed to have a bivariate normal distribution with mean zero and covariance matrix Σ , and ϵ_{ij} is the residual error for subject i at time point t_{ij} with a normal distribution $N(0, \sigma_e^2)$, which is assumed to be independent of the random effects. The above model can be extended to include other known predictors (could be fixed or time varying), such as baseline characteristics.

The sub-model for the binary outcome, D_i , is a logistic regression model that incorporates the predicted random effects, b_{0i} and b_{1i} , as predictors to obtain the risk of the future binary outcome. So, if π_i represents the probability of subject i developing the outcome conditional on the biomarker values until the last follow up time, T_i , i.e., $P(D_i = 1)$, the likelihood L for the joint model combining the two sub-models can be written as follows:

$$L = \prod_{i=1}^N \int \prod_{j=1}^{T_i} \varphi(\epsilon_{ij}) \pi_i^{D_i} (1 - \pi_i)^{1-D_i} \varphi(b_i) db_i, \quad (2)$$

where

$$\pi_i = 1 / (1 + e^{-(\alpha_0 + \alpha_1 b_{0i} + \alpha_2 b_{1i})}),$$

$$\epsilon_{ij} = y_i(t_{ij}) - (\beta_0 + b_{0i} + (\beta_1 + b_{1i})t_{ij}),$$

$b_i = (b_{0i}, b_{1i})^T$, and $\varphi(\epsilon_{ij})$ represents a normal density distribution with mean 0 and variance σ_e^2 . Likewise, $\varphi(b_i)$ is a bivariate normal density with mean zero and covariance matrix Σ .

The above joint model can be used to obtain (possibly updated) predictions for a new subject i , using the available biomarker information up until the time at which a prediction is to be made, say at timepoint t , $\hat{P}(D_i = 1 | y_i(t))$. First, the random effects $\hat{b}_{it}$ (reflecting the longitudinal profile measured up until time point t) are predicted from the linear mixed model in equation (1). Second, the predicted random effects $\hat{b}_{it}$ are used as predictors in equation (2).

In our simulations, we implement the Bayesian framework to the JM-Bin approach. Since our prior knowledge is limited, we use proper but vague prior distributions which are commonly used for the model's location and dispersion parameters, see Dandis et al., 2019 for more details

(2). We apply the MCMC technique, where two chains are initiated with 1,000 burn-in iterations and are run for 10,000 iterations. We use **rjags** package (version 4.9) which provides an interface between **R** software (version 3.6.2) and the **JAGS** library to perform the analysis (7-9).

2.1.2 Longitudinal discriminant analysis (LoDA)

Unlike JM-Bin, LoDA models the longitudinal pattern of the biomarker for the diseased and healthy subjects separately. Assuming normally distributed random effects and error terms, the $y_i(t_{ij})$'s follow different multivariate normal distributions for patients and healthy subjects as follows,

-For patients: $(y_i(t_{ij})|D_i = 1) \sim N(\mu_{1i}(t_{ij}), \Sigma_{1i}(t_{ij}))$

-For healthy subjects: $(y_i(t_{ij})|D_i = 0) \sim N(\mu_{0i}(t_{ij}), \Sigma_{0i}(t_{ij}))$,

where the means vectors μ and the covariance matrices Σ are both obtained from two linear mixed models using all the available biomarker measurements.

For a new subject i , we use the above distributions to obtain the predicted probabilities that the longitudinal data known up until time point t belong to the patients' or to the healthy subjects' group. First, we predict $\hat{P}(y_i(t)|D_i = 1)$ and $\hat{P}(y_i(t)|D_i = 0)$. Then we apply Bayes' rule to obtain the posterior probability $\hat{P}(D_i = 1|y_i(t))$ and $\hat{P}(D_i = 0|y_i(t))$, as follows.

$$\hat{P}(D_i = 1|y_i(t)) = \frac{\hat{P}(y_i(t)|D_i = 1) \hat{P}(D_i = 1)}{\hat{P}(y_i(t)|D_i = 0) \hat{P}(D_i = 0) + \hat{P}(y_i(t)|D_i = 1) \hat{P}(D_i = 1)}$$

and

$$\hat{P}(D_i = 0|y_i(t)) = \frac{\hat{P}(y_i(t)|D_i = 0) \hat{P}(D_i = 0)}{\hat{P}(y_i(t)|D_i = 0) \hat{P}(D_i = 0) + \hat{P}(y_i(t)|D_i = 1) \hat{P}(D_i = 1)}$$

where $\hat{P}(D_i = 0)$ and $\hat{P}(D_i = 1)$ are the prior probabilities of belonging to each group and are estimated using the proportions of the groups in the study historic data.

In our simulations, the linear mixed models per group are fitted using the **GLMM_MCMC** function, and LoDA is performed using the **GLMM_longitDA2** function, both from the **R** package **mixAK** (5).

2.2 Simulation design: general overview

We based our simulations on a real-life example obtained from the Dutch Central Registry for hydatidiform moles at the Radboud university medical center in Nijmegen (Radboudumc). We simulate data to predict women's post-molar gestational trophoblastic neoplasia (GTN) status (binary outcome) using weekly serum human chorionic gonadotropin (hCG) measurements (longitudinal biomarker data). The GTN status was determined after week 7. Figure 1 represents the longitudinal profiles of the 10 log-transformed hCG measurements over time from the original dataset. In general, the hCG levels for the subjects who develop GTN (GTN patients) are higher and more stable compared to the hCG levels for the healthy subjects, which start at lower levels and decrease with time. These data have been analysed elsewhere (2, 10, 11).

We generated the data using parameters for different scenarios described in Table 1 and as visualized in Figure 2. For each scenario 300 data sets were generated. Each of these data sets consists of 400 subjects to be used for model development and 400 subjects to test the predictive performance of the two approaches, using out-of-sample validation. For each subject, six biomarker values were simulated from a linear mixed model (LMM) (random intercept and random slope model). Missing values were simulated to be completely missing at random by

randomly removing 20 percent of the repeated measurements for all the subjects, so some subjects might have more missing values than others.

Both JM-Bin and LoDA use the same information for prediction, however they handle them differently. In JM-Bin, predictions typically only are based on the subject-specific deviations from the overall intercept and slope, while LoDA models the distribution of intercepts, slopes and residual errors separately per group, and in principle uses more information. Therefore, we expect that due to these structural differences between JM-Bin and LoDA, the event rate (proportion of patients in the study) and the variability of the biomarker might have an influence on the predictive performance of the two approaches. In order to examine that, we use a combination of different event rates: 10%, 30% and 50%, and different within and/or between subjects' variability. We also examine the predictive performance when the amount of variability differs between the two groups, i.e., in situations where the between- and/or within-subjects variability is larger in patients than in healthy subjects for example.

In each scenario, both JM-Bin and LoDA are fitted using all the available measurements in the training dataset (up to 6 measurements per subject) and the resulting models are used to provide out-of-sample dynamic predictions based on the available longitudinal measurements till time point t_{ij} , where $j = 1, 2, \dots, 6$. To keep it short, we obtain predictions based on the data known up until three time points, i.e., t_{i3} . The dynamic predictive accuracy is evaluated using time varying measures: the mean squared error of prediction (MSEP) and the area under the receiver operator characteristic curve (AUC). The reported prediction accuracies for both approaches are based on the averages over the 300 simulated data sets.

Table 1. Description of the different scenarios used in simulating longitudinal biomarker profiles for patients and healthy subjects.

Simulation Scenario	Patients						Healthy subjects					
	α	β	σ_{ϵ}^2	d_{11}	d_{22}	d_{12}	α	β	σ_{ϵ}^2	d_{11}	d_{22}	d_{12}
Reference dataset	2.58	0.007	0.04	0.63	0.15	-0.01	2.44	-0.31	0.04	0.63	0.15	-0.01
Increased between-subjects variability	2.58	0.007	0.04	1.26	0.30	-0.04	2.44	-0.31	0.04	1.26	0.30	-0.04
Increased within-subjects variability	2.58	0.007	4.00	0.63	0.15	-0.01	2.44	-0.31	4.00	0.63	0.15	-0.01
Increased within- and between-subjects variability	2.58	0.007	4.00	1.26	0.30	-0.04	2.44	-0.31	4.00	1.26	0.30	-0.04
Increased between-patients variability	2.58	0.007	0.04	1.26	0.30	-0.04	2.44	-0.31	0.04	0.63	0.15	-0.01
Increased within-patient variability	2.58	0.007	4.00	0.63	0.15	-0.01	2.44	-0.31	0.04	0.63	0.15	-0.01
Increased within- and between-patients variability	2.58	0.007	4.00	1.26	0.30	-0.04	2.44	-0.31	0.04	0.63	0.15	-0.01

The above scenarios were repeated for different event rates (i.e. percentage of patients) of 10%, 30% and 50%.

α = Fixed Intercept, β =Fixed slope, d_{11} =standard deviation for random intercept b_0 , d_{22} = standard deviation for random slope b_1 , d_{12} =Random effects covariance, σ_{ϵ}^2 = Residual variance. 300 simulations per scenario were conducted.

3 Results

Figure 3 displays examples of the biomarker profiles for 10 randomly selected subjects per outcome group (patients or healthy subjects) for each simulation scenario. The predictive performance measures for the different simulation scenarios for JM-Bin and LoDA are presented in Table 2 and Figure 4 and 5. The MSEP's and AUC's were calculated on the validation datasets and based on predictions using the available biomarker measurements up until the third time point. In general, LoDA produces more accurate predictions of the future event, with higher AUC's and lower MSEP's in all the studied simulation scenarios. In most of simulation scenarios, the increase in the number of patients does not influence the performance of the two approaches.

Increasing the between and/or within-subject variability for both patients and healthy subjects by increasing the variances of the random effects and/or the variance of the residuals, reduces the AUC's of both JM-Bin and LoDA approximately to the same extent. The LoDA results in slightly higher AUC's and lower MSEP's than JM-Bin when we increase the variability equally for the two outcome groups. The predictive performance of LoDA is especially better than JM-Bin when the biomarker's (within- and/or between-subjects) variability is different between the patients and the healthy subjects.

Table 2. The area under the ROC curve and mean squared error of prediction using both JM-Bin and LoDA and using different event rates and variability among patients and healthy subjects.

Simulation scenario	Event rate	LoDA		JM-Bin	
		AUC	MSEP	AUC	MSEP
Reference dataset	10%	0.86	0.07	0.86	0.07
Increased between-subjects variability		0.73	0.08	0.74	0.08
Increased within-subjects variability		0.77	0.08	0.77	0.08
Increased within- and between-subjects variability		0.68	0.09	0.68	0.09
Increased between-patients variability		0.84	0.06	0.78	0.07
Increased within-patients variability		0.90	0.05	0.80	0.07
Increased within- and between-patients variability		0.91	0.04	0.73	0.07
Reference dataset	30%	0.87	0.13	0.86	0.13
Increased between-subjects variability		0.74	0.18	0.74	0.18
Increased within-subjects variability		0.78	0.17	0.78	0.18
Increased within- and between-subjects variability		0.68	0.19	0.69	0.19
Increased between-patients variability		0.84	0.13	0.78	0.15
Increased within-patients variability		0.90	0.10	0.80	0.15
Increased within- and between-patients variability		0.91	0.09	0.73	0.16
Reference dataset	50%	0.87	0.15	0.87	0.15
Increased between-subjects variability		0.74	0.21	0.74	0.21
Increased within-subjects variability		0.78	0.19	0.78	0.20
Increased within- and between-subjects variability		0.68	0.22	0.69	0.23
Increased between-patients variability		0.84	0.16	0.79	0.19
Increased within-patients variability		0.90	0.12	0.80	0.19
Increased within- and between-patients variability		0.91	0.11	0.74	0.21

Event rates = percentage of patients; AUC=area under the ROC curve; MSEP= mean squared error of prediction.

219

220

4 Discussions

In this paper, we have compared two approaches to obtain (dynamic) predictions for a future long-term dichotomous outcome based on longitudinal biomarker data, specifically the longitudinal discriminant analysis (LoDA) and joint modeling (JM-Bin) for longitudinal and binary data. We gave an overview of the two approaches and investigated their predictive performance over different simulation scenarios, mainly by varying event rates, and biomarker's within- and /or between-subjects variability.

Both LoDA and JM-Bin resulted in equivalently high predictive performance in the reference dataset. Increasing the subject-specific variability equally among the two groups (patients and healthy subjects) decreased the predictive performance in the two approaches to the same extent. This implies that when the variability in the application of interest is similar between patients and healthy subjects, then LoDA and JM-Bin perform equally well.

When the subject-specific variability is noticeably higher in the patients compared to the healthy subjects, the LoDA provides higher prediction accuracy when compared to JM-Bin. The difference is especially large when both the within- and the between subjects' variability is larger in only one of the groups (in our study is the patient's group). This could be due to the structural difference between the two approaches. The joint model will not capture the differences in variability between the two groups since we only fit one mixed effects model for both groups, whereas in the LoDA the variability is better captured using a mixed model per outcome group. One way to improve the discrimination power of the joint model is to incorporate the within-subjects variability i.e., the subject-specific residual variance, σ_i^2 , as a predictor in the logistic submodel, similar to the work done by Parker et al (2019), where they

jointly model the individual trajectories, within-individual variability and a later continuous outcome (12).

We also investigated the effect of changing the event rate on the prediction accuracy of the two approaches, and found surprisingly, that despite the structural difference between the models, the increase in the number of patients relative to healthy subjects did not influence the prediction accuracy. We expected that LoDA would perform worse than JM-Bin in the low event rate scenarios as the LoDA in this case uses less information in the patients' group model and more information in the healthy subjects' model. However, our results were based on a minimum event rate of 10% (corresponding to 40 patients), which seems to be sufficient to capture the needed information to fit the models per group in LoDA. However, this result may not hold for smaller event rates; in fact, we found non-convergence when using smaller event rates.

Compared to the classical regression models and discriminant analyses, both LoDA and JM-Bin are more flexible, allowing to incorporate longitudinal biomarker information which typically is measured intermittently and at unbalanced time points. In addition, both can be extended by including an additional fixed or time-varying patient characteristic which may improve prediction without altering the presented methodology. Both models can be fitted using available software, however, fitting the LoDA is easier and faster than JM-Bin.

To our knowledge, this is the first simulation study to compare the predictive performance of the longitudinal discriminant analysis and the joint modeling. We used realistic simulation scenarios with different degrees of variability that are likely to be found in real world data. However, like all studies, generalization of the current findings is limited to the studied conditions. Our results are based on fitting linear mixed effects models for linear profiles, which fit the data very well. We expect to get the same results with other types of longitudinal profiles

as long as the models are correctly specified. Such scenarios require further investigation in future research. Furthermore, our interest is in predicting a future fixed outcome, known after a particular time of the follow up. However, if the time at which the outcome occurred is known and is of interest, then the joint model for longitudinal and survival data will be more suitable than our presented methodology (1).

5 Conclusions

In conclusion, our study shows that LoDA significantly outperforms JM-Bin in prediction accuracy when there is a difference in variability between outcome groups such as patients and healthy subjects. We suggest that a data analyst first investigates the variability of longitudinal biomarker for the outcome groups using plots and a mixed effects model. If there are differences in the groups' variability, we recommend using LoDA over JM-Bin.

280 **6 List of Abbreviations**

LoDA	Longitudinal Discriminant Analysis
JM-Bin	Joint Model for a binary outcome
SREM	Shared Random Effects Model
PMM	Pattern Mixture Model
GTN	Gestational Trophoblastic Neoplasia
hCG	human Chorionic Gonadotropin
LMM	Linear Mixed Model
MSEP	Mean Squared Error Of Prediction
AUC	Area Under The Receiver Operator Characteristic Curve

281

282

7 Declarations

7.1 Ethics approval and consent to participate

Not applicable

7.2 Consent for publication

Not applicable

7.3 Availability of data and materials

We used only summarized data from the following paper :

Dandis R, Teerenstra S, Massuger L, Sweep F, Eysbouts Y, IntHout J. A tutorial on dynamic risk prediction of a binary outcome based on a longitudinal biomarker. Biom J. 2020;62(2):398-413.

7.4 Competing interests

The authors declare that they have no competing interests

7.5 Funding

Not applicable

7.6 Authors' Contributions

RD performed the simulations and the analysis, interpreted the results and was a major contributor in writing the manuscript. ST and JH interpreted the results and contributed in writing the discussion section. KR contributed in writing the discussion. All authors read and approved the final manuscript.

7.7 Acknowledgements

Not applicable

8 References

1. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. In: Series CHCB, editor. New York 2012. p. 275
2. Dandis R, Teerenstra S, Massuger L, Sweep F, Eysbouts Y, IntHout J. A tutorial on dynamic risk prediction of a binary outcome based on a longitudinal biomarker. *Biom J.* 2020;62(2):398-413.
3. Lix LM, Sajobi TT. Discriminant analysis for repeated measures data: a review. *Front Psychol.* 2010;1:146.
4. Komárek A, Hansen BE, Kuiper EM, van Buuren HR, Lesaffre E. Discriminant analysis using a multivariate linear mixed model with a normal mixture in the random effects distribution. *Stat Med.* 2010;29(30):3267-83.
5. Komárek A, Komárková L. Capabilities of R Package mixAK for Clustering Based on Multivariate Continuous and Discrete Longitudinal Data. *Journal of Statistical Software; Vol 1, Issue 12 (2014).* 2014.
6. Hughes DM, Komárek A, Czanner G, Garcia-Fiñana M. Dynamic longitudinal discriminant analysis using multiple longitudinal markers of different types. *Stat Methods Med Res.* 2018;27(7):2060-80.
7. R. Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
8. Plummer M. rjags: Bayesian Graphical Models using MCMC. R package version 4-6. 2016.
9. Plummer M. JAGS: A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling. 3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria. 2003;124.
10. Eysbouts YK, L.F.A.G. Massuger LFAG, Ottevanger PB, IntHout J, Sweep FCGJ, C.M.G. T. Nomogram-based prediction of post-molar gestational trophoblastic neoplasia with serum human chorionic gonadotropin. Unpublished manuscript. 2017.

- 329 11. Eysbouts Y, Brouwer R, Ottevanger P, Massuger L, Sweep F, Thomas C, et al. Serum Human
330 Chorionic Gonadotropin Normogram for the Detection of Gestational Trophoblastic Neoplasia. *Int J*
331 *Gynecol Cancer*. 2017;27(5):1035-41.
- 332 12. Parker RMA, Leckie G, Goldstein H, Howe LD, Heron J, Hughes AD, et al. Joint modelling of
333 individual trajectories, within-individual variability and a later outcome: systolic blood pressure
334 through childhood and left ventricular mass in early adulthood. *medRxiv*. 2019:19008474.
- 335
- 336

337 **Figure Legends**

- Figure 1** Biomarker profiles for 100 randomly selected subjects from the original reference dataset (solid lines = patients, dashed lines = healthy subjects)
- Figure 2** A pictorial representation of the simulation procedure using the first three biomarker measurements as example. MSE_P= mean squared error of prediction and AUC=the area under the receiver operator characteristic curve.
- Figure 3** Biomarker profiles for 10 randomly selected subjects per simulation scenario per outcome group (left = patients, right = healthy subjects in each pair of subfigures). Thick lines with grey error bands represent biomarker means per group based on all simulated subjects.
- Figure 4** The area under the ROC curve for JM-Bin and LoDA using the scenarios from Table 1 and a 30 % event rate.
- Figure 5** The mean squared error of prediction for both JM-Bin and LoDA and using different subject-variability and 30 % event rate

338

339

Figures

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5