Supplementary Material for Vertex Exchange Method for non-parametric estimation of mixing distributions in logistic mixed models

Louise Marquart [1, 2], and Geert Verbeke [3]

[1] QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

[2] Institute for Social Science Research, University of Queensland, Brisbane, Queensland, Australia

[3] Interuniversity Institute for Biostatistics and Statistical Bioinformatics, KU Leuven, Leuven, Belgium.

Address for correspondence: Louise Marquart, QIMR Berghofer Medical Research Institute, 300 Herston Road, Brisbane 4006, Australia.

E-mail: `louise.marquart@qimrberghofer.edu.au`.

Phone: (+61) 738 453 555.

Fax: (+61) 738 453 503.

Supplementary Material 1

To assess the impact of the choice of the initial starting values and the initial grid used in Vertex Exchange Method (VEM) estimation of the random intercept logistic models, a sensitivity analysis was performed by re-fitting the VEM algorithm to the case study with alternative starting values and initial grids. Two sets of starting values were considered, parameter coefficients of the equivalent model with random intercepts assumed to be normally distributed (random intercept logistic-normal model) (Table S1), or no random intercepts (binary logistic model) (Table S2). A total of eight different combinations of initial grid size and grid range were considered, consisting of four initial number grid points, K=101, 201, 301 or 401, and the range of the grid was either 5 or 7 standard deviations of the equivalent model with assumed normal random intercepts.

The VEM approach produced similar deviance for all models. There were some minor differences observed for coefficients relating to the random effects. The intercept coefficient ranged from 3.526 to 3.900, and the random intercept variance estimate ranged from 9.93 to 20.47. The larger random intercept variance estimates were for the fits with the binary logistic model as starting values and grid sizes of 301 and 401 with range defined as $\pm 7\sigma_b$. These fits had support for the random intercept at extreme values near the boundary. For the two different starting values, there were minimal differences for the coefficients and standard errors for the parameters unrelated to the random intercept, particularly for the larger grid points of 301 and 401. The computational time for estimating the model was dependent on the number of initial grid points, with models based on 101 initial grid points required approximately an hour of CPU time, compared to over 8 hours required for the models with 401 initial grid points.

Hence, we conclude that the VEM applied to estimate the unspecified distribution of the random intercept logistic model is generally robust to the choice of initial grid, including the grid size and the range of grid points, and the choice of starting values. The parameters related to the random intercept were sensitive to the number of grid points, with boundary solutions identified for three scenarios resulting in large random intercept variance estimates.

Table S1: HILDA data: Sensitivity of VEM to the choice of initial grid, including the number of grid points (K=101, 201, 301 or 401) and the range of the grid points (5 or 7 standard deviations), when applied to the HILDA case study with starting values based on the random intercept logistic model with normal distributed random intercepts. The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance with the CPU computational time.

Starting Values	**Random intercept logistic-normal model**															
Grid Points	**101**				**201**				**301**				**401**			
Grid Range	$5\sigma_b$		$7\sigma_b$		$5\sigma_b$		$7\sigma_b$		$5\sigma_b$		$7\sigma_b$		$5\sigma_b$		$7\sigma_b$	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Constant	3.864		3.823		3.864		3.870		3.846		3.890		3.861		3.900	
(Wave-1)/10	1.038	0.095	1.053	0.095	1.040	0.095	1.044	0.095	1.041	0.095	1.040	0.095	1.040	0.095	1.041	0.095
(Baseline Age-30)/10	0.198	0.161	0.220	0.164	0.203	0.160	0.202	0.160	0.212	0.161	0.198	0.161	0.204	0.161	0.198	0.161
Marital Status																
Married/Defacto																
Sep/Div/Wid	-0.319	0.130	-0.315	0.130	-0.316	0.130	-0.316	0.130	-0.319	0.130	-0.318	0.130	-0.319	0.130	-0.315	0.130
Single	-0.052	0.257	-0.069	0.265	-0.049	0.254	-0.076	0.251	-0.038	0.256	-0.058	0.255	-0.042	0.257	-0.046	0.255
Highest Education																
Bachelor or higher																
Year 12/Dip/Cert	-1.459	0.179	-1.407	0.181	-1.457	0.178	-1.424	0.177	-1.450	0.178	-1.441	0.178	-1.455	0.178	-1.444	0.177
Year 11 or less	-2.618	0.177	-2.568	0.177	-2.613	0.176	-2.578	0.176	-2.607	0.176	-2.594	0.176	-2.611	0.176	-2.593	0.176
Dependent Children																
None																
Youngest<5 years	-2.361	0.114	-2.325	0.113	-2.358	0.114	-2.346	0.114	-2.354	0.114	-2.355	0.114	-2.358	0.114	-2.353	0.114
Youngest 5-24 years	-0.419	0.103	-0.393	0.104	-0.418	0.103	-0.407	0.104	-0.415	0.103	-0.413	0.104	-0.416	0.103	-0.415	0.104
Random Effect																
Variance	9.93		10.35		10.06		10.25		10.19		10.39		10.16		10.68	
Deviance (2×log-likelihood)	11502		11502		11502		11502		11502		11502		11502		11502	
Computational Time (hh:mm)	0:46		1:10		2:33		2:40		5:07		5:15		8:51		9:03	

Table S2: HILDA data: Sensitivity of VEM to the choice of initial grid, including the number of grid points (K=101, 201, 301 or 401) and the range of the grid points (5 or 7 standard deviations), when applied to the HILDA case study with starting values based on the binary logistic model. The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance with the CPU computational time.

Starting Values	**Binary Logistic Regression**															
Grid Points	**101**				**201**				**301**				**401**			
Grid Range	$5\sigma_b$		$7\sigma_b$		$5\sigma_b$		$7\sigma_b$		$5\sigma_b$		$7\sigma_b$		$5\sigma_b$		$7\sigma_b$	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Intercept	3.878		3.865		3.822		3.845		3.814		3.637		3.838		3.526	
(Wave-1)/10	1.037	0.095	1.044	0.095	1.043	0.095	1.046	0.095	1.043	0.095	1.048	0.094	1.039	0.095	1.047	0.095
(Baseline Age - 30)/10	0.189	0.159	0.200	0.159	0.211	0.161	0.214	0.160	0.215	0.162	0.229	0.161	0.199	0.161	0.221	0.160
Marital Status																
Married/Defacto																
Sep/Div/Wid	-0.313	0.130	-0.313	0.130	-0.317	0.130	-0.315	0.130	-0.318	0.130	-0.316	0.130	-0.318	0.130	-0.314	0.130
Single	-0.041	0.249	-0.065	0.250	-0.031	0.258	-0.027	0.253	-0.036	0.260	-0.016	0.257	-0.031	0.256	-0.023	0.256
Highest Education																
Bachelor or higher																
Year 12/Dip/Cert	-1.462	0.176	-1.417	0.176	-1.446	0.178	-1.431	0.176	-1.445	0.178	-1.425	0.177	-1.452	0.176	-1.431	0.177
Year 11 or less	-2.617	0.175	-2.566	0.175	-2.600	0.176	-2.584	0.175	-2.599	0.176	-2.578	0.174	-2.604	0.176	-2.585	0.174
Dependent Children																
None																
Youngest<5 years	-2.363	0.115	-2.344	0.114	-2.352	0.114	-2.341	0.113	-2.350	0.114	-2.334	0.113	-2.359	0.114	-2.339	0.113
Youngest 5-24 years	-0.422	0.104	-0.407	0.104	-0.414	0.103	-0.408	0.103	-0.413	0.103	-0.404	0.103	-0.418	0.104	-0.407	0.103
Random Effect																
Variance	10.31		10.33		10.75		10.69		10.75		16.27		10.41		20.47	
Deviance (2×log-likelihood)	11502		11502		11502		11502		11502		11502		11502		11502	
Computational Time (hh:mm)	0:52		0:55		2:54		2:27		6:15		5:28		8:34		9:52	

Supplementary Material 2

To assess the impact of the initial starting values and the initial grid used in VEM estimation for the random intercept and random slope logistic mixed model, a sensitivity analysis was performed by re-fitting the VEM algorithm to the case study with alternative initial grids and starting values of the fixed effect parameters. Two alternative sets of starting values for the fixed effect parameters were considered, either the starting values were the estimated coefficients of the equivalent logistic mixed model assuming bivariate normal random effects (Table S3), or the standard binary logistic model (Table S4). A total of eight different combinations of initial grid size and grid range were considered. The initial grids consisted of either 11, 21, 31 or 41 initial grid points in each dimension, such that the two dimensional grid for the two random effects were either K=11×11, 21×21, 31×31 or 41×41, and the range of the grid was based on the Cholesky decomposition multiplied by a factor of 5 or 7, $\pm 5\hat{\boldsymbol{S}}_{\boldsymbol{b}}$ or $\pm 7\hat{\boldsymbol{S}}_{\boldsymbol{b}}$. Therefore eight different combinations of the initial grid and two initial starting values were considered.

The VEM approach was robust to the initial grid choice, though coefficient estimates differed depending on the initial starting values of the coefficients. Although there were differences in the coefficients and standard error estimates, at the 5% significance level, the same inferential conclusions would be made for all VEM models. Within each initial starting value subgroup, the VEM approach was robust to the number of initial grid points resulting in similar deviances. The deviance ranged from 10738 to 11114 for models with starting values based on the logistic mixed model, and ranged from 10745 to 11071 for models with starting values based on the standard logistic model. The VEM approach was susceptible to boundary solutions, and therefore, the choice of the initial grid range impacted the estimation of parameters related to the random effects. In comparison to the grid range of $\pm 5\hat{\boldsymbol{S}}_{\boldsymbol{b}}$, the larger grid range of $\pm 7\hat{\boldsymbol{S}}_{\boldsymbol{b}}$ resulted in larger estimates for all components of the variance-covariance matrix. The CPU time required for model convergence increased with the initial number of grid points and the grid range.

The VEM used to estimate the random intercept and random slope logistic model was robust to the choice of starting values and to the choice of the initial grid, including the number of grid points and the range of the initial grid. Increasing the number of initial grid points to consist of 41 equally spaced support points resulted in a marginal improvement of the residual deviance, however the CPU time required to fit the model was approximately 4 times longer.

Table S3: HILDA data: Sensitivity of VEM to the choice of initial grid, including the number of grid points (K=11×11 to 41×41) and the range of the grid points ($\pm\, 5\hat{S}_b$ or $7\hat{S}_b$) for starting values based on coefficients of a random intercept and random slope logistic model assuming bivariate normal random effects, when applied to the HILDA case study. The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance, along with the CPU computational time.

Starting Values	**Logistic-normal mixed model**															
Grid Points	**11 × 11**				**21 × 21**				**31 × 31**				**41 × 41**			
Grid Range	$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$		$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$		$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$		$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Constant	5.695		5.779		3.275		3.131		3.599		3.332		3.322		3.153	
(Wave-1)/10	2.274		3.521		2.215		2.651		2.820		2.725		2.231		3.211	
(Age Baseline-30)/10	0.042	0.169	0.155	0.186	1.105	0.144	0.806	0.138	0.705	0.146	0.742	0.149	1.075	0.138	0.833	0.141
Marital Status																
Married/Defacto																
Sep/Div/Wid	-0.586	0.174	-0.830	0.175	-0.357	0.183	-0.589	0.161	-0.466	0.162	-0.468	0.175	-0.373	0.165	-0.415	0.163
Single	-0.304	0.220	0.004	0.230	0.322	0.248	0.247	0.241	0.260	0.228	0.270	0.253	0.294	0.233	0.296	0.228
Highest Education																
Bachelor or higher																
Year 12/Dip/Cert	-2.140	0.197	-2.429	0.178	-1.166	0.166	-1.036	0.166	-1.316	0.162	-1.273	0.167	-1.268	0.157	-1.262	0.161
Year 11 or less	-3.723	0.174	-3.849	0.208	-2.939	0.172	-2.282	0.165	-2.750	0.171	-2.861	0.170	-2.864	0.166	-2.723	0.172
Dependent Children																
None																
Youngest<5	-3.816	0.161	-3.829	0.153	-2.716	0.124	-2.501	0.133	-2.746	0.123	-2.771	0.124	-2.701	0.119	-2.695	0.124
Youngest 5-24	-1.554	0.147	-1.388	0.154	-0.793	0.122	-0.770	0.133	-0.812	0.124	-0.847	0.124	-0.760	0.119	-0.786	0.123
Random Effects																
$\sigma^2_{b_0}$	24.46		46.72		30.42		39.09		32.44		44.00		31.27		44.75	
$\sigma^2_{b_1}$	65.15		110.74		89.56		121.23		98.40		123.85		89.20		133.64	
σ_{b_0,b_1}	-20.77		-42.04		-25.36		-33.83		-25.50		-39.01		-26.41		-36.29	
Deviance (2×log-likelihood)	10894		11114		10752		10773		10740		10741		10740		10738	
Computational Time (hh:mm)	1:20		0:45		16:17		11:13		37:44		56:32		87:01		102:08	

Table S4: HILDA data: Sensitivity of VEM to the choice of initial grid, including the number of grid points (K=11×11 to 41×41) and the range of the grid points (± $5\hat{S}_b$ or $7\hat{S}_b$) for starting values based on coefficients of a binary logistic model, when applied to the HILDA case study. The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance, along with the CPU computational time.

Starting Values	**Binary Logistic model**															
Grid Points	**11 × 11**				**21 × 21**				**31 × 31**				**41 × 41**			
Grid Range	$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$		$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$		$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$		$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Constant	2.909		3.402		3.141		3.218		3.099		3.434		2.954		2.830	
(Wave-1)/10	1.800		3.215		2.831		3.318		2.011		2.205		2.170		2.690	
(Age Baseline-30)/10	1.251	0.154	0.936	0.122	0.869	0.160	0.764	0.138	1.258	0.140	0.930	0.167	1.235	0.132	1.259	0.135
Marital Status																
Married/Defacto																
Sep/Div/Wid	-0.717	0.160	-0.592	0.145	-0.514	0.161	-0.500	0.162	-0.410	0.166	-0.369	0.166	-0.414	0.161	-0.402	0.165
Single	0.347	0.300	0.101	0.239	0.242	0.246	0.076	0.259	0.229	0.240	0.235	0.258	0.321	0.239	0.293	0.239
Highest Education																
Bachelor or higher																
Year 12/Dip/Cert	-0.890	0.159	-0.836	0.143	-1.222	0.181	-0.813	0.163	-1.205	0.161	-1.440	0.183	-1.192	0.154	-1.181	0.156
Year 11 or less	-2.403	0.185	-2.145	0.173	-2.449	0.198	-2.011	0.163	-2.800	0.175	-2.970	0.222	-2.779	0.164	-2.804	0.173
Dependent Children																
None																
Youngest<5 years	-2.420	0.137	-2.189	0.119	-2.518	0.140	-2.390	0.137	-2.620	0.122	-2.721	0.140	-2.622	0.119	-2.552	0.122
Youngest 5-24 years	-0.684	0.140	-0.501	0.120	-0.728	0.137	-0.642	0.133	-0.749	0.120	-0.804	0.138	-0.726	0.118	-0.685	0.120
Random Effects																
$\sigma^2_{b_0}$	19.86		28.97		28.63		50.76		30.41		39.03		34.21		48.75	
$\sigma^2_{b_1}$	66.93		100.81		91.56		142.16		86.65		114.40		85.36		123.85	
σ_{b_0,b_1}	-11.94		-27.17		-11.68		-15.48		-21.15		-31.88		-24.27		-36.13	
Deviance (2×log-likelihood)	10867		11071		10778		10783		10749		10754		10745		10745	
Computational Time (hh:mm)	0:47		0:40		11:18		10:19		44:48		43:41		107:58		109:34	