

**Sample Size Calculation and Optimal Design for Regression-Based Norming of Tests
and Questionnaires.**

Online Supplement A

Francesco Innocenti¹, Frans E. S. Tan¹, Math J. J. M. Candel¹, and Gerard J. P. van
Breukelen^{1,2}

¹Department of Methodology and Statistics, Care and Public Health Research Institute
(CAPHRI), Maastricht University

²Department of Methodology and Statistics, Graduate School of Psychology and
Neuroscience, Maastricht University

Author Note

Francesco Innocenti  <https://orcid.org/0000-0001-6113-8992>

Math J. J. M. Candel  <https://orcid.org/0000-0002-2229-1131>

Gerard J. P. van Breukelen  <https://orcid.org/0000-0003-0949-0272>

We have no conflict of interest to disclose. A summary of this study was presented at
the 41st Annual Conference of the International Society for Clinical Biostatistics.

Correspondence concerning this article should be addressed to Francesco Innocenti,
Department of Methodology and Statistics, Care and Public Health Research Institute
(CAPHRI), Maastricht University, P.O. Box 616, 6200MD, Maastricht, the Netherlands.
Email: francesco.innocenti@maastrichtuniversity.nl

Table of Contents

Literature Review of Regression-Based Normative Studies	4
Simulation Studies	7
Checking the First-Order Taylor Series Approximations of the Sampling Variances...7	
<i>Simulation Design</i>	7
<i>Simulation Results</i>	12
Assessing the Bias of the Variance Estimators	23
<i>Simulation Design</i>	23
<i>Simulation Results</i>	24
Outline of the R Codes for the Simulation Studies	34
Optimal and Robust Design	37
Derivation of the D-Optimal Designs in Table 2 (main text).....	37
Robust Design.....	39
<i>Maximin Design Based on the Relative Efficiency Criterion</i>	39
<i>Maximin Design Based on the Efficiency Criterion</i>	41
 List of Figures and Tables	
Figure S.A.1.....	6
Figure S.A.2.....	10
Figure S.A.3.....	15

Figure S.A.4.....	16
Figure S.A.5.....	17
Figure S.A.6.....	18
Figure S.A.7.....	19
Figure S.A.8.....	20
Figure S.A.9.....	21
Figure S.A.10.....	22
Figure S.A.11.....	26
Figure S.A.12.....	27
Figure S.A.13.....	28
Figure S.A.14.....	29
Figure S.A.15.....	30
Figure S.A.16.....	31
Figure S.A.17.....	32
Figure S.A.18.....	33
Table S.A.1.....	6
Table S.A.2.....	9
Table S.A.3.....	42

Literature Review of Regression-Based Normative Studies

A literature review of normative studies under the regression-based approach was performed to have an overview of the following features: (i) sample size, (ii) violation of the normality and/or homoscedasticity assumptions, and (iii) most common predictors, and type of interaction between age and sex. The search was performed (in September 2019) using two archives: PubMed, an archive of biomedical and life sciences journals, and PsycINF, a database of abstracts in psychology produced by the American Psychological Association. The keywords “normative data” and “regression-based” were used and the search was restricted to documents in English (but no year range restriction), giving 109 results in PubMed and 68 in PsycINF. The list of results obtained from PubMed contained all titles from PsycINF, with the exception of two results (a corrigendum and a dissertation) which were excluded from the review. Another 44 papers were discarded for various reasons, such as being methodological papers or longitudinal studies. The remaining 65 papers dealt with norming psychological tests for assessing various cognitive functions, mostly in adults (58/65 papers). Online supplement B contains the list of the 109 papers, the reasons why the 44 papers were discarded, and the information extracted from the 65 considered papers.

In most of the 65 considered studies, a number of outcome variables were normed separately (i.e. with different regression models). These outcomes were either independent psychological tests or sub-scores of the same test, for a total of 396 outcome variables. Due to missing data, outliers or other technical reasons, the sample size varied across outcome variables. The distribution of the size of the normative sample across the 65 studies is shown in Figure S.A.1.

Table S.A.1 shows how frequently the assumptions of normality and homoscedasticity were (i) satisfied, (ii) not satisfied, or (iii) their validity/non-validity was

not reported (i.e. “unknown”), over the 396 multiple linear regression models. For only 165 of the 396 models, the validity/non-validity of both assumptions were stated. Of these 165 models, 117 (71%) satisfied both assumptions, 23 (14%) satisfied one assumption, and 25 (15%) none. When normality and homoscedasticity were both violated, a number of solutions was adopted. The most popular was, first, to compute the standard deviation of the residuals per quartile of the predicted values, then to standardize the residuals with these standard deviations and, finally, to estimate percentiles from the empirical distribution of the standardized residuals. As shown in Table S.A.1, for 231 of the 396 outcomes the users of the normative data are left with the uncertainty about whether the provided norms are sound or not, because the validity/non-validity of at least one assumption was not stated. These concerns are partially attenuated by noting that, among those studies that did not state the assumption checks, a number of them used the empirical cumulative distribution of the (non-standardized) residuals or other non-parametric methods to estimate percentiles, which at least addresses possible non-normality (but not heteroscedasticity).

In the 65 studies, age, sex, and education were the most common predictors. The model used for norming included age for 349/396 outcomes, sex for 178/396 outcomes, and education for 321/396 outcomes (of which 226 treated education as quantitative and 95 as categorical). Of these 396 models, 101 included the main effects of age and sex only, 56 included the main effects of age and sex, and a quadratic age effect, one model included the main effects of age and sex, and their interaction, and one model included the main effects of age and sex, their interaction, and a quadratic age effect. None of the models included the interaction between sex and the quadratic effect of age. These results should be taken with caution because not all studies tested the presence of all possible interactions or of quadratic effects.

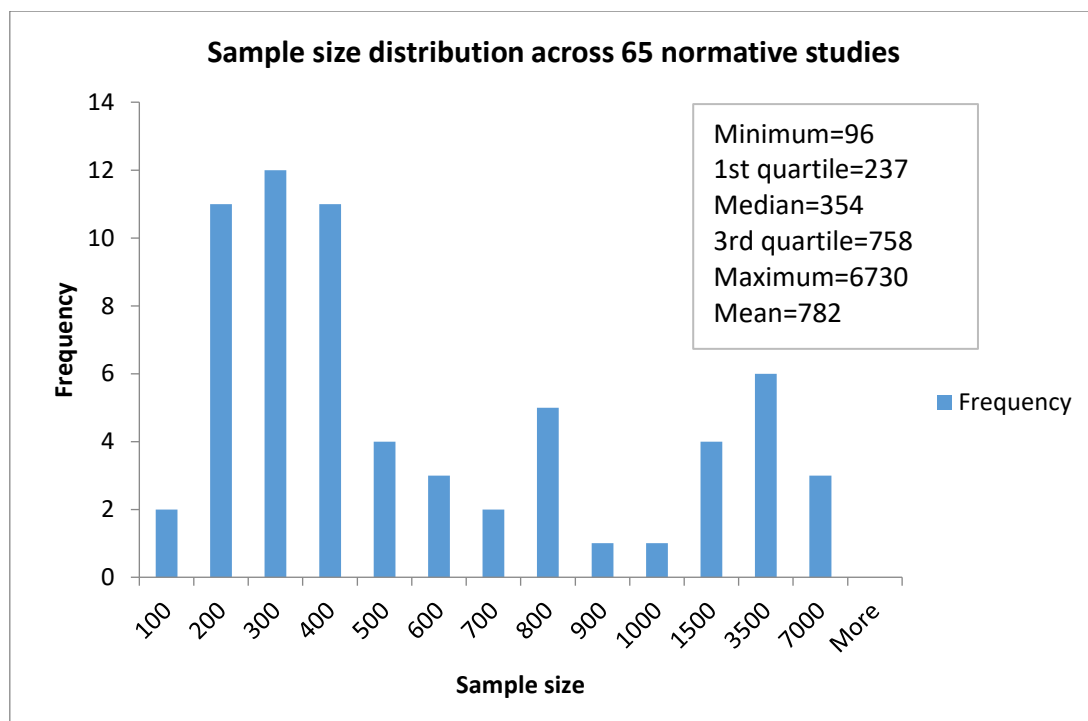
Table S.A.1

Frequencies of the Normality and Homoscedasticity Assumptions in the 396 Multiple Regression Models

	Homoscedasticity			
Normality	Satisfied	Not satisfied	Unknown	Total
Satisfied	117	7	67	191
Not satisfied	16	25	26	67
Unknown	0	0	138	138
Total	133	32	231	396

Figure S.A.1

Distribution of the Size of the Normative Sample Across the 65 Normative Studies



Simulation Studies

Checking the First-Order Taylor Series Approximations of the Sampling Variances

Simulation Design

Equations (7) and (8) in the main text are based on first-order Taylor series approximations, so their bias must be assessed. This is done through a simulation study in which equations (7) and (8) are compared with the true $V(\hat{Z}_0)$ and $V(PR(\hat{Z}_0))$, for which no analytical expressions are known. To generate the true $V(\hat{Z}_0)$ and $V(PR(\hat{Z}_0))$, one needs to artificially reproduce what would happen in practice: (i) a normative sample is drawn from the reference population, (ii) the model parameters are estimated using the normative sample, and (iii) the raw outcome of an individual (not belonging to the normative sample) is translated into an estimated Z-score and a PR-score. To compute $V(\hat{Z}_0)$ and $V(PR(\hat{Z}_0))$, these three steps must be repeated as many times as possible (here, $S = 20,000$ times) while keeping the individual's raw outcome in step (iii) constant. An additional goal of the simulation study is to assess the bias of \hat{Z}_0 and $PR(\hat{Z}_0)$. While \hat{Z}_0 is not expected to be biased because $\hat{\beta}$ and $\hat{\sigma}_\varepsilon^2$ are unbiased (Johnson & Wichern, 1998, pp. 389-390), there might be bias in $PR(\hat{Z}_0)$, since equation (6) is a nonlinear function of \hat{Z}_0 .

In the simulation study, the considered true Z-scores were all values from -3 to 3 with increment 0.5 , yielding $M = 13$ values. The considered PR-scores were not those corresponding to the 13 Z-scores but $PR_0 \in \{1, 2.5, 5, 10, 90, 95, 97.5, 99\}$, because the latter set of values is more relevant in practice. For $X_1 = \text{age}$, all the values from 20 to 80, with step 5, were considered, giving $Q = Q_1 \times Q_2 = 13 \times 2 = 26$ combinations of age and sex (age was centered before computing quadratic and interaction effects with models (1)-(5), and sex was coded 0/1). Hence, $M \times Q$ was the number of combinations of age, sex, and Z-score, as

well as the number of individuals and raw scores Y_0 to whom the norms are applied. For the size of the generated normative samples, four values were considered: $N = \{338, 676, 1690, 3380\}$. These four values correspond to four values of $L = \{1, 2, 5, 10\}$, where L denotes the number of replications per combination of age, sex, and Z-score, so $N = M \times Q \times L$. These N -values are in line with literature. Specifically, Oosterhuis et al. (2016) found in their literature review that the sample size $N \in [122, 96585]$, and for 68% of the normative samples $N \in [500, 2500]$. In the literature review of this paper, $N \in [96, 6730]$ and the first, second, and third quartiles of the sample size distribution were 237, 354, and 758, respectively, while the average N was 782 (see Figure S.A.1).

With respect to the choice of true regression parameter values used to generate normative samples, Table S.A.2 shows three sets of true values. For $\beta_0, \beta_1, \beta_2, \beta_3$, and σ_ε , the true values were taken from the estimates of the corresponding parameters in three normative studies, where the assumptions of normality and homoscedasticity were both met (see Goretti et al., 2014; Parmenter et al., 2010; Van der Elst et al., 2006). Unfortunately, estimates for β_4 and β_5 were not reported in the aforementioned studies, because the associated predictors were not significant. Therefore, the true values of β_4 and β_5 were chosen such as to have three different patterns of $E(Y)$ relative to sex and age, as shown in Figure S.A.2. Specifically, the difference in $E(Y)$ between men and women (i) decreases as age increases for the Profession Naming verbal fluency test (PNVFT), (ii) increases as age increases for the Delis-Kaplan Executive Function System Sorting test (DKEFS), (iii) is larger for extreme values of age for the Symbol Digit Modalities Test (SDMT).

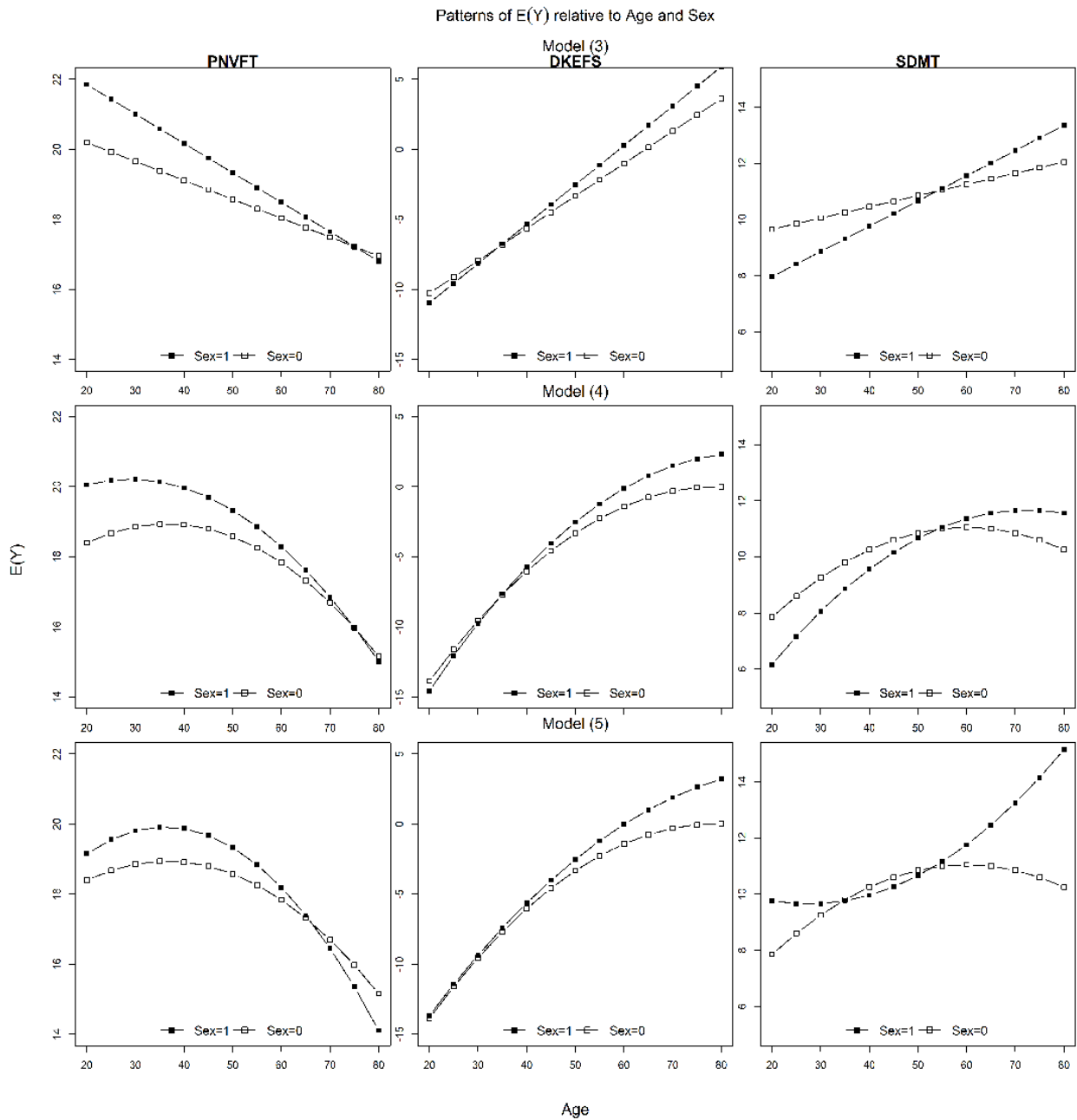
Table S.A.2*True Values of the Model Parameters*

		Table 2 of Van der Elst et al. (2006)	Tables 2 and 4 of Parmenter et al. (2010)	Tables 4-5 of Goretti et al. (2014)
	Y	Profession Naming verbal fluency test (PNVFT)	Delis-Kaplan Executive Function System (DKEFS) Sorting test	Symbol Digit Modalities Test (SDMT)
SD residuals	σ_{ε}	4.818	2.605	2.658
Intercept	β_0	18.579	-3.316	10.854
Age	β_1	-0.054	0.231	0.040
Sex	β_2	0.753	0.794	-0.192
Age ²	β_3	-0.002	-0.004	-0.002
Age×Sex	β_4	-0.030	0.050	0.050
Age ² ×Sex	β_5	-0.001	0.001	0.004
Range and coding of the predictors in the original study		Age(years - 50) ∈ [-25, 30], Sex ∈ {0,1}	Age(years) ∈ [20, 60], Sex ∈ {1, 2}	Age(years) ∈ [18, 65], Sex ∈ {1, 2}

Figure S.A.2

Expected Value of the Outcome Variable as a Function of Age, for Different Values of Sex

(curves), Different Models (rows), and Different Outcome Variables (columns)



The steps of the simulation study are the following (for technical details, see the Outline of the R codes for the Simulation Studies section).

1. *Choose population parameter values.*

- a. Choose a set of true values for the model parameters $(\boldsymbol{\beta}, \sigma_\varepsilon^2)$ from Table S.A.2, the joint distribution of the predictors age and sex and the set of true Z-score values (and the corresponding true PR-scores) as described before.
- b. For each combination of age, sex, and Z-score compute the corresponding raw score Y_0 (i.e. by adding $\varepsilon_0 = \sigma_\varepsilon Z_0$ to $\mathbf{x}_0' \boldsymbol{\beta}$), and compute the values of equations (7) and (8). These $M \times Q$ raw scores and values of equations (7) and (8) belong to the $M \times Q$ individuals to whom the norms will be applied (i.e. not the members of the normative sample).

2. *Generate S normative samples of size $N = M \times Q \times L$.*

Generate S normative samples of size N (i.e. draw N residuals ε from $N(0, \sigma_\varepsilon^2)$ and do this S times). For each of these S normative samples, compute the N raw scores \mathbf{y} (i.e. by adding the N residuals ε to $\mathbf{X}\boldsymbol{\beta}$, where \mathbf{X} is obtained by replicating L times each combination of age, sex, and Z-score), and estimate the model parameters. Note that the same S samples of N residuals ε was used for each model parameter sets in Table S.A.2.

3. *Calculate the true average and variance of \hat{Z}_0 and $PR(\hat{Z}_0)$ over the S normative samples.*

For each individual and thus for each raw score Y_0 from step 1b (i.e. not members of the normative sample), compute:

- a. \hat{Z}_0 and $PR(\hat{Z}_0)$ S times (i.e. using $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_\varepsilon^2$ obtained from each of the S normative samples),
- b. the true variances of \hat{Z}_0 and $PR(\hat{Z}_0)$ over the S normative samples, and

- c. the averages of \hat{Z}_0 and $PR(\hat{Z}_0)$ over the S normative samples,
4. Calculate the (relative) bias of the first-order Taylor series approximation of the statistics.

Calculate the relative biases of equations (7) and (8)

(i.e. $R.B. = \frac{\text{approximated variance} - \text{true variance}}{\text{true variance}}$), and the (absolute) biases of \hat{Z}_0 and

$PR(\hat{Z}_0)$. Here, the approximated variance is as obtained with equations (7) and (8) in step 1b and the true variance is obtained in step 3b.

Since each individual from step 1b corresponds to a specific combination of age, sex, and Z-score, one obtains the relative biases of equations (7) and (8), and the absolute biases of \hat{Z}_0 and $PR(\hat{Z}_0)$, for each combination of true Z-score/true PR-score with age and sex.

Simulation Results

Figures S.A.3-S.A.10 show the results of the simulation study for PNVFT under models (1) and (5), which were chosen to show the patterns of the bias as the complexity of the model increases. These figures show the results only for the Z-scores and PR-scores that are most useful in practice, that is, $Z_0 \in \{-2.5, -2, -1.5, +1.5, +2, +2.5\}$ and $PR_0 \in \{1, 2.5, 5, 10, 90, 95, 97.5, 99\}$. Online supplement B contains the results obtained under the other models (Figures S.B.1.1-S.B.1.12), and for the PR-scores corresponding to $Z_0 \in \{-2.5, -2, -1.5, +1.5, +2, +2.5\}$ (Figures S.B.1.25-S.B.1.34). Online supplement B provides the results for DKEFS (Figures S.B.2.1-S.B.2.20), and SDMT (Figures S.B.2.21-S.B.2.40). All figures have the same structure. Each figure refers to a specific model, and is composed of eight panels. Within a figure, each column refers to a value of sex, while each row refers to a sample size. Each panel shows a bias per Z- or PR-score that is plotted against age. For each combination of age, sex, and true Z-score, the corresponding \hat{Z}_0 was the same across the three sets of true β and σ_ϵ values in Table S.A.2. As a result, the same was true for

the relative biases of equations (7) and (8), and the absolute biases of \hat{Z}_0 and $PR(\hat{Z}_0)$ (see online supplement B). This follows from having used the same S samples of N residuals for each of the three sets of true model parameter values in Table S.A.2. Hence, only the figures for PNVFT are mentioned in presenting the simulation results.

Z-Scores. Figures S.A.3-S.A.4 (and S.B.1.1-S.B.1.3 in online supplement B) show the relative bias ($R.B.$) of $V(\hat{Z}_0)$ (i.e. equation (7)). In all the considered scenarios, this relative bias was always within the interval $(-3\%, 3\%)$, which means that $V(\hat{Z}_0)$ (i.e. equation (7)) is accurate even for small sample sizes (e.g. $N = 338$). It is not clear how the relative bias is affected by the predictors or the sample size. Specifically, increasing the sample size the bias should decrease, since the Delta method is a large-sample approximation, but in some scenarios (e.g. Figure S.A.3), this is not the case. This might entail that the considered sample sizes are not large enough for the bias to vanish. But in any case and most importantly, $R.B. \in (-3\%, 3\%)$ in all scenarios. Figures S.A.5-S.A.6 (and S.B.1.4-S.B.1.6 in online supplement B) show that, as expected, \hat{Z}_0 is (nearly) unbiased (i.e. absolute bias $\in (-0.01, 0.01)$) for any Z-score, age by sex combination, sample size, and model.

PR-Scores. Figures S.A.7-S.A.8 (and S.B.1.7-S.B.1.9 in online supplement B) show the relative bias of $V(PR(\hat{Z}_0))$ (i.e. equation (8)). Equation (8) tends to have a larger relative bias than $V(\hat{Z}_0)$ (i.e. equation (7)). The relative bias decreases as the sample size increases and becomes acceptable (i.e. $R.B. \in [-5\%, 3\%]$) if $N \geq 1690$ for models (3)-(5) or if $N \geq 676$ for models (1)-(2), except for the combinations of extreme values of age (i.e. age = 20 and 80) with extreme PR-scores ($PR_0 = 1$ or 99), for which combinations the relative bias is within the range $[-10\%, 3\%]$. Furthermore, the relative bias of $V(PR(\hat{Z}_0))$ (i.e. equation (8)) increases as age moves away from the average age of 50 years toward 20 or 80, and as

the PR-score moves away from 50 toward 1 or 99, and as the complexity of the model increases (i.e. moving from model (1) to model (5)). The bias of $V\left(PR(\hat{Z}_0)\right)$ does not depend in a systematic way on sex, which makes sense as there are only two values of sex and these are equally far removed from the average due to the 50:50 distribution of sex. Figures S.A.9-S.A.10 (and S.B.1.10-S.B.1.12 in online supplement B) show the absolute bias of $PR(\hat{Z}_0)$ (i.e. equation (6)). The bias decreases as the sample size increases and is always within the interval $[-0.1, 0.1]$, on the scale from 0 to 100, if $N \geq 1690$. Furthermore, $PR(\hat{Z}_0)$ overestimates $PR(Z_0)$ if $PR_0 < 50$ and underestimates it if $PR_0 > 50$. The bias increases when moving from $PR_0 = 1$ to $PR_0 = 10$ or from $PR_0 = 99$ to $PR_0 = 90$, or as age is further away from the average age, or as model complexity increases, and does not systematically depend on sex.

Figure S.A.3

Relative Bias of Equation (7), $V(\hat{Z}_0)$, Under Model (1) for $Y = PNVFT$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $Z_0 \in \{\pm 1.5, \pm 2, \pm 2.5\}$ (curves)

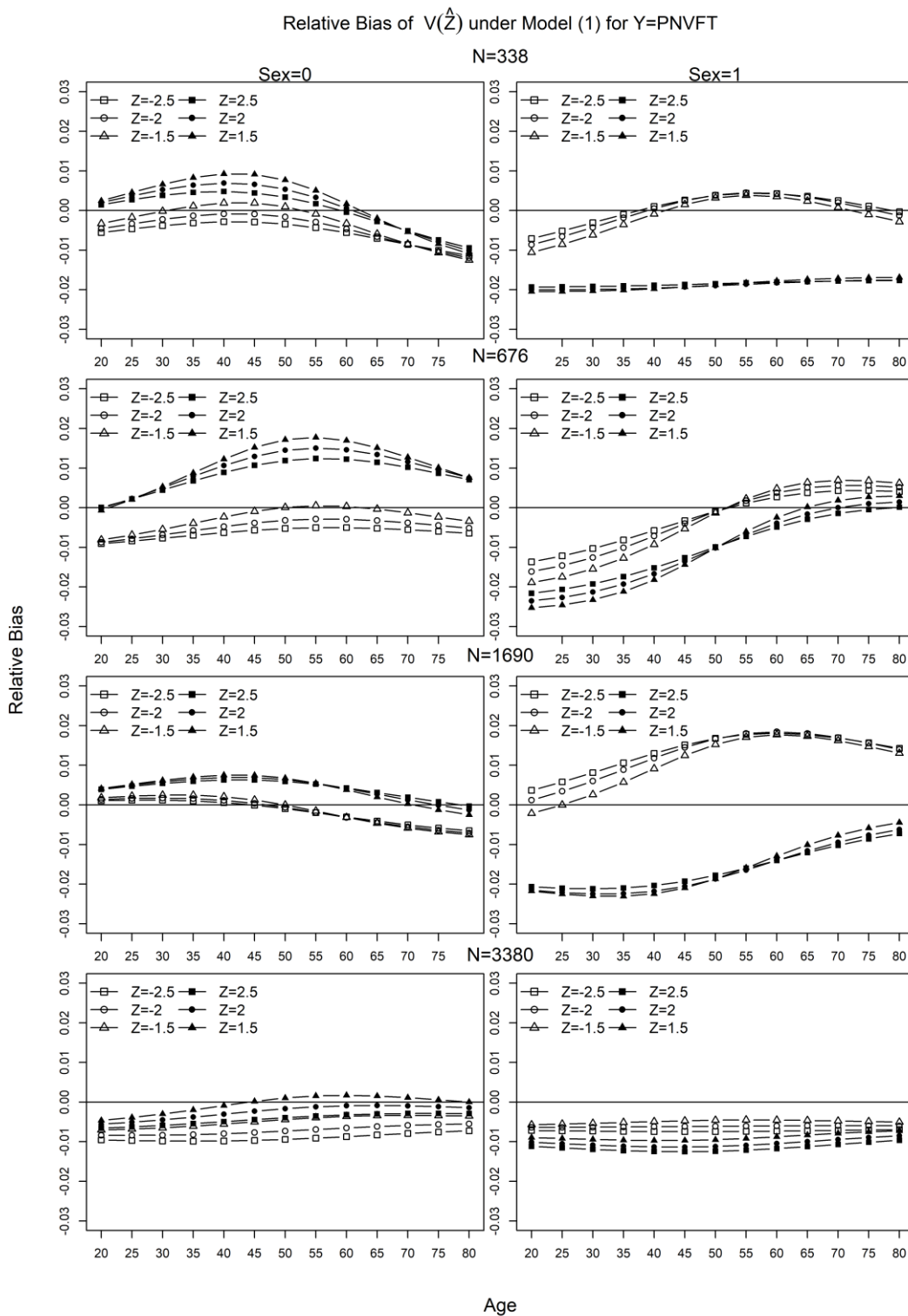


Figure S.A.4

Relative Bias of Equation (7), $V(\hat{Z}_0)$, Under Model (5) for $Y = \text{PNVFT}$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $Z_0 \in \{\pm 1.5, \pm 2, \pm 2.5\}$ (curves)

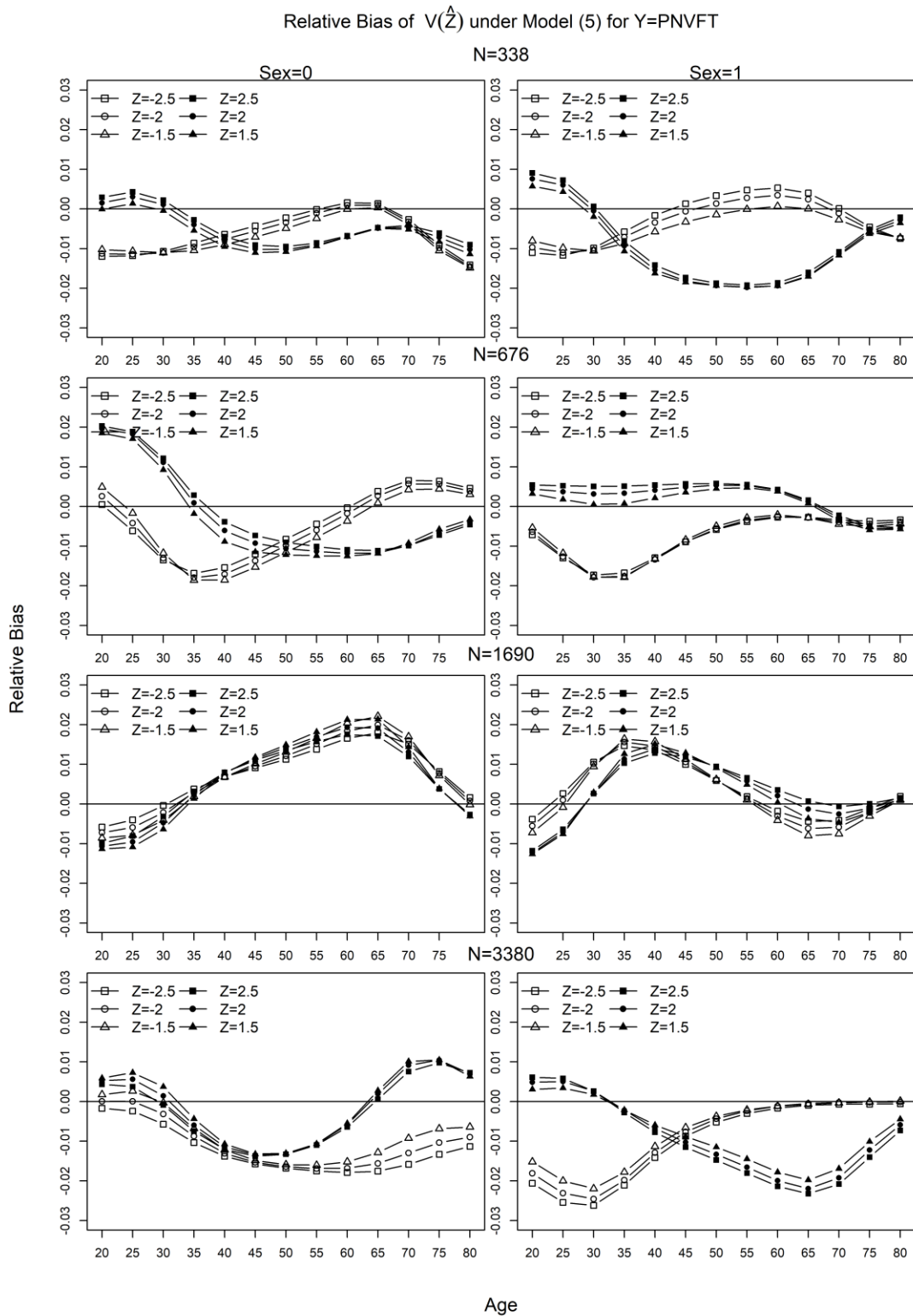
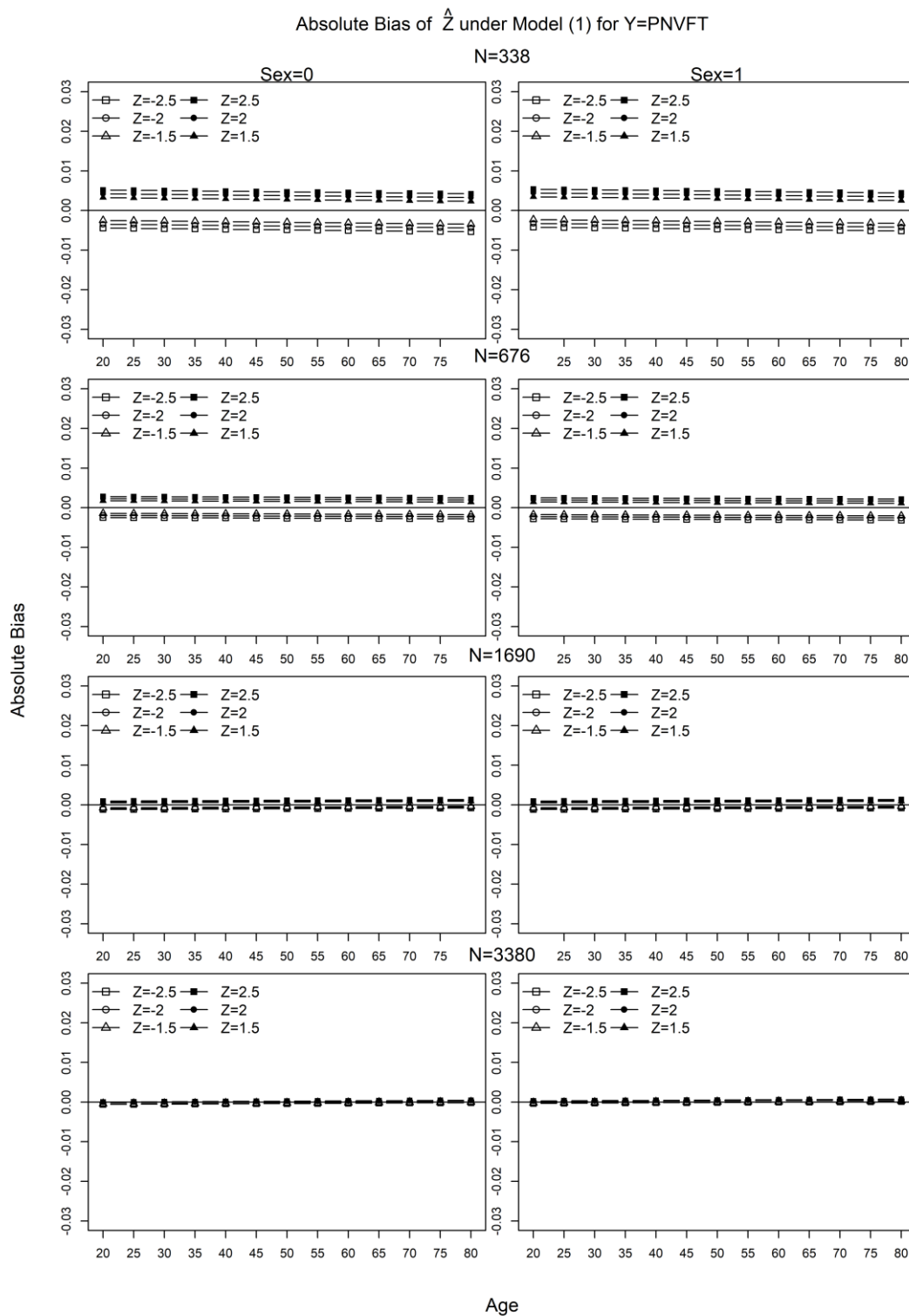


Figure S.A.5

Absolute Bias of \hat{Z}_0 , Under Model (1) for $Y = PNVFT$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $Z \in \{\pm 1.5, \pm 2, \pm 2.5\}$ (curves)



Age

Figure S.A.6

Absolute Bias of \hat{Z}_0 , Under Model (5) for $Y = \text{PNVFT}$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $Z \in \{\pm 1.5, \pm 2, \pm 2.5\}$ (curves).

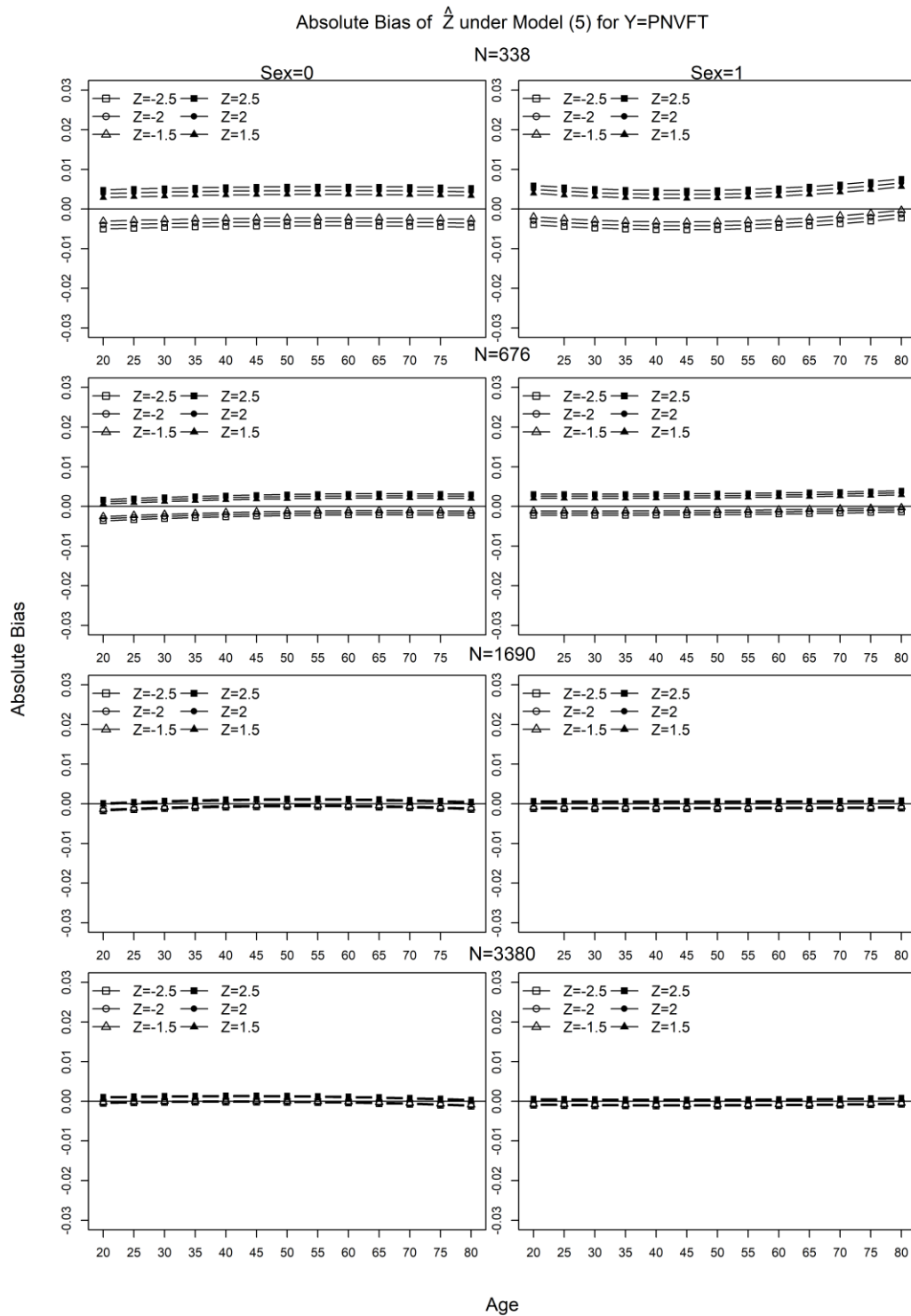
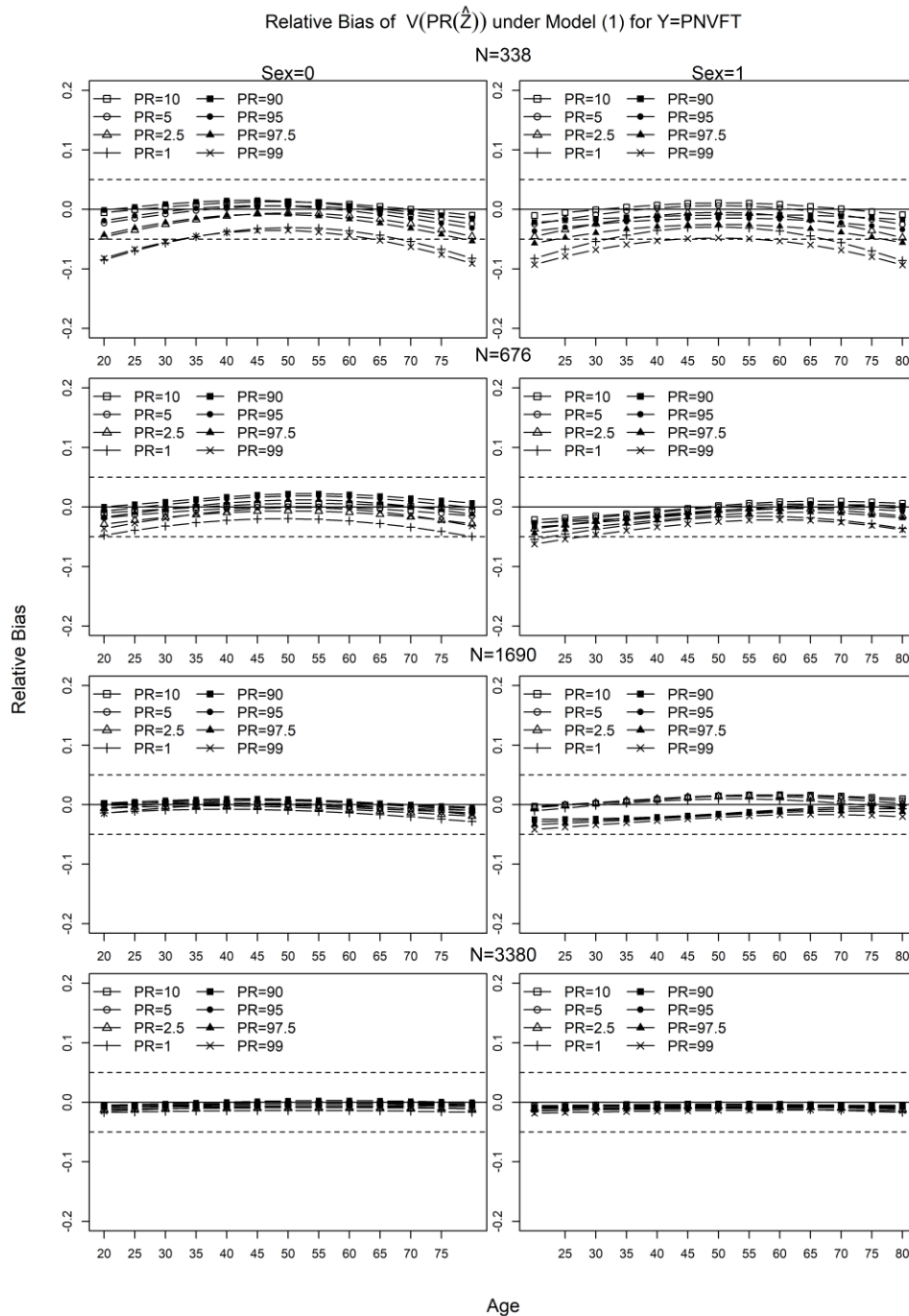


Figure S.A.7

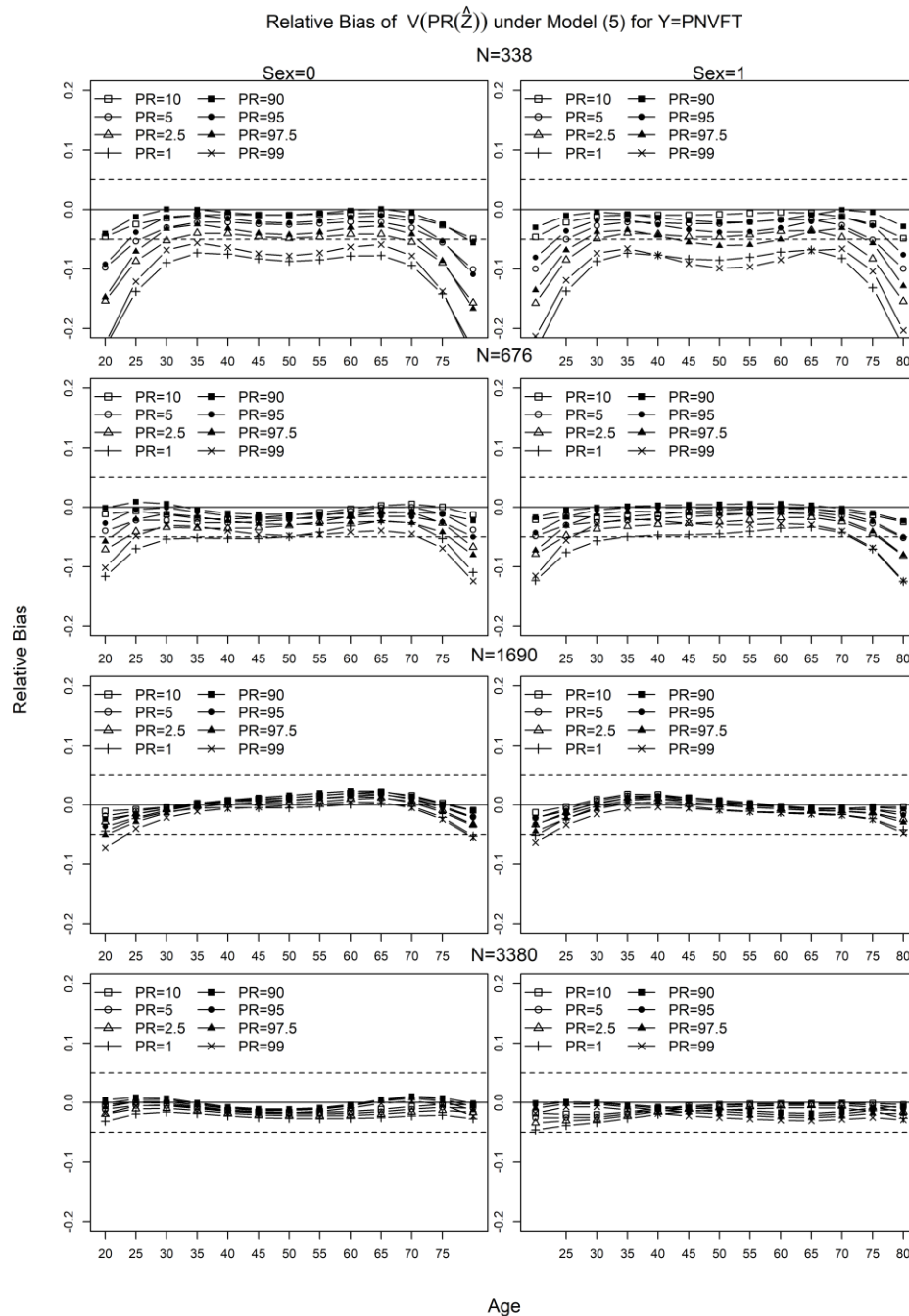
Relative Bias of Equation (8), $V\left(PR(\hat{Z}_0)\right)$, Under Model (1) for $Y = PNFT$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $PR_0 \in \{1,2.5,5,10,90,95,97.5,99\}$ (curves)



Note. The relative bias can be considered as acceptable within the range $[-5\%, 5\%]$ (horizontal dotted lines)

Figure S.A.8

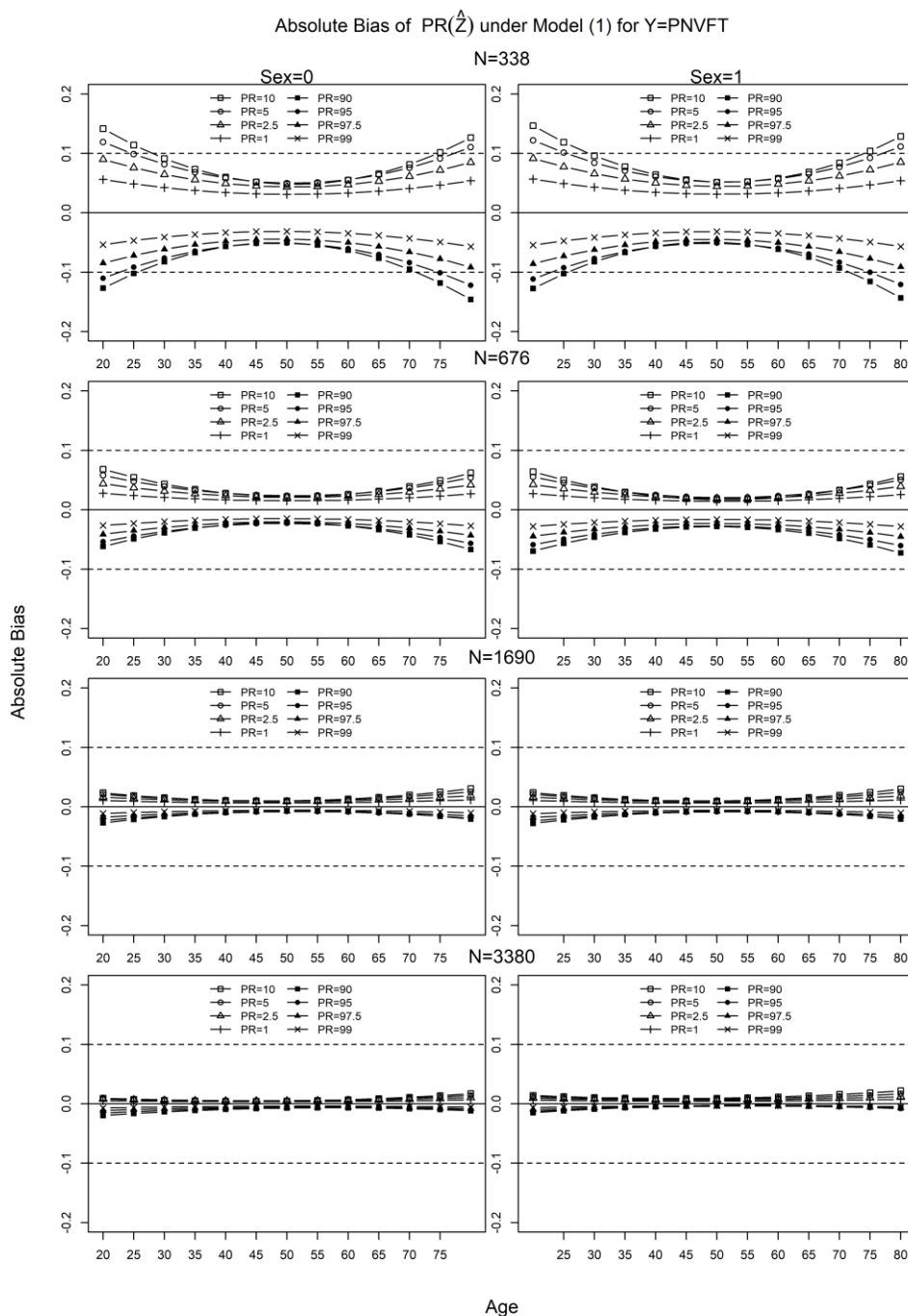
Relative Bias of Equation (8), $V\left(PR(\hat{Z}_0)\right)$, Under Model (5) for $Y = PNVFT$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $PR_0 \in \{1,2.5,5,10,90,95,97.5,99\}$ (curves)



Note. The relative bias can be considered as acceptable within the range $[-5\%, 5\%]$ (horizontal dotted lines).

Figure S.A.9

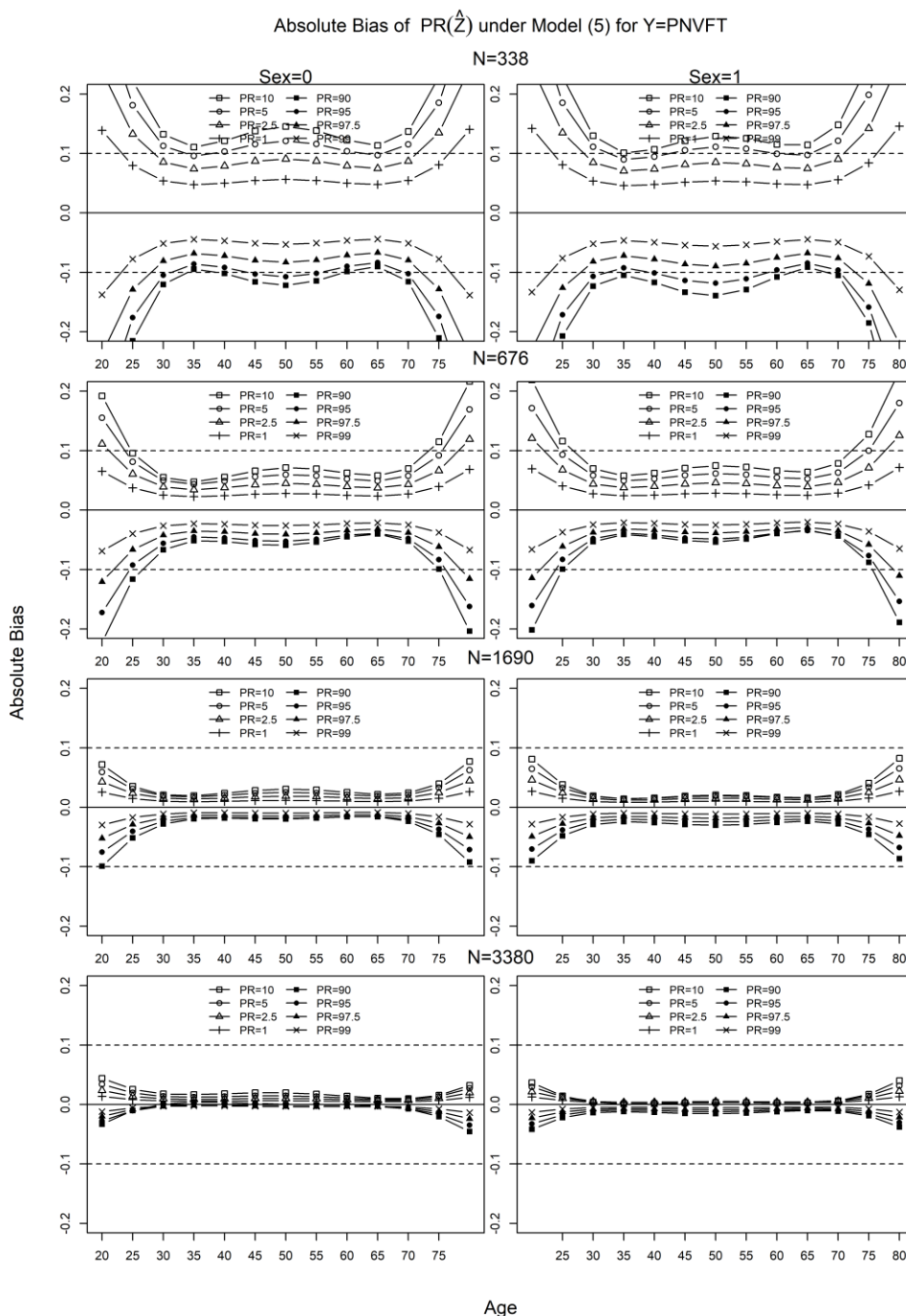
Absolute Bias of Equation (6), $PR(\hat{Z}_0)$, Under Model (1) for $Y = PNVFT$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $PR_0 \in \{1,2.5,5,10,90,95,97.5,99\}$ (curves)



Note. The absolute bias can be considered as acceptable within the range $[-0.1,0.1]$ on the scale from 0 to 100 (horizontal dotted lines).

Figure S.A.10

Absolute Bias of Equation (6), $PR(\hat{Z}_0)$, Under Model (5) for $Y = PNVFT$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $PR_0 \in \{1,2.5,5,10,90,95,97.5,99\}$ (curves)



Note. The absolute bias can be considered as acceptable within the range $[-0.1,0.1]$ on the scale from 0 to 100 (horizontal dotted lines).

Assessing the Bias of the Variance Estimators

Simulation Design

To assess the bias of $\hat{V}(\hat{Z}_0)$ and $\hat{V}(PR(\hat{Z}_0))$ (i.e. equations (9) and (10)) relative to the true variances of \hat{Z}_0 and $PR(\hat{Z}_0)$, a new simulation study is needed to generate the true (unknown) variances. The simulation procedure follows the same steps as that described in the previous section up to step 3 (for details, see the Outline of the R codes for the Simulation Studies section). In the new step 3, the individual's estimated Z-scores (so \hat{Z}_0) are used to compute equations (9) and (10) instead of using their true Z-scores to compute equations (7) and (8), for each of the S normative samples. The averages of equations (9) and (10) are then computed across the S normative samples per combination of predictors and true Z-score of interest (this averaging was not needed for equations (7) and (8) that do not depend on the normative sample). In the new step 4, the relative biases of equations (9) and (10) are computed from these averages as $\frac{\text{average of estimated variance} - \text{true variance}}{\text{true variance}}$, where the true variance is computed as the variance of \hat{Z}_0 and $PR(\hat{Z}_0)$ across the S normative samples (like in the previous simulation study). Furthermore, for each combination of predictors and Z-score of interest, the coverage of the 95% confidence interval for Z_0 and PR_0 was computed across the S normative samples. The coverage was defined as the proportion of normative samples in which the 95% confidence interval, computed using equations (9) and (10), contained the true Z_0 and PR_0 .

The M relevant Z-scores are the same as those of the previous simulation study, and so are the values of $X_1 = \text{age}$ and $X_2 = \text{sex}$. The true values of β and σ_ε^2 are those shown in Table S.A.2 for PNVFT only, because the previous simulation study has shown that these true values affect neither the bias of equations (7) and (8) nor that of \hat{Z}_0 . Since equation (7)

for $V(\hat{Z}_0)$ was already accurate for $N = 338$, the sample sizes used to check the bias of equation (9) are the same as those of the previous simulation study. In contrast, for equation (10) the sample sizes are $N = \{676, 1690, 2028, 3380\}$, which correspond to $L = \{2, 5, 6, 10\}$ replications per combination of age, sex and true PR_0 , because the bias of equation (8) for $V(PR(\hat{Z}_0))$ was too large in most scenarios if $N = 338$.

Simulation Results

The results of this simulation study are shown in Figures S.A.11-S.A.18 for models (1) and (5), and in Figures S.B.1.13-S.B.1.24 in online supplement B for the other models. The structure of the figures is the same as that of the previous simulation study. The focus is on $Z_0 = \{-2.5, -2, -1.5, +1.5, +2, +2.5\}$ and $PR_0 \in \{1, 2.5, 5, 10, 90, 95, 97.5, 99\}$. Results for the PR-scores corresponding to $Z_0 = \{-2.5, -2, -1.5, +1.5, +2, +2.5\}$ are in Figures S.B.1.35-S.B.1.44 (online supplement B).

Z-Scores. Figures S.A.11-S.A.12 (and S.B.1.13-S.B.1.15 in online supplement B) show the relative bias of $\hat{V}(\hat{Z}_0)$ (i.e. equation (9)), which has similar patterns to that of $V(\hat{Z}_0)$ (i.e. equation (9)). Furthermore, the relative bias of $\hat{V}(\hat{Z}_0)$ (i.e. equation (9)) is always within the range $(-3\%, 3\%)$ in all scenarios, just like the relative bias of $V(\hat{Z}_0)$ (i.e. equation (7)). Figures S.A.13-S.A.14 (and S.B.1.16-S.B.1.18 in online supplement B) show the coverage of the 95% confidence interval for Z_0 obtained using $\hat{V}(\hat{Z}_0)$. In all considered scenarios (models, sample sizes), the coverage is close to the nominal coverage probability (i.e. 0.95) for all true Z_0 scores.

PR-Scores. Figures S.A.15-S.A.16 (and S.B.1.19-S.B.1.21 in online supplement B) show the relative bias of $\hat{V}(PR(\hat{Z}_0))$ (i.e. equation (10)). This relative bias tends to decrease as sample size increases and is always within the interval $[-3\%, 10\%]$. For $N \geq 1690$, the

bias is always within $[-3\%, 5\%]$. In general, the patterns of the bias of $\hat{V}\left(PR(\hat{Z}_0)\right)$ tend to be less smooth than those of the bias of $V\left(PR(\hat{Z}_0)\right)$. Figures S.A.17-S.A.18 (and S.B.1.22-S.B.1.24 in online supplement B) show the coverage of the 95% confidence interval for PR_0 obtained using equation (10). If $N = 676$, the coverage for PR_0 tends to be worse than that for Z_0 , but never smaller than 93%. The coverage tends to decrease moving from model (1) to model (5), but it is always close to 95% if $N \geq 1690$.

Figure S.A.11

Relative Bias of Equation (9), $\hat{V}(\hat{Z}_0)$, Under Model (1) for $Y = \text{PNVFT}$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $Z_0 \in \{\pm 1.5, \pm 2, \pm 2.5\}$ (curves)

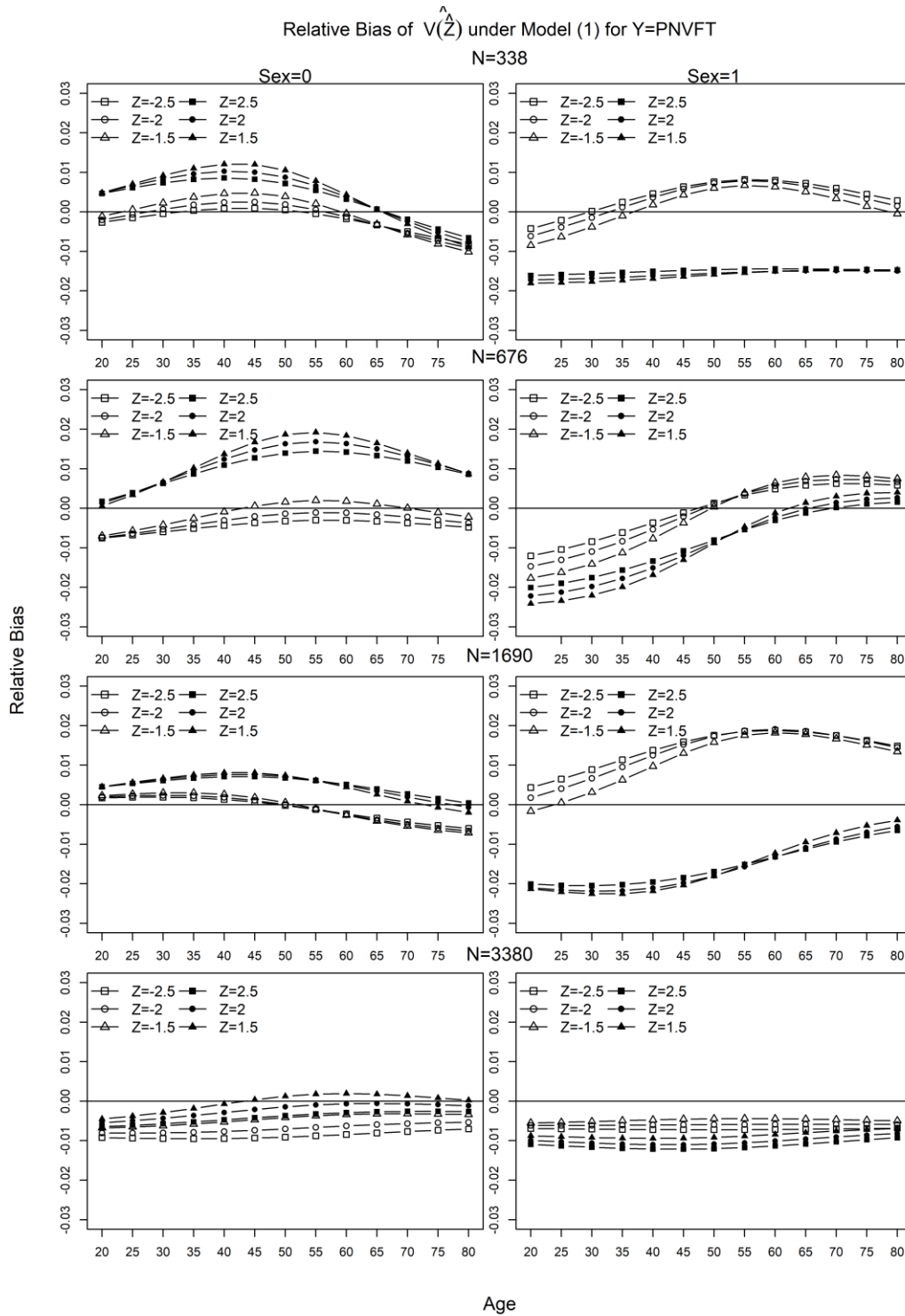


Figure S.A.12

Relative Bias of Equation (9), $\hat{V}(\hat{Z}_0)$, Under Model (5) for $Y = \text{PNVFT}$, as a Function of Age $\in [20,80]$ (x -axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $Z_0 \in \{\pm 1.5, \pm 2, \pm 2.5\}$ (curves)

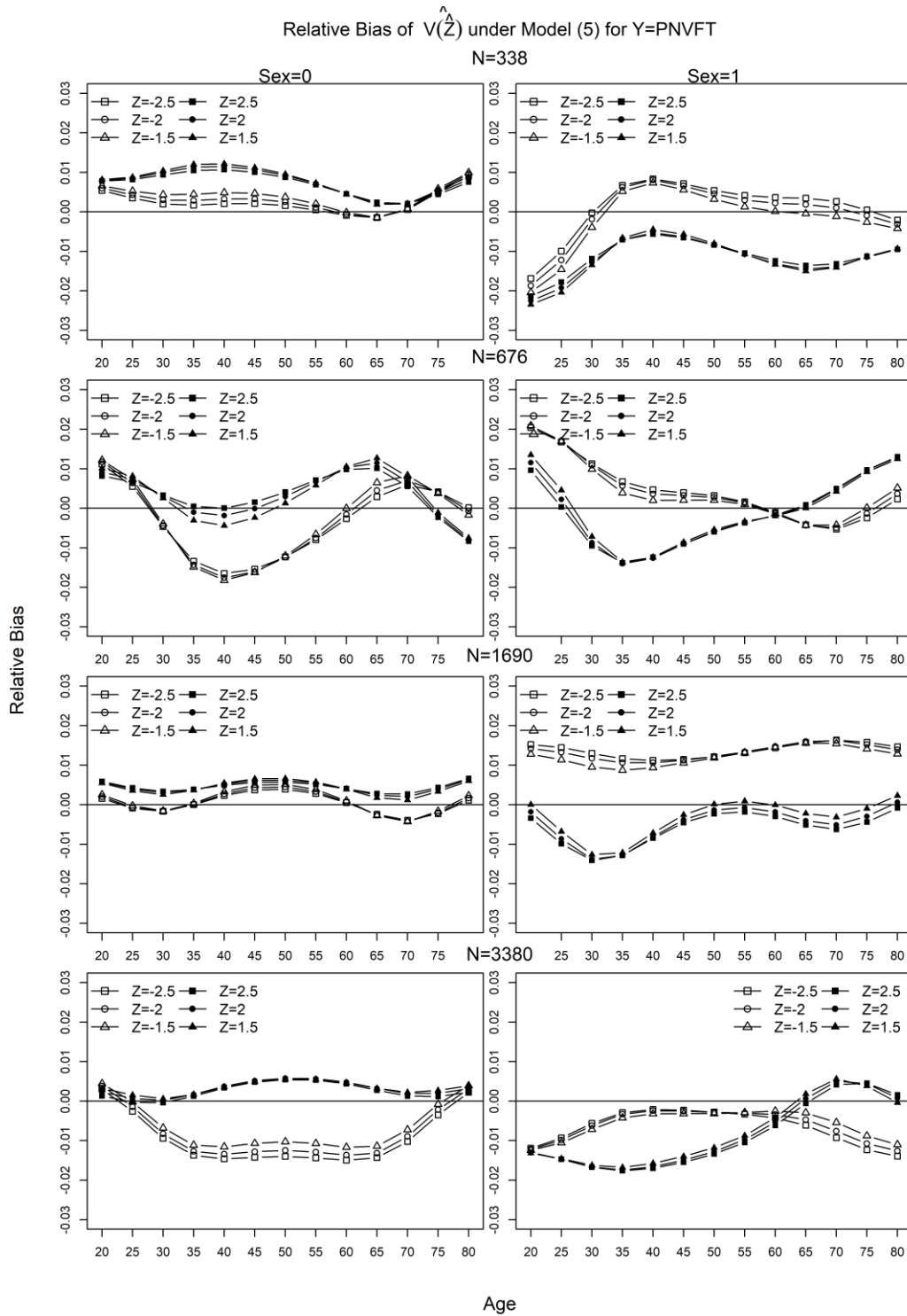


Figure S.A.13

Coverage of the 95% Confidence Interval for Z_0 Computed with Equation (9), $\hat{V}(\hat{Z})$, Under Model (1) for $Y = \text{PNVFT}$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $Z_0 \in \{\pm 1.5, \pm 2, \pm 2.5\}$ (curves)

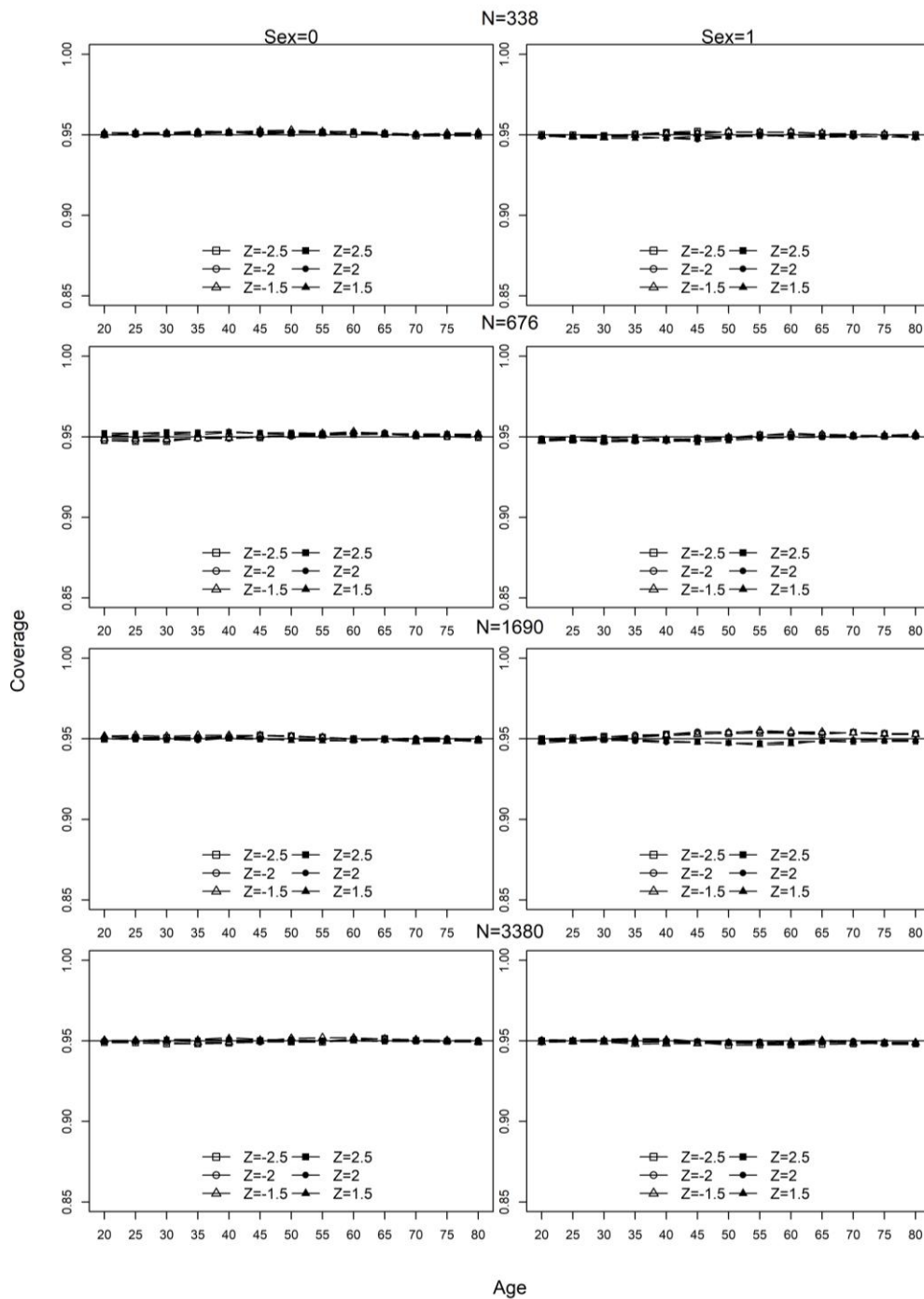


Figure S.A.14

Coverage of the 95% Confidence Interval for Z_0 Computed with Equation (9), $\hat{V}(\hat{Z})$, Under Model (5) for $Y = \text{PNVFT}$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $Z_0 \in \{\pm 1.5, \pm 2, \pm 2.5\}$ (curves)

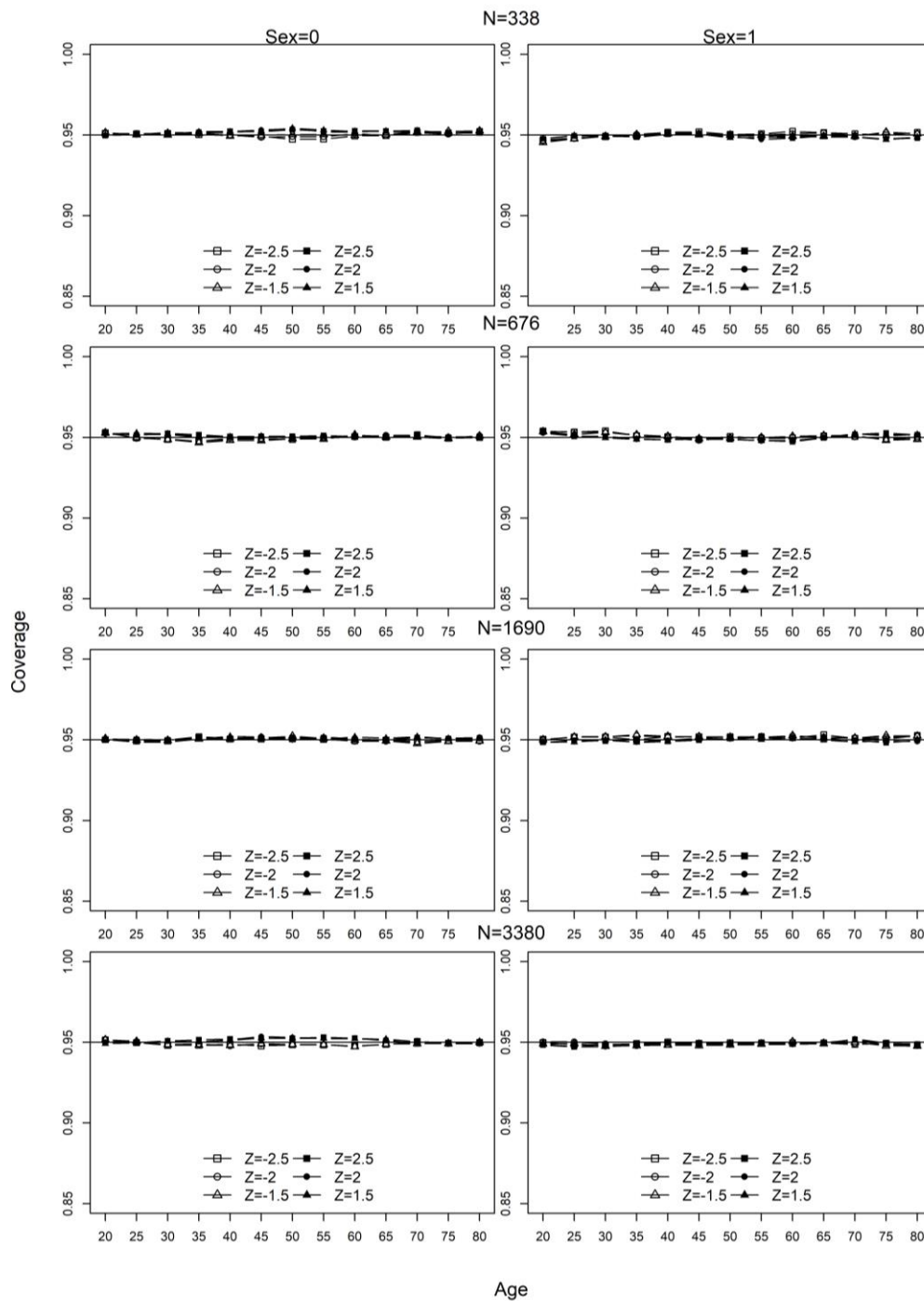
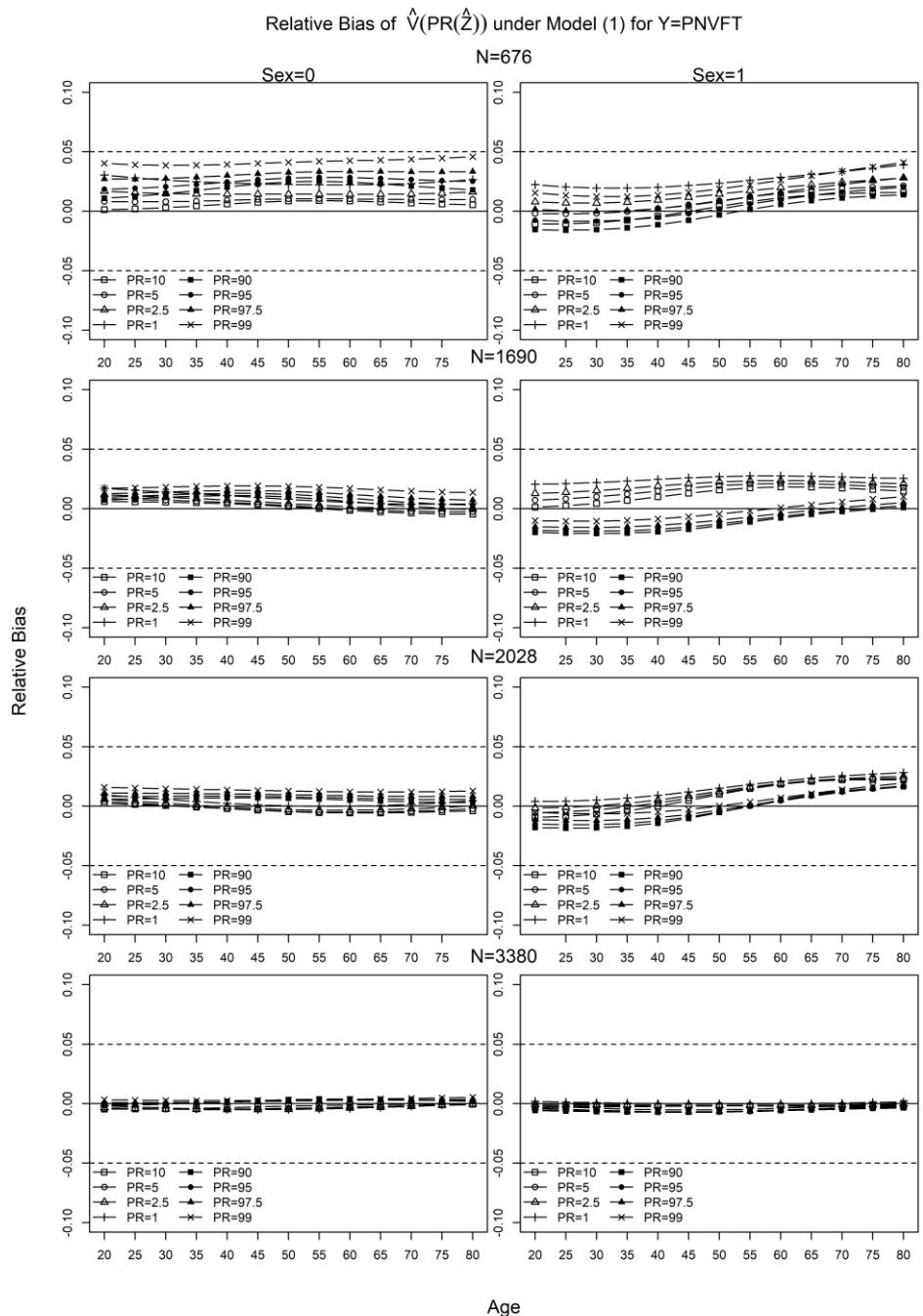


Figure S.A.15

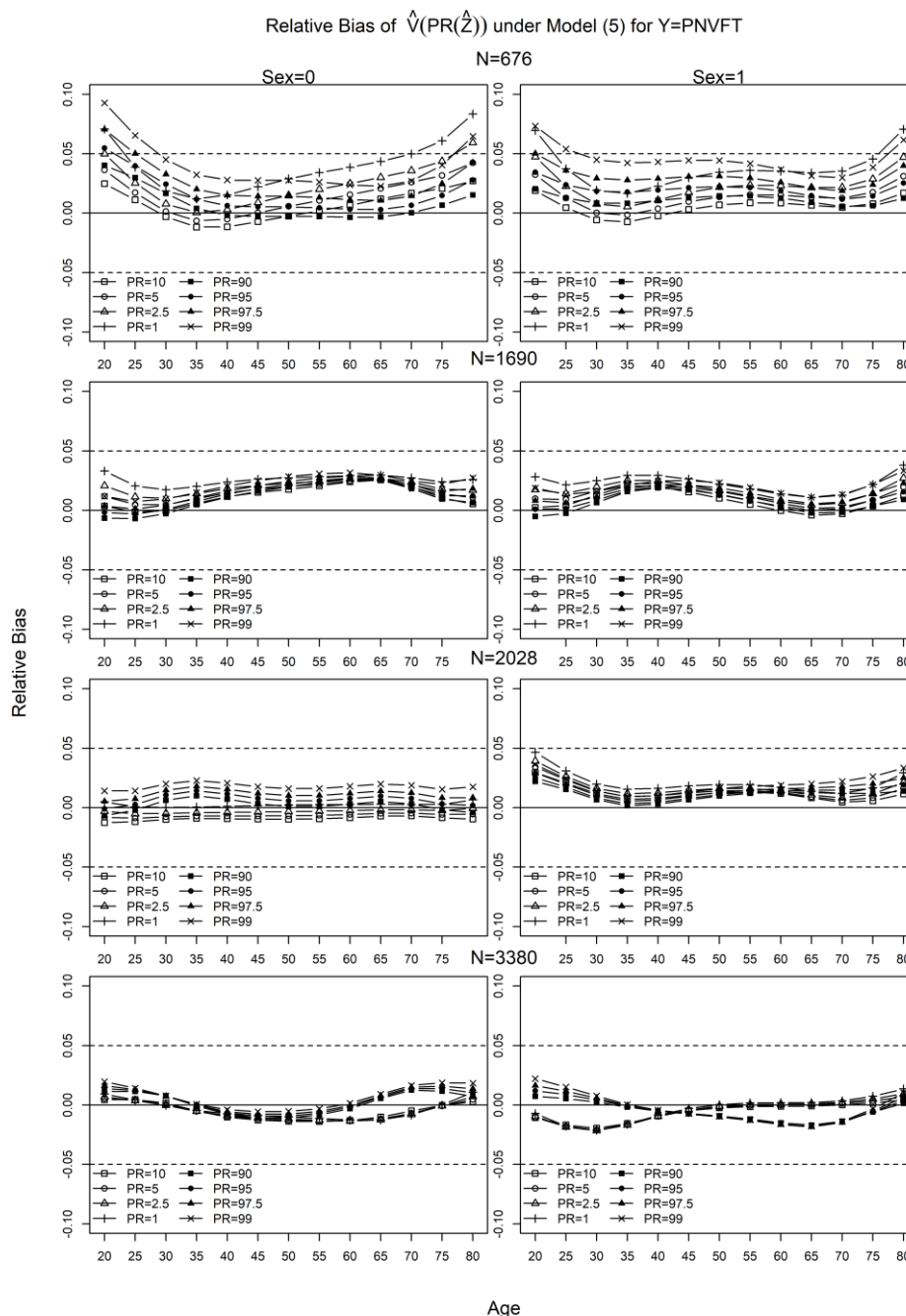
Relative Bias of Equation (10), $\hat{V}(PR(\hat{Z}_0))$, Under Model (1) for $Y = PNVFT$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $PR_0 \in \{1,2.5,5,10,90,95,97.5,99\}$ (curves)



Note. The relative bias can be considered as acceptable within the range $[-5\%, 5\%]$ (horizontal dotted lines).

Figure S.A.16

Relative Bias of Equation (10), $\hat{V}(PR(\hat{Z}_0))$, Under Model (5) for $Y = PNVFT$, as a Function of Age $\in [20,80]$ (x-axis) and Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{338,676,1690,3380\}$ (rows), and $PR_0 \in \{1,2.5,5,10,90,95,97.5,99\}$ (curves)



Note. The relative bias can be considered as acceptable within the range $[-5\%, 5\%]$ (horizontal dotted lines).

Figure S.A.17

Coverage of the 95% Confidence Interval for PR_0 Computed with Equation (10),

$\hat{V}(PR(\hat{Z}_0))$, Under Model (1) for $Y = PNVFT$, as a Function of Age $\in [20,80]$ (x-axis) and

Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{676,1690,2028,3380\}$ (rows), and

$PR_0 \in \{1,2.5,5,10,90,95,97.5,99\}$ (curves)

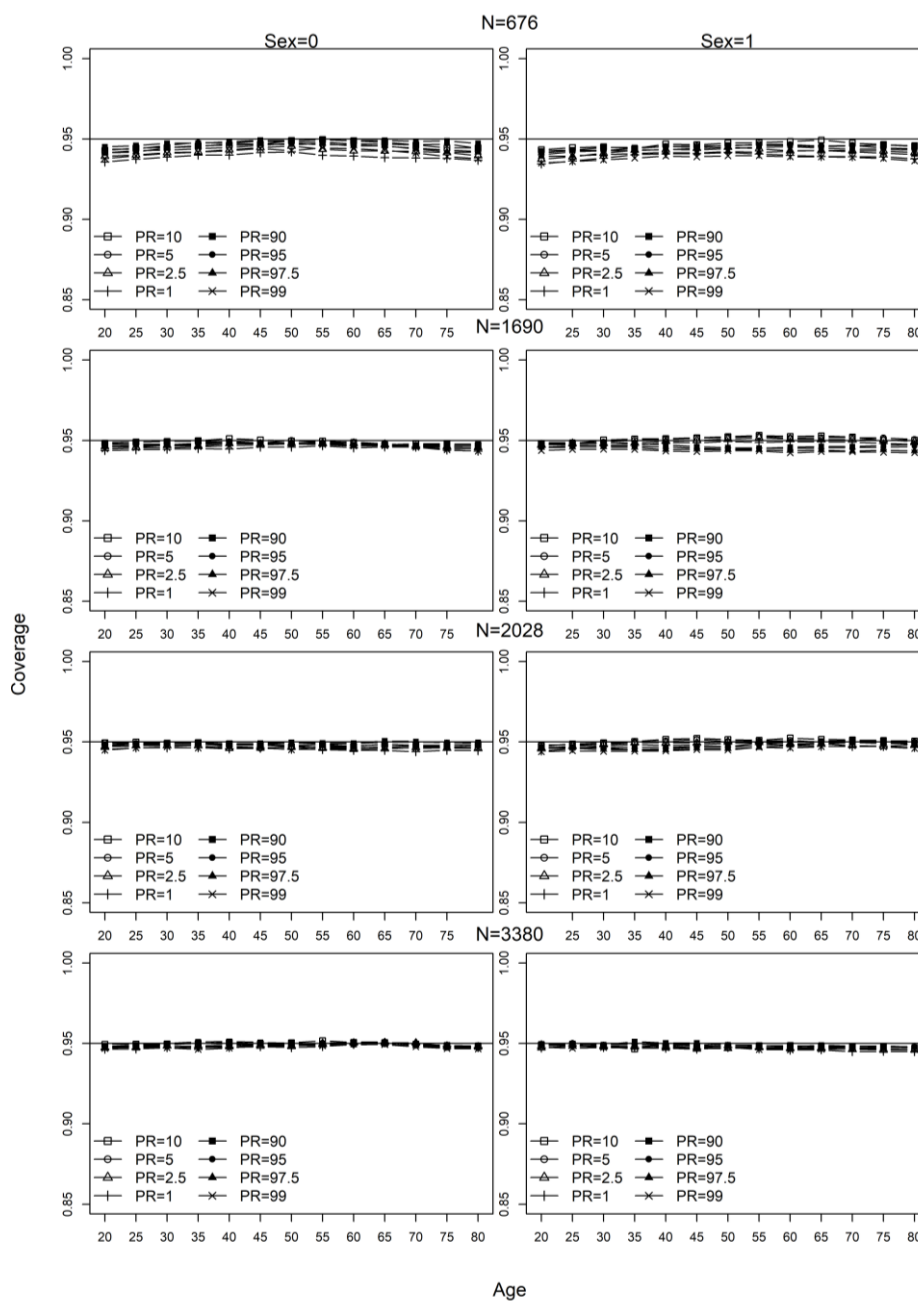


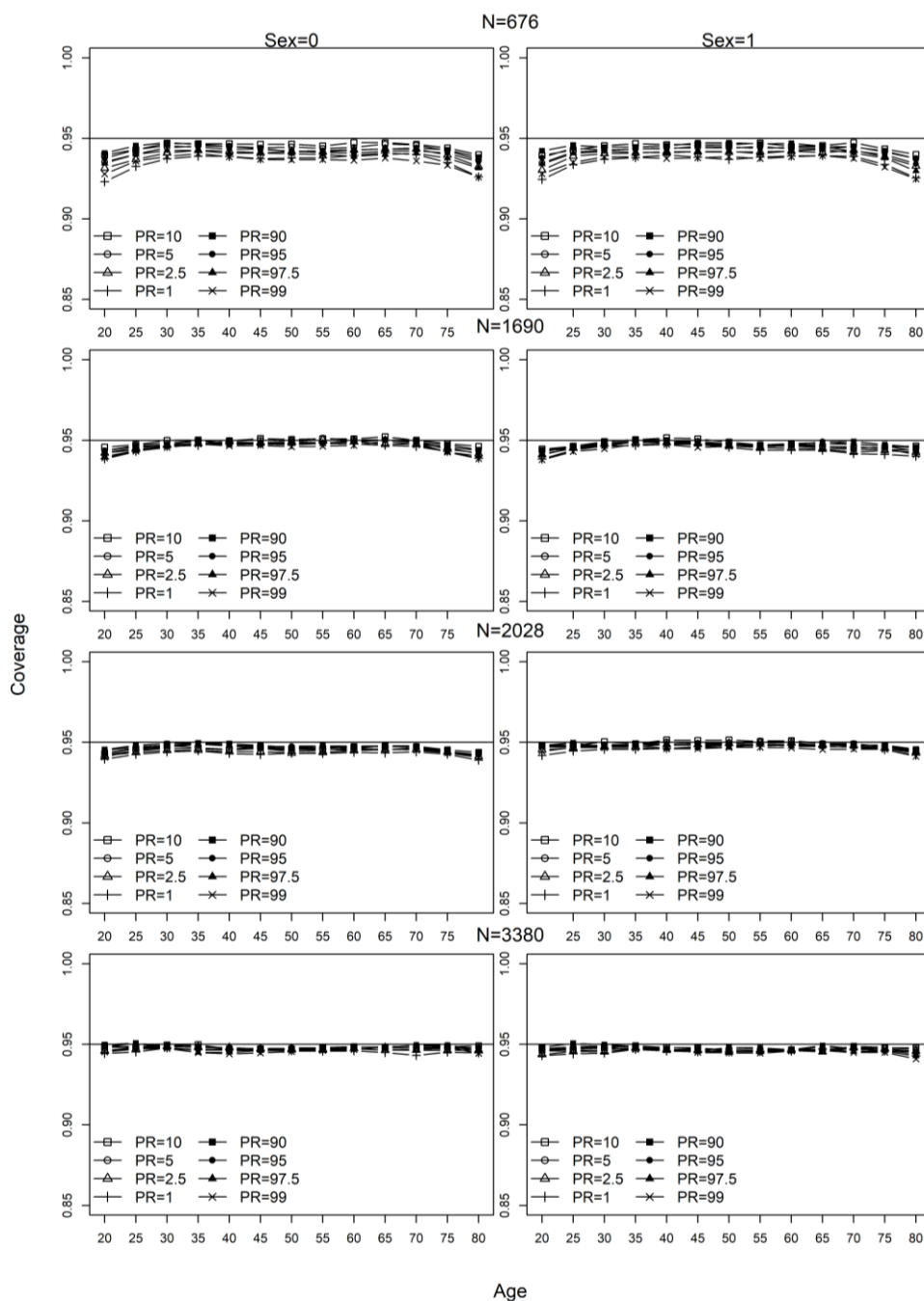
Figure S.A.18

Coverage of the 95% Confidence Interval for PR_0 Computed with Equation (10),

$\hat{V}(PR(\hat{Z}_0))$, Under Model (5) for $Y = PNVFT$, as a Function of Age $\in [20,80]$ (x-axis) and

Sex $\in \{0,1\}$ (columns), for Different Sample Sizes $N \in \{676,1690,2028,3380\}$ (rows), and

$PR_0 \in \{1,2.5,5,10,90,95,97.5,99\}$ (curves)



Outline of the R Codes for the Simulation Studies

Steps of the R Code for the First Simulation Study

The bias of equations (7) and (8) is assessed with the following procedure that must be repeated for each model considered in the main text, and set of parameter $(\boldsymbol{\beta}, \sigma_\varepsilon)$ values in Table S.A.2.

0. Choose a set of M Z-score values that are relevant in practice. Denote by \mathbf{z}_0 the $M \times 1$ vector of these relevant Z-scores. Compute $\boldsymbol{\varepsilon}_0 = \sigma_\varepsilon \mathbf{z}_0$ and $\mathbf{PR}(\mathbf{z}_0) = \boldsymbol{\Phi}(\mathbf{z}_0) \times 100$. Define all possible combinations of the levels of $X_1 = \text{age}$ and $X_2 = \text{sex}$. Denote by Q the number of all possible combinations for $X_1 = \text{age}$ and $X_2 = \text{sex} \in \{0,1\}$. Center X_1 (i.e. $X_1 - \bar{X}_1$) prior to the computation of the quadratic effect of X_1 and the interaction terms. Generate $\mathbf{X}_{N \times (k+1)}$, that is, the design matrix of the S normative samples, by replicating L times each combination of $Z_0 \in \mathbf{z}$, X_1 , and X_2 , so that the sample size of the normative sample is $N = L \times M \times Q$. In this way the size of the normative sample is not fully determined by the number of combinations of Z_0 , X_1 , and X_2 (i.e. $M \times Q$).
1. Individuals to whom the norms will be applied (i.e. not members of the normative sample). Each individual ($j = 1, \dots, M \times Q$), to whom the norms will be applied, corresponds to a different combination of Z_0 , X_1 , and X_2 . For each individual j (i.e. for each combination of Z_0 , X_1 , and X_2), compute the observed raw score y_{0j} by filling $\boldsymbol{\beta}$, $\boldsymbol{\varepsilon}_{0j} = \sigma_\varepsilon Z_{0j}$, $\mathbf{x}'_{0j} = [1 \ X_{1,0j} \ X_{2,0j}]$ in the chosen model. This gives the $(M \times Q) \times 1$ vector \mathbf{y}_0 (these $M \times Q$ raw scores will be translated into estimated Z-scores and PR-scores once $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_\varepsilon$ are obtained from the normative sample).
2. For each individual j from step 1 (i.e. for each combination of Z_0 , X_1 , and X_2 , with $j = 1, \dots, M \times Q$), compute equations (7) and (8) by filling \mathbf{x}_0' , $\mathbf{X}_{N \times (k+1)}$, Z_{0j} , N , and k , in

these formulae. Denote the variances computed with equations (7) and (8) by $V(\hat{Z}_{0j})_{D.M.}$

and $V\left(PR(\hat{Z}_{0j})\right)_{D.M.}$, where ‘‘D.M’’ means ‘‘Delta method’’.

3. Generate S normative samples. To create a normative sample ($s = 1, \dots, S$, with $S = 20,000$), sample a new $N \times 1$ vector $\boldsymbol{\varepsilon}^{(s)} \sim N(0, \sigma_{\varepsilon}^2)$. Compute the $N \times 1$ vector $\mathbf{y}^{(s)}$ from the chosen model, using $\boldsymbol{\beta}$, $\mathbf{X}_{N \times (k+1)}$, and $\boldsymbol{\varepsilon}^{(s)}$.

4. For each normative sample ($s = 1, \dots, S$), fit the model, obtaining then $\hat{\boldsymbol{\beta}}^{(s)} =$

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}^{(s)}, \hat{\boldsymbol{\varepsilon}}^{(s)} = \mathbf{y}^{(s)} - \mathbf{X}\hat{\boldsymbol{\beta}}^{(s)}, \text{ and } \hat{\sigma}_{\varepsilon}^{(s)} = \sqrt{\frac{\hat{\boldsymbol{\varepsilon}}^{(s)'}\hat{\boldsymbol{\varepsilon}}^{(s)}}{N-k-1}}.$$

5. For each individual j from step 1 (i.e. for each combination of Z_0 , X_1 , and X_2 , with $j = 1, \dots, M \times Q$), estimate the Z_0 score, based on the s -th normative sample, with $\hat{Z}_{0j}^{(s)} =$

$$\frac{Y_{0j} - \hat{Y}_j^{(s)}}{\hat{\sigma}_{\varepsilon}^{(s)}} = \frac{Y_{0j} - \mathbf{x}'_{0j}\hat{\boldsymbol{\beta}}^{(s)}}{\hat{\sigma}_{\varepsilon}^{(s)}}, \text{ and the } PR \text{ score with } PR(\hat{Z}_{0j})^{(s)} = \Phi(\hat{Z}_{0j}^{(s)}) \times 100.$$

6. For each individual j from step 1 (i.e. for each combination of Z_0 , X_1 , and X_2 , with $j = 1, \dots, M \times Q$), compute $\bar{Z}_{0j} = \sum_{s=1}^S \hat{Z}_{0j}^{(s)} / S$, $\overline{PR(\hat{Z}_{0j})} = \sum_{s=1}^S PR(\hat{Z}_{0j})^{(s)} / S$, $V(\hat{Z}_{0j}) =$

$$\frac{\sum_{s=1}^S (\hat{Z}_{0j}^{(s)} - \bar{Z}_{0j})^2}{S-1}, \text{ and } V\left(PR(\hat{Z}_{0j})\right) = \frac{\sum_{s=1}^S (PR(\hat{Z}_{0j})^{(s)} - \overline{PR(\hat{Z}_{0j})})^2}{S-1}.$$

7. For each individual j from step 1 (i.e. for each combination of Z_0 , X_1 , and X_2 , with $j = 1, \dots, M \times Q$), compute the relative biases of equations (7) and (8), that is,

$$R.B. \text{ (eq. (7))} = \frac{V(\hat{Z}_{0j})_{D.M.} - V(Z_{0j})}{V(\hat{Z}_{0j})} \text{ and } R.B. \text{ (eq. (8))} = \frac{V(PR(\hat{Z}_{0j}))_{D.M.} - V(PR(Z_{0j}))}{V(PR(\hat{Z}_{0j}))},$$

respectively, where $V(\hat{Z}_{0j})_{D.M.}$ and $V\left(PR(\hat{Z}_{0j})\right)_{D.M.}$ have been computed in step 2.

Furthermore, compute the absolute biases of \hat{Z}_0 and of $PR(\hat{Z}_{0j})$, that is, $\bar{Z}_{0j} - Z_{0j}$ and $\overline{PR(\hat{Z}_{0j})} - PR(Z_{0j})$.

Steps of the R Code for the Second Simulation Study

The simulation procedure is the same as that in the previous section apart from the following steps.

5. For each individual j from step 1 (i.e. for each combination of Z_0 , X_1 , and X_2 , with $j =$

$1, \dots, M \times Q$), estimate the Z_0 score, based on the s -th normative sample, with $\hat{Z}_{0j}^{(s)} =$

$$\frac{Y_{0j} - \hat{Y}_j^{(s)}}{\hat{\sigma}_\varepsilon^{(s)}} = \frac{Y_{0j} - \mathbf{x}'_{0j} \hat{\boldsymbol{\beta}}^{(s)}}{\hat{\sigma}_\varepsilon^{(s)}}. \text{ Compute equations (9) and (10) by filling } \mathbf{x}_0', \mathbf{X}_{N \times (k+1)}, \hat{Z}_{0j}^{(s)}, N,$$

and k , in these formulas. Denote the variances computed with equations (9) and (10) by

$$\hat{V}(\hat{Z}_{0j})_{D.M.}^{(s)} \text{ and } \hat{V}(PR(\hat{Z}_{0j}))_{D.M.}^{(s)}, \text{ respectively.}$$

6. For each individual j from step 1 (i.e. for each combination of Z_0 , X_1 , and X_2 , with $j =$

$1, \dots, M \times Q$), compute the Monte Carlo averages $\overline{\hat{V}(\hat{Z}_{0j})}_{D.M.} = \sum_{s=1}^S \hat{V}(\hat{Z}_{0j})_{D.M.}^{(s)} / S$ and

$$\overline{\hat{V}(PR(\hat{Z}_{0j}))}_{D.M.} = \sum_{s=1}^S \hat{V}(PR(\hat{Z}_{0j}))_{D.M.}^{(s)} / S, \text{ and the true variances } V(\hat{Z}_{0j}) =$$

$$\frac{\sum_{s=1}^S (\hat{Z}_{0j}^{(s)} - \bar{\hat{Z}}_{0j})^2}{S-1} \text{ and } V(PR(\hat{Z}_{0j})) = \frac{\sum_{s=1}^S (PR(\hat{Z}_{0j})^{(s)} - \overline{PR(\hat{Z}_{0j})})^2}{S-1}.$$

7. For each individual j from step 1 (i.e. for each combination of Z_0 , X_1 , and X_2 , with $j =$

$1, \dots, M \times Q$), compute the relative biases of equations (9) and (10), that is,

$$R.B. \text{ (eq. (9))} = \frac{\overline{\hat{V}(\hat{Z}_{0j})}_{D.M.} - V(\hat{Z}_{0j})}{V(\hat{Z}_{0j})} \text{ and } R.B. \text{ (eq. (10))} = \frac{\overline{\hat{V}(PR(\hat{Z}_{0j}))}_{D.M.} - V(PR(\hat{Z}_{0j}))}{V(PR(\hat{Z}_{0j}))},$$

respectively, as well as the coverage of 95% confidence intervals for Z_{0j} and $PR(Z_{0j})$,

that is, the number of times $Z_{0j} \in \left[\hat{Z}_{0j} \pm 1.96 \sqrt{\hat{V}(\hat{Z}_{0j})_{D.M.}^{(s)}} \right]$ and $PR(Z_{0j}) \in$

$$\left[PR(\hat{Z}_{0j}) \pm 1.96 \sqrt{\hat{V}(PR(\hat{Z}_{0j}))_{D.M.}^{(s)}} \right] \text{ divided by } S.$$

Optimal and Robust Design

Derivation of the D-Optimal Designs in Table 2 (main text).

The derivation of the D-optimal designs in Table 2 (main text) depends on two sets of assumptions. First, normality and homoscedasticity of the residual ε must hold. Second, for each predictor the range of possible values (e.g. age $\in [20, 80]$) and its regression function (e.g. quadratic regression for age) do not depend on other predictors (Schwabe, 1996, p. 36). In other words, the range of values and the model for age (e.g. quadratic regression) are the same for men and women under the model assumed to plan the study.

The derivation of the D-optimal designs in Table 2 (main text) can be summarized in the following steps:

1. **Find the Marginal Models.** The multiple regression model (Table 2, leftmost column, main text) is decomposed into as many components as there are predictors. Specifically, for each predictor the corresponding component is obtained by deleting from the full model (Table 2, leftmost column, main text) the other predictors, all interactions, and the intercept if already included in another component (see Example 6.4, p. 83, in Schwabe, 1996). Each component can be then rewritten as a single predictor model for Y and, for this reason, is called the *marginal model* for that predictor (see Table 2, second column, main text). For example, deleting $\beta_2 X_2$ and interaction term $\beta_4 X_1 X_2$ from model (3) (i.e. from $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_1 X_2$), yields the marginal model for age which is $E_1(Y) = \beta_0 + \beta_1 X_1$ and, consequently, the marginal model for sex is $E_2(Y) = \beta_2 X_2$.
2. **Derive the Optimal Designs for the Marginal Models.** For models (1)-(3) and (5), the D-optimal designs for the marginal models for X_1 (Table 2, third column, main

text) are taken from Table 3.5, p. 67, in Berger and Wong (2009). For model (4), the optimal design for the marginal model for X_1 (Table 2, third column, main text) is taken from Table 6.1, p. 94, in Schwabe (1996). This latter design is not D-optimal but is obtained by optimizing a particular criterion (i.e. the determinant of the information matrix for the marginal model of the quantitative predictor times a function of the categorical predictor) such that the design for the full model is D-optimal (this criterion is given in Corollary 6.14.i, p. 91, in Schwabe, 1996). The D-optimal design for the marginal model for X_2 (Table 2, fourth column, main text) is taken from p. 24 in Schwabe (1996).

3. **Construct the D-Optimal Design for the Full Model.** For models (1)-(5), the D-optimal design ξ^* (Table 2, rightmost column, main text) is obtained from the optimal designs of the marginal models in step 2 as follows. The support points of ξ^* are all possible combinations of support points of the designs from step 2, and the corresponding design weights of ξ^* are the product of the design weights of the optimal designs from step 2.

One could verify that the designs obtained in step 3 are D-optimal by making use of the equivalence theorem (Atkinson et al., 2007, p. 122), which states that a design ξ is D-optimal if and only if $d(\mathbf{X}, \xi) = (k + 1)$ at its support points and $d(\mathbf{X}, \xi) < (k + 1)$ over the rest of the design region (see Figure 2, main text). Schwabe (1996) has analytically proven that designs obtained from steps 1-3 are D-optimal. Specifically, Schwabe (1996) has shown that the D-optimal design of a model with either complete or no interaction is given by the Kronecker product of the D-optimal designs for the marginal models (Theorem 4.2, p. 41, and Theorem 5.2, p. 52, in Schwabe, 1996). This result still holds when $k \geq 2$ predictors are included into these two types of models (Corollary 4.6, p. 44, and Corollary 5.3, p.53, in Schwabe, 1996), which makes the extension to the inclusion of education in these types of

models straightforward. Furthermore, Schwabe (1996) has shown that the Kronecker product of the optimal designs for the marginal models is the D-optimal design for models with partial interaction with two predictors of which one is categorical (Corollary 6.14, p. 91, in Schwabe, 1996). However, in this case the designs for the marginal models might not be D-optimal. For the categorical predictor with Q_2 levels (e.g. sex), the optimal design for the marginal model is the design that gives equal weight $\frac{1}{Q_2}$ to each level of the predictor (Schwabe, 1996, p. 83), which is D-optimal (Schwabe, 1996, Corollary 3.16.i, p. 24). In contrast, the optimal design for the marginal model for the other predictor (e.g. age) is not D-optimal, because it is obtained by maximizing a different optimality criterion given in Corollary 6.14.i, p. 91, in Schwabe (1996). For models with partial interaction as in model (4) where age, but not age², is assumed to interact with sex, extension to $k \geq 2$ predictors is complicated and beyond the scope of this paper.

Robust Design

Maximin Design Based on the Relative Efficiency Criterion

As explained in the main text, the optimal design depends on the assumed model. In presence of uncertainty about the “true” model, a robust alternative to the optimal design is the maximin design. Based on the relative efficiency criterion, the maximin design is defined as that design that maximizes the minimum relative efficiency. The relative efficiency is defined as the ratio of $V(\hat{Z}_0)$ or $V(PR(\hat{Z}_0))$ (i.e. equations (7) and (8) in the main text) under the optimal design ξ^* to $V(\hat{Z}_0)$ or $V(PR(\hat{Z}_0))$ under the non-optimal design ξ . This variance ratio can be shown to reduce to equation (11) in the main text:

$$RE(\xi \text{ vs } \xi^*) = \frac{V(PR(\hat{Z}_0)|\xi^*)}{V(PR(\hat{Z}_0)|\xi)} = \frac{100^2 \phi(Z_0)^2 \left[\frac{d(X, \xi^*)}{N} + \frac{1}{2(N-k-1)} Z_0^2 \right]}{100^2 \phi(Z_0)^2 \left[\frac{d(X, \xi)}{N} + \frac{1}{2(N-k-1)} Z_0^2 \right]} =$$

$$= \frac{V(\hat{Z}_0|\xi^*)}{V(\hat{Z}_0|\xi)} = \frac{\frac{d(\mathbf{X},\xi^*) + \frac{1}{2(N-k-1)}Z_0^2}{N}}{\frac{d(\mathbf{X},\xi) + \frac{1}{2(N-k-1)}Z_0^2}{N}} \approx \frac{d(\mathbf{X},\xi^*) + \frac{Z_0^2}{2}}{d(\mathbf{X},\xi) + \frac{Z_0^2}{2}},$$

where $d(\mathbf{X}, \xi) = N\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$ is the standardized prediction variance under design ξ , and the approximation follows from $N - k - 1 \approx N$. Recall from the Simulation Studies section that equation (7) (main text) is an accurate approximation of the true $V(\hat{Z}_0)$ for $N \geq 338$, and equation (8) (main text) is an accurate approximation of the true $V\left(PR(\hat{Z}_0)\right)$ for $N \geq 1690$. For these lower-bounds for the sample size $N - k - 1 \approx N$, yielding equation (11) in the main text.

Note that $RE(\xi \text{ vs } \xi^*)$ (i.e. equation (11)) is a function of the design of the normative sample (through \mathbf{X} in $d(\mathbf{X}, \xi)$), and of the individual's \mathbf{x}_0 and Z . Hence, the maximin design is obtained by first finding, for each design, the lowest value of equation (11) over \mathbf{x}_0 for a given Z_0 , for each model, and then choosing the design with the highest minimum relative efficiency across all models. For the designs and models in Table 3 (main text), a numerical evaluation of equation (11) for several values of \mathbf{x}_0 and $Z_0 \in \{\pm 3, \pm 2.5, \pm 2, \pm 1.5, \pm 1, \pm 0.5, 0\}$, has shown that the maximin design does not depend on Z_0 , because in all considered scenarios the design in the second row of Table 3 (main text) was the design with the highest minimum relative efficiency across all models. In Table 3 (main text), the lowest value of equation (11) over \mathbf{x}_0 are shown only for $Z_0 = 0$ and $Z_0 = \pm 2$, because $Z_0 = 0$ yields the maximum efficiency loss for ξ , and $Z_0 = \pm 2$ is often used in practice as a cut-off to identify extreme performance or symptoms. Both for $Z_0 = 0$ and $Z_0 = \pm 2$, the minimum of equation (11) was obtained for an \mathbf{x}_0 that is a support point of the optimal design under the “true” model (e.g. $\text{age} \in \{-1, 0, +1\}$ and $\text{sex} = 0/1$). The R code of the numerical evaluation is given in online supplement B.

Maximin Design Based on the Efficiency Criterion

Based on the efficiency criterion, the maximin design is defined as that design that maximizes the minimum efficiency, that is, that design that minimizes the maximum sampling variance (i.e. equations (7) and (8) in the main text). Note that in equations (7) and (8) (main text) the only component that depends on the normative sample is $\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 = \frac{d(\mathbf{X}, \xi)}{N}$, where $d(\mathbf{X}, \xi)$ is the standardized prediction variance. Hence, maximizing $d(\mathbf{X}, \xi)$ maximizes equations (7) and (8), given N and Z_0 . Therefore, to find the maximin design based on the efficiency criterion, Table S.A.3 shows the maximum $d(\mathbf{X}, \xi)$ for each design (row) under the “true” model (column). As can be seen in Table S.A.3, model (5) is the model that yields the maximum $d(\mathbf{X}, \xi)$ for all designs (except the two age levels design under which the quadratic age effect is not identifiable), and the optimal design for model (5) (i.e. Table S.A.3, second row) is the maximin design, because it yields the smallest maximum $d(\mathbf{X}, \xi)$ (and thereby the smallest maximum of equations (7) and (8)). The R code of the numerical evaluation to obtain the results in Table S.A.3 is given in online supplement B.

Table S.A.3

Maximum Standardized Prediction Variance $d(\mathbf{X}, \xi) = N\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$ for Each Design

(row) Under the “True” Model (column)

Design	“True” model				
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
2 Age levels. Equal weight $w^* = \frac{1}{4}$ to -1 and 1 of $X_1 = \text{Age}$, for each level of $X_2 = \text{Sex}$	3		4		
3 Age levels. Equal weight $w^* = \frac{1}{6}$ to $-1, 0$, and 1 of $X_1 = \text{Age}$, for each level of $X_2 = \text{Sex}$	3.5	4	5	5.5	6
3 Age levels. Equal weight $\frac{3}{16}$ to -1 and 1 of $X_1 = \text{Age}$, and weight $\frac{1}{8}$ to 0 of $X_1 = \text{Age}$, for each level of $X_2 = \text{Sex}$	3.33	5	4.67	5	8
4 Age levels. Equal weight $w^* = \frac{1}{8}$ to $-1, -0.333, 0.333$ and 1 of $X_1 = \text{Age}$, for each level of $X_2 = \text{Sex}$	3.8	4.8	5.6	6.6	7.6
5 Age levels. Equal weight $w^* = \frac{1}{10}$ to $-1, -0.5, 0, 0.5$ and 1 of $X_1 = \text{Age}$, for each level of $X_2 = \text{Sex}$	4	5.43	6	7.43	8.86
6 Age levels. Equal weight $w^* = \frac{1}{12}$ to $-1, -0.6, -0.2, 0.2, 0.6$ and 1 of $X_1 = \text{Age}$, for each level of $X_2 = \text{Sex}$	4.14	5.93	6.29	8.07	9.86
13 Age levels. Equal weight $w^* = \frac{1}{26}$ to $-1, -0.833, -0.667, -0.5, -0.333, -0.167, 0, 0.167, 0.333, 0.5, 0.667, 0.833$, and 1 of $X_1 = \text{Age}$, for each level of $X_2 = \text{Sex}$	4.57	7.71	7.14	10.29	13.43

Note. Minimum efficiency (i.e. maximum $d(\mathbf{X}, \xi)$) per design in boldface. Recall that

$\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$ is the component of equations (7) and (8) that depends on the normative sample.

References

- Atkinson, A. C., Donev, A. N., & Tobias, R. D. (2007). *Optimum experimental designs, with SAS*. Oxford University Press.
- Berger, M. P. F., & Wong, W. K. (2009). *An introduction to optimal designs for social and biomedical research*. John Wiley & Sons.
- Goretti, B., Niccolai, C., Hakiki, B., Sturchio, A., Falautano, M., Minacapelli, E., Martinelli, V., Incerti, C., Nocentini, U., Murgia, M., Fenu, G., Cocco, E., Marrosu, M. G., Garofalo, E., Ambra, F. I., Maddestra, M., Consalvo, M., Viterbo, R. G., Trojano, M., ... Amato, M. P. (2014). The brief international cognitive assessment for multiple sclerosis (BICAMS): Normative values with gender, age and education corrections in the Italian population. *BMC Neurology*, *14*, 171-176. <https://doi.org/10.1186/s12883-014-0171-6>
- Johnson, R. A., & Wichern, D. W. (1998). *Applied multivariate statistical analysis* (4th ed.). Pearson Prentice Hall.
- Oosterhuis, H. E. M., Van der Ark, L. A., & Sijtsma, K. (2016). Sample size requirements for traditional and regression-based norms. *Assessment*, *23*(2), 191-202. <https://doi.org/10.1177/1073191115580638>
- Parmenter, B. A., Testa, S. M., Schretlen, D. J., Weinstock-Guttman, B., & Benedict, R. H. B. (2010). The utility of regression-based norms in interpreting the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). *Journal of the International Neuropsychological Society*, *16*, 6-16. <https://doi.org/10.1017/S1355617709990750>
- Schwabe, R. (1996). *Optimum designs for multi-factor models*. Springer-Verlag.

Van der Elst, W., Van Boxtel, M. P. J., Van Breukelen, G. J. P., & Jolles, J. (2006).

Normative data for the animal, profession and letter *M* naming verbal fluency tests for Dutch speaking participants and the effects of age, education, and sex. *Journal of the International Neuropsychological Society*, 12, 80-89.

<https://doi.org/10.1017/S1355617706060115>