Article

A Simulation Study of Categorizing Continuous Exposure Variables Measured with Error in Autism Research: Small Changes with Large Effects

Karyn Heavner [1,*] and Igor Burstyn [1,2,3]

[1] Department of Environmental and Occupational Health, School of Public Health,
    Drexel University, Philadelphia, PA 19104, USA; E-Mail: Igor.Burstyn@drexel.edu

[2] A.J. Drexel Autism Institute, School of Public Health, Drexel University, Philadelphia,
    PA 19104, USA

[3] Department of Epidemiology and Biostatistics, School of Public Health, Drexel University,
    Philadelphia, PA 19104, USA

* Author to whom correspondence should be addressed; E-Mail: karynkh@aol.com;
    Tel.: +1-413-559-0160.

Abstract: Variation in the odds ratio (OR) resulting from selection of cutoffs for categorizing continuous variables is rarely discussed. We present results for the effect of varying cutoffs used to categorize a mismeasured exposure in a simulated population in the context of autism spectrum disorders research. Simulated cohorts were created with three distinct exposure-outcome curves and three measurement error variances for the exposure. ORs were calculated using logistic regression for 61 cutoffs (mean ± 3 standard deviations) used to dichotomize the observed exposure. ORs were calculated for five categories with a wide range for the cutoffs. For each scenario and cutoff, the OR, sensitivity, and specificity were calculated. The three exposure-outcome relationships had distinctly shaped OR (*versus* cutoff) curves, but increasing measurement error obscured the shape. At extreme cutoffs, there was non-monotonic oscillation in the ORs that cannot be attributed to "small numbers." Exposure misclassification following categorization of the mismeasured exposure was differential, as predicted by theory. Sensitivity was higher among cases and specificity among controls. Cutoffs chosen for categorizing continuous variables can have profound

effects on study results. When measurement error is not too great, the shape of the OR curve may provide insight into the true shape of the exposure-disease relationship.

Keywords: categorization; autism spectrum disorders; epidemiology methods; misclassification; dichotomization

---

1. Introduction

The cutoff used to dichotomize a continuous exposure variable may have a profound effect on the measures of association and interpretation of study results [1,2]. It is well-understood that, as the threshold for definition of "exposure" changes, the magnitude of the effect estimates, such as odds ratio (OR), will vary as well even though the true relative risk per unit of exposure is constant. For example, if one studies the association between obesity measured as body-mass index (BMI) and myocardial infarction, and categorizes BMI, then if (a) we set a threshold at BMI = 25 then we compare "underweight" and "overweight", but if we (b) set a threshold at BMI = 30, then we compare "obese" patients with the rest. It is reasonable to expect different estimates of risk in the two analytical approaches even though they arise from the same underlying association of obesity and myocardial infarction. Frequently, data are analyzed using multiple cutoffs but only one of these categorizations is reported, creating publication bias *in situ* [3]. Presenting results for a wide range of plausible cutoffs, by plotting effect estimates against cutoffs, minimizes the potential for fishing for particular results, makes study results more useful to readers, and avoids the pitfalls of using data driven cutoffs or theory or guidance driven cutoffs that may change [4]. Analyzing continuous exposure data using more than one cutoff (two categories) magnifies the problem of selecting cutoffs as many different combinations of cutoffs may be plausible and estimates become increasingly unstable in finite datasets as the number of categories increases [5]. However, it is routinely done, with some justification, to explore the shapes of exposure-response associations, although even this has been criticized, with alternatives involving non-parametric techniques being suggested [6,7]. In addition, if more than two categories are analyzed, the cutoffs are frequently data driven, which has its own limitations.

Environmental exposures and gestational age are examples of exposures that are inherently continuous but are frequently categorized and for which the recommended cutoffs tend to change over time or are under debate. For example, environmental regulations tend to lower the exposure threshold at which a change in risk is of interest, rendering studies presented with a limited range of cutoffs (perhaps based on exposure limits in existence at the time of the study) less useful for future risk assessment or meta-analysis. Likewise, the clinical thresholds for shortened gestation and "term" pregnancy have evolved over time [8,9] and underscore the importance of investigating more than one cutoff for gestational age. Our previous research presented estimates of effect for a wide range of cutoffs used to dichotomize the exposure of interest on an OR curve, plotting OR *versus* the cutoff [1,4]. An "OR curve" was used in the context of autism spectrum disorders (ASD) research with results presented for a range of gestational age cutoffs to help clarify heterogeneity of the ASD-gestational age associations reported in the literature [4]. Our previous work presented OR curves for associations for which the shape of the underlying association was unknown.

There are also methodological caveats with dichotomization of an exposure that is imperfectly measured [10,11]. Even when measurement error in a continuous exposure is non-differential before categorization, the process of categorization creates dependence of the resulting misclassified binary variable on the outcome when (1) exposure categories are created from a continuous exposure that is assessed with error and (2) the exposure and outcome are not independent, producing heterogeneity of risk within exposure categories [12,13]. This study is an investigation of the first condition. In practice one can never be sure of the second condition (independence of exposure and outcome), and therefore evaluation of the degree of association must consider the possibility of differential misclassification. However, if one was certain that exposure and outcome are independent, then there would be no need to investigate their association using any exposure metric and the study investigating their association would not take place, rendering condition 2 of no import here.

The extent to which such sources of differential misclassification produce important biases in epidemiologic analyses is currently not known as it can be expected to depend on multiple factors, including the degree of measurement error, the true strength of the effect of the exposure on the outcome, and the choice of the cutoff value. The contribution of model misspecification to bias in studies that categorize a mismeasured exposure variable is not known and it is beyond the scope of the present study to gauge the relative importance of the various sources of error. The influence of the true shape on the exposure-outcome association biases that arise from dichotomization is currently unknown as all prior research was conducted for "linear" models (in a sense that log (relative risk) was linear with exposure in simulations). It has long been recognized that this is a topic of some concern (e.g., [13]) but comprehensive evaluation of the degree of non-differential misclassification created by dichotomization is lacking.

In this study we investigated the effects of changing the cutoffs for a hypothetical causal exposure, assessed with and without measurement error, on the association with ASD in a simulated population similar to the source population for the Early Autism Risk Longitudinal Investigation (EARLI). We also build on our previous work (in which the true shape of the underlying association was unknown) by investigating whether the shape of the OR curve provides insight into the true underlying shape of the exposure-outcome relationship. EARLI is an enriched-risk pregnancy cohort (mothers of a child affected by an ASD enrolled at the start of a subsequent pregnancy) intended to investigate environmental exposures potentially important in the etiology of ASD (http://www.earlistudy.org/) [14]. Our previous work that was informed by the EARLI described the potential impact of measurement error in exposures and outcomes under different scenarios of study size, precision of exposure measurement and magnitude of the true association [15]. Importantly, numerous continuous environmental exposures for which ideal cutoffs are either unknown (e.g., brominates, flame retardants, bisphenol-A) or contentious (e.g., mercury and lead) are of interest to EARLI due to suspicion of their neurotoxicity that is relevant to ASD. Although our work is motivated by EARLI, the simulations we conduct and conclusions should be of general interest; keeping the simulations closely connected to EARLI not only serves our own interest, but also illustrates how our simulation study can be adapted to the complexities of a specific problem rather than presenting conclusions for some unrealistically simplistic set-up.

Table 1. True and observed values in the simulated population.

Values in the Population (Notation)	True Values	Observed Values	Measurement Error	Postulated True Association with Latent Measure of Outcome [a]	Cutoffs Range (Increments)
Environmental exposure 1 (X_1)	$X_1 \sim N(0,1)$, correlated with X_2 by Pearson correlation $\rho = 0.7$	$W_1 = X_1 + \varepsilon_1$	$\varepsilon_1 \sim N(0, \sigma^2)$, where $\sigma^2 \in \{0.0625, 0.25, 1\}$	$\{0.15, 0.25, 0.5\}$	-3 to 3 (in increments of 0.1) See Figure 1 for cutoffs for 5 categories
Environmental exposure 2 (X_2)	$X_2 \sim N(0,1)$, correlated with X_1 by Pearson correlation $\rho = 0.7$	$W_2 = X_2 + \varepsilon_2$	$\varepsilon_2 \sim N(0, 0.25)$	0	<1 *vs.* ≥ 1
Sex (Z)	$Z \sim \text{Binomial}(0.5, 1)$	Z	None	1	
Gestational age (X_{ga})	$X_{ga} = (43 - \gamma)$, where $\gamma \sim \chi^2(3)$ 1 week was subtracted from the above gestational age for 5% of males	$W_{ga} = R((X_{ga} + \varepsilon_{ga}); 23, 43)$, Where $R(.)$ [b] is function that round expression to integers, and then truncated to 23 to 43 weeks.	$\varepsilon_{ga} \sim N(0, \frac{1}{7^2})$	0.1	<37 *vs.* ≥ 37
Autism endophenotype (latent, Y)	Linear model: $Y_{Linear} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z + \beta_4 X_{ga} + \varepsilon_y$ "Threshold" (semi-linear) model: If $x_1 < \text{mean}_{x1} - \text{standard deviation}_{x1}$ then $Y_{Threshold} = \beta_2 X_2 + \beta_3 Z + \beta_4 X_{ga} + \varepsilon_y$ If $x_1 \geq \text{mean}_{x1} - \text{standard deviation}_{x1}$ then $Y_{Threshold} = 1.5 \times \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z + \beta_4 X_{ga} + \varepsilon_y$ "Saturation" (semi-linear) model: If $x_1 < \text{mean}_{x1} - \text{standard deviation}_{x1}$ then $Y_{Saturation} = 1.5 \times \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z + \beta_4 X_{ga} + \varepsilon_y$ If $x_1 \geq \text{mean}_{x1} - \text{standard deviation}_{x1}$ then $Y_{Saturation} = 0.5 \times \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z + \beta_4 X_{ga} + \varepsilon_y,$ $\varepsilon_y \sim N(0,1)$	Y [b] $= R(T(y); 0, 18)$, where $T(.)$ is a function that is transformed to a Y log-normal distribution that matched the observed AOSI in EARLI	due to rounding by $R(.)$ [b]	Not applicable	0–6, 7–18

[a] coefficients of linear regression, see text and bottom of the table for details, β's; [b] $R(f(.); \min, \max)$ is the function that rounds values of function $f(.)$ to integers and truncates values (retains only values) that fall within interval [min, max].

2. Materials and Methods

2.1. Simulated Population

The population parameters used in this and our previous work [15] are listed in Table 1. Briefly, a population cohort of 1,000,000 children was generated with observed sex (Z) and gestational age (W_{ga}) distributions similar to the general US population. We generated such a large simulated dataset in order to evaluate how our method performs when instabilities due to sampling error and small strata are not an issue. We simulated true gestational age X_{ga} (in weeks) in the cohort as $43 - \gamma$ where $\gamma \sim \sim\chi^2(3)$ distribution, after which 5% of boys were assigned gestational age that was 1 week shorter to account for shortened gestation, on average, among boys [16]. Completed weeks of gestational age were considered (as is typically reported) and only live births between 23 and 43 weeks (inclusive) gestational age were included in the simulation.

The cohort had simulated exposure to two agents: exposure 1, represented by X_1, and exposure 2, represented by X_2, that both follow standard normal distributions and are correlated (Pearson $\rho = 0.7$). Only exposure 1 exerts causal influence on ASD-related phenotype. This constellation of covariates assumed that there is no true confounding, since the outcome is independent of X_2 and all other covariates are independent. The simulated population was created and analyses of the entire population (not a sample thereof) were conducted using SAS version 9.2 (SAS Institute, Cary, NC, USA). Ethics committee review and informed consent were not necessary as this was a simulation study. Our SAS code will be made available to other researchers upon request.

2.2. Covariate Measurement Error

We assumed that the two environmental exposures are observed with classical measurement error. The first ($W_1 = X_1 + \varepsilon_1$, $\varepsilon_1 \sim N(0,\sigma^2)$, where $\sigma^2 \in \{0.0625, 0.25, 1\}$ and will be referred to as the measurement error variance) is a causal exposure. The extent of measurement error spanned the plausible range: From good precision of environmental measurements with error approximately 6% of true exposure variability to poor precision with error variance equal to the true exposure variability. We present in detail only results for σ^2 of 0.25 and 1, with the full details given in the Appendices. The second exposure (W_2) was not causal ($W_2 = X_2 + \varepsilon_2$, where $\varepsilon_2 \sim N(0,0.25)$). Gestational age was subject to two sources of error: (a) Classical error in the observed continuous measure of the length of gestation, $W_{ga} = X_{ga} + \varepsilon_{ga}$, where $\varepsilon_{ga} \sim N(0,\sigma^2_{ga})$ and (b) round-off error as gestational age was reported in completed weeks (23−43 weeks). All errors were conditionally independent of each other.

2.3. Linear and Semi-Linear Risk Models

In addition to X_1, sex (sex ratio of 4 males: 1 female [17]) and gestational age [18–20] exerted some causal influence on dichotomous ASD status and continuous ASD-related phenotype (Y). We verified that these yielded parameter estimates within the expected range. The true shape of the exposure-response curve is unknown so the simulated disease was modeled using the following three forms:

Linear model:

$Y_{Linear} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z + \beta_4 X_{ga} + \varepsilon_y$

"Threshold" (semi-linear) model:

If $x_1 <$ mean$_{x1}$-standard deviation$_{x1}$ then $Y_{Threshold} = \beta_2 X_2 + \beta_3 Z + \beta_4 X_{ga} + \varepsilon_y$

If $x_1 \geq$ mean$_{x1}$-standard deviation$_{x1}$ then $Y_{Threshold} = 1.5 \times \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z + \beta_4 X_{ga} + \varepsilon_y$

"Saturation" (semi-linear) model:

If $x_1 <$ mean$_{x1}$-standard deviation$_{x1}$ then $Y_{Saturation} = 1.5 \times \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z + \beta_4 X_{ga} + \varepsilon_y$

If $x_1 \geq$ mean$_{x1}$-standard deviation$_{x1}$ then $Y_{Saturation} = 0.5 \times \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z + \beta_4 X_{ga} + \varepsilon_y$,

The slope of the causal association with X_1 above and below the inflection point defined by (mean$_{x1}$ − standard deviation$_{x1}$) was altered to create deviation from linearity. For example, in the saturation model, the effect of X_1 is three times greater below than above the inflection point ($1.5 \times \beta_1 X_1 / 0.5 \times \beta_1 X_1 = 3$). These multipliers were chosen arbitrarily but have numerical values that yield plausible "average" associations if the linear model was fitted to the data. For all models $\varepsilon_y \sim N(0,1)$, β_2 was 0 since X_2 had no true effect on Y. The latent values of ASD-related phenotype ($Y_{Linear, Threshold, Saturation}$) were computed for each of three $\beta_1 \in \{0.15, 0.25, 0.5\}$ that correspond to "weak", "moderate", and "strong" underlying associations (see Heavner *et al.*, [15] for more details on the simulation parameters). We present results only for the weak and strong associations in text, with all the details available in the Appendices.

2.4. Case Definition and Outcome Misclassification

The values of Y_{Linear}, $Y_{Threshold}$, and $Y_{Saturation}$ were transformed to yield values in the valid range (0–18) for Autism Observation Scale in Infants (AOSI) scores [21], the outcome measure of interest. Latent continuous ASD-related endophenotype, was subjected to rounding error and categorized using a single clinically relevant cutoff. For Y_{Linear}, $Y_{Threshold}$ and $Y_{Saturation}$, $Y_c^* = 0$ if Y^* is 0–6 and $Y_c^* = 1$ if Y^* is 7–18 (Table 1). For all exposure-disease models, subjects with a high AOSI score were considered to be cases. An AOSI score of at least seven has been shown to have clinical and/or etiologic significance. It must be noted that we do not equate high AOSI with a clinical diagnosis of ASD. We merely assert that ASD is a collection of traits that naturally occur on a continuous scale and are segregated into binary disease and healthy groups based on some criteria. This is consistent with the current conceptualization of ASD as a "spectrum" of phenotypes, not a definitive state common to all cases, as is true of other conditions (e.g., death from a cardiac event or acquisition of an infection). Various AOSI sub-scales may be more related to environmental exposure than others but this is not our focus here and our argument applies to any measures of continuous traits that have AOSI-like properties.

2.5. Investigating the Effect of Changing the Cutoff(s) for the Observed Causal Exposure (W_1)

The mismeasured exposure was dichotomized in the first set of analysis and then a separate analysis was conducted where it was divided into five categories.

Two Categories

W_1 was *dichotomized*, yielding W_{1c}, at 61 different cutoffs (cutoff = φ) ranging from the population mean plus or minus three standard deviations in increments of 0.1 such that $W_{1c} = 1$ if $W_1 \geq$ φ, and $W_{1c} = 0$ if $W_1 <$ φ. The reference group for these dichotomized variables was $W_{1c} = 0$.

Five Categories

W_1 was then divided into *five categories* based on the number of standard deviations away from the mean value in the cohort. Four cutoffs (φ1 – φ4) were used such that the reference group was between φ2 and φ3 (referent group: φ2 ≤ W_1 < φ3) (Figure 1). The distance between two consecutive cutoffs varied (distance between cutoffs = width of the reference group = δ) from 0.5 standard deviations to 1.5 standard deviations. The midpoint of category 3 (the reference category) was the estimated population mean (0) of W_1. This resulted in three groups of equal width δ (the 2nd, 3rd (reference) and 4th groups) and two groups of varying width at the extremes of the exposure distribution (groups 1 and 5). A number of other multi-category approaches are reasonable and used in epidemiology (e.g., categories formed to ensure equal number of "cases", categories reflecting percentiles of exposure distribution in disease-free individuals, categories based on hypothesized thresholds or other quantities with intuitive interpretation such as those relevance to clinical practice) but it is beyond the scope of this paper to compare them. We present an approach that bases the categories on the distribution of the exposure in the entire cohort; we make our SAS code available upon request for other researchers to explore alternative scenarios.

δ= width of the reference group = 0.5 to 1.5 standard deviations (1 standard deviation = 1)

Category 1: W_1< (Mean – 1.5*δ)
Category 2: (Mean – 1.5*δ) ≤ W_1< (Mean - 0.5*δ)
Category 3 (referent): (Mean – 0.5* δ) ≤ W_1< (Mean + 0.5*δ)
Category 4: (Mean + 0.5*δ) ≤ W_1< (Mean + 1.5*δ)
Category 5: (Mean + 1.5*δ) ≤ W_1
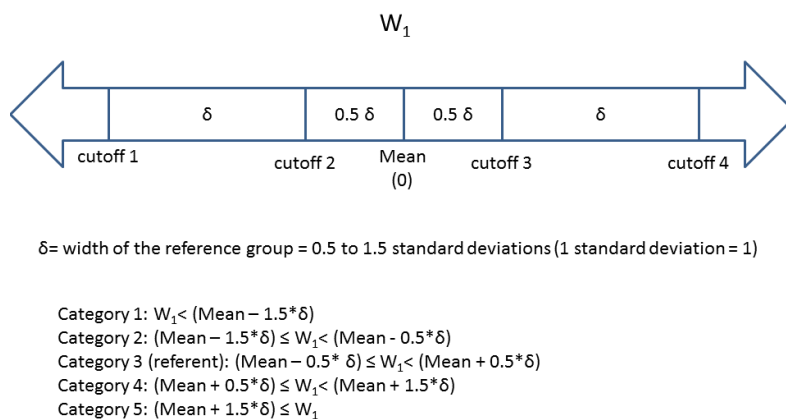
Figure 1. Descriptions of four cutoffs and five categories for W_1.

2.6. Assessing the Effect of Variation in the Cutoff for W_1

For each W_{1c}, a logistic model was computed in the form:

$$LOGIT\left(pr(Y_c^* = 1)\right) = c_0 + c_1 W_{1c} + c_2 W_{2c} + c_3 Z + c_4 W_{gac} \; (logistic).$$

Thus, for logistic models, exp (c_1) is the odds ratio of the estimated effect of W_{1c} (OR_1) on odds of AOSI ≥ 7. All independent variables were considered as dichotomized predictors ($W_{2c} = 1$ if $W_2 \geq 1$, $W_{2c} = 0$ if $W_2 < 1$ and $W_{gac} = 1$ if $W_{ga} \geq 37$, $W_{gac} = 0$ if $W_{ga} < 37$) of the dichotomized AOSI score. Although these variables are not true confounders, their inclusion in the logistic models realistically simulates regression analyses that are common in ASD research. In addition, the inclusion of these variables is not detrimental to the model from a statistical perspective as the size of the population is

very large. We also fitted analogous logistic regression models with values of the exposures that were not contaminated by measurement error, e.g., with X_1 instead of W_1, *etc*. These models were repeated for the threshold and saturation models.

OR curves were created for the analyses of changing the cutoff used to dichotomize W_1. These curves were created by plotting the OR_1 derived from logistic regression on the y-axis and the cutoff used to dichotomize W_1 (φ) on the x-axis. A separate set of plots of the ORs *versus* the width of the reference group (δ) was created for the scenarios with five categories of W_1 (yielding points for four ORs for each width of the reference group).

We examined the degree of misclassification and its dependence on the outcome created by dichotomization of W_1. The true classification with respect to the cutoff was the model without measurement error, *i.e.*, when X_1 is dichotomized. For each cutoff of W_1, the sensitivity and specificity were calculated separately for cases and controls. Sensitivity and specificity were defined as follows:

$$Sensitivity = \frac{\# \ exposed \ both \ without \ measurement \ error \ and \ with \ measurement \ error}{\# \ exposed \ without \ measurement \ error}$$

$$Specificity = \frac{\# \ unexposed \ both \ without \ measurement \ error \ and \ with \ measurement \ error}{\# \ unexposed \ without \ measurement \ error}$$

3. Results and Discussion

The characteristics of the population are similar to what was expected based on the simulation parameters (data not shown). The effect of changing the cutoff for dichotomizing W_1 is shown in Figure 2 (and Figure A1). Each of the three exposure-outcome relationships had a distinctly shaped OR curve. Even with such a large population (n = 1,000,000), at the extreme cutoffs, there was non-monotonic oscillation in the effect estimates that is open to various interpretations. (In all scenarios, the greatest cutoff for the mismeasured exposure yielded more than 750 exposed cases, which is rarely attained even in 'large' epidemiology studies. For example, in the linear model with a weak true association and small measurement error, there were 846 cases with a mismeasured exposure at least as great as the highest cutoff.) The linear models had U-shaped graphs with slightly lower ORs for very large cutoffs compared to very small ones. For all strengths of the true association, the OR was increasingly underestimated (compared to the scenario with no measurement error) and the OR curve became flatter as measurement error increased. The highest estimates of OR tend to occur when cutoffs are made at the extremes of the exposure distribution, with changes from an easily dismissed OR of <2 tending to occur with the cutoff in the center of the distribution and a clear "signal" of OR >2 apparent with categorizations at the extremes when there was a moderate underlying association.

The semi-linear (threshold and saturation) models had a true inflection point (at one standard deviation below the mean). This inputted inflection point was obvious in the OR curves constructed under the semi-linear models without measurement error (Figure 2B,C). The OR curves appeared more linear and the inflection point was more obscured as the measurement error variance increased. In general, the OR became closer to the null as the measurement error increased. It is noteworthy that for the threshold model, contrary to intuition, dichotomization near the inflection point results in overestimation, rather than attenuation, of the OR in the presence of measurement error.

Weak true association　　　　　**Strong true association**

(**A**) Linear model

(**B**) Threshold (semi-linear) model

(**C**) Saturation (semi-linear) model
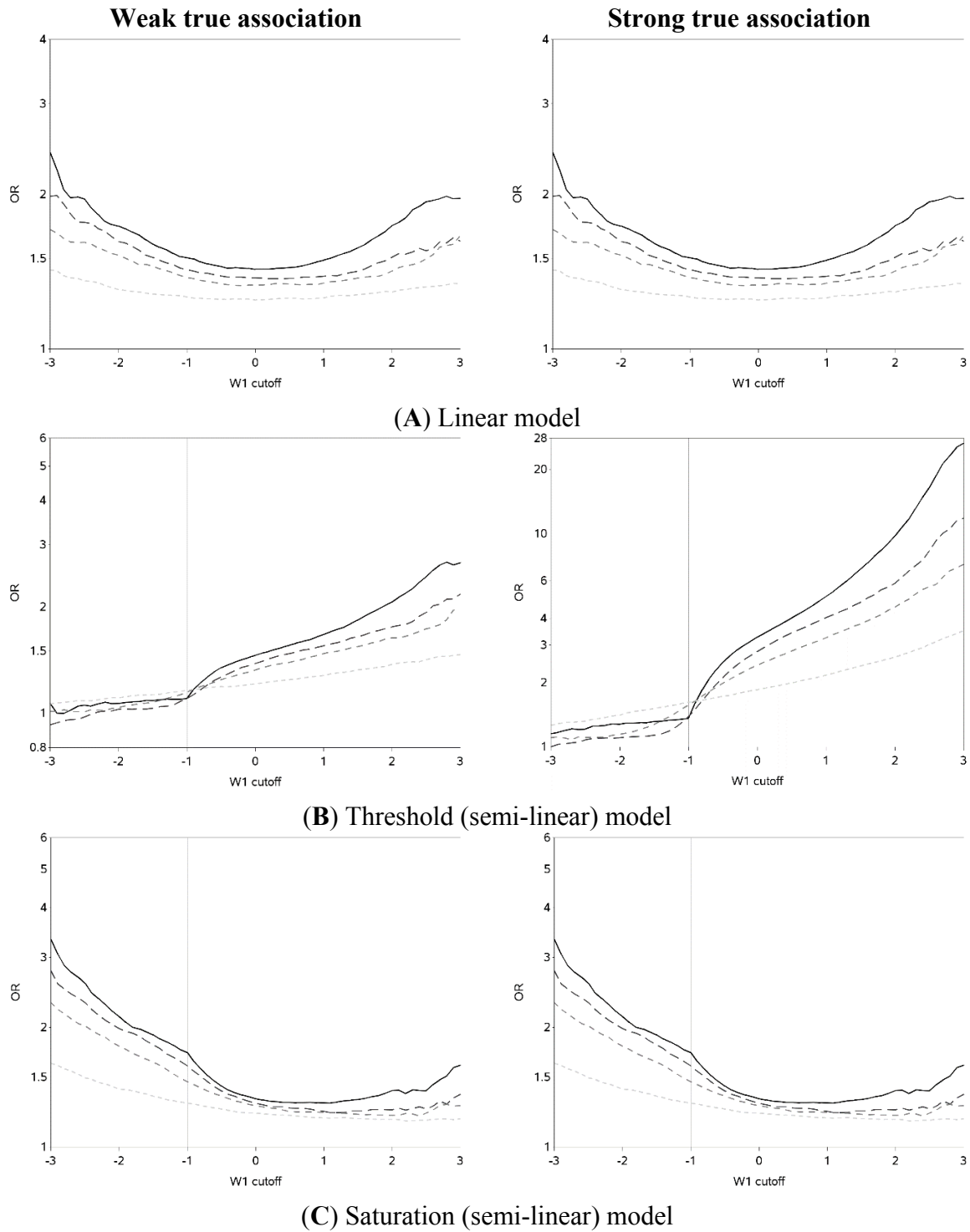
No measurement error ———————

Measurement error variance = 0.0625 — — — — — — —

Measurement error variance = 0.25 - - - - - - - - - - - -

Measurement error variance = 1 - - - - - - - - - - - - -

The W_1 cutoff was changed from −3.0 to 3.0 in increments of 0.1. Exposed $W_1 \geq$ cutoff, referent: $W_1 <$ cutoff.

The vertical reference line for the semi-linear models is the inputted inflection point.

Figure 2. Odds ratios (ORs) for different cutoffs (between mean ± 3 standard deviations) used to dichotomize a mismeasured exposure (W_1).

The two distinct slopes (below and above the mean minus one standard deviation) were clear for both the threshold and saturation models without measurement error. There was an apparent inflection point for the threshold models with moderate or large variance in the measurement error. The shape of the graph changed and the inflection point seemed to move as the measurement error variance increased. The shape of the curves for the threshold model for cutoffs greater than the mean resemble the steeper curves of the equivalent part of the curves for the linear models, as expected. It is also apparent that the highest ORs are obtained when the cutoff is made in the steeper section of the true exposure-response curve: this can be used to distinguish the linear from non-linear exposure-response associations. The influence of choice of cutoff on the overall results is perhaps most clear in the case of the saturation model: for a moderate true association, cutoffs below the inflection point tend to yield evidence of a "strong" association (OR > 3) while cutoffs above the inflection point yield an OR less than two. ORs less than two are frequently dismissed as unimportant in the epidemiology literature.

The effect of changing the cutoffs used to create five categories of the causal exposure is illustrated in Figure 3 and Figure A2. The graphs were distinct for each of the three exposure-response models. For the linear models, as the width of the reference group (δ) increased, the ORs were farther from the null for all scenarios. In addition, as the measurement error variance increased, the ORs for all four categories moved closer to the null. Again, we note that whereas under low measurement error the graphs can be used to distinguish between associations of different shapes, this identification of shapes becomes problematic as measurement error increases, even with the increase in number of categories from two to five. In practice, this may lead to both under- and over-interpretation of information contained in the data about the shape of the exposure-response relationship (e.g., dismissing non-linearity when it is present, or positing a curvilinear shape when the exposure-response is linear, respectively), even though it would leave little doubt that there is an association between the exposure and outcome.

When the distance between cutoffs was shortened it was possible to gain more insight into possible thresholds and inflection points (Figure A3 contains OR curves when the true association was weak, varying the cutoffs used to create five categories by smaller increments (0.01)).

The sensitivity and specificity for each cutoff used for dichotomization is illustrated in Figure 4 and Figure A4. There was a clear pattern of trade-off between sensitivity and specificity, with cutoffs at higher apparent exposure values producing classifiers that favored specificity over sensitivity and *vice versa*. Cutoffs in the center of the apparent exposure distribution produced values of sensitivity and specificity that were similar to each other. The curves for sensitivity and specificity cross below the mean for controls and above the mean for cases. The specificity is unstable for smaller cutoff values and the sensitivity is unstable for larger cutoff values. Sensitivity tended to be higher among cases and specificity higher among controls. The deviation from non-differential misclassification was most pronounced when dichotomization was done at the extremes of the exposure distribution. Furthermore, the degree of divergence from non-differential misclassification increased with the strength of the true exposure-outcome association as well as the magnitude of the measurement error variance. Receiver operator characteristic (ROC) curves for selected cutoffs (−3, −2, −1, 0, 1, 2, 3) are shown in Figure A5. In general, the ROC curves varied little by the cutoff used to dichotomize W_1, particularly when the true association was weak.
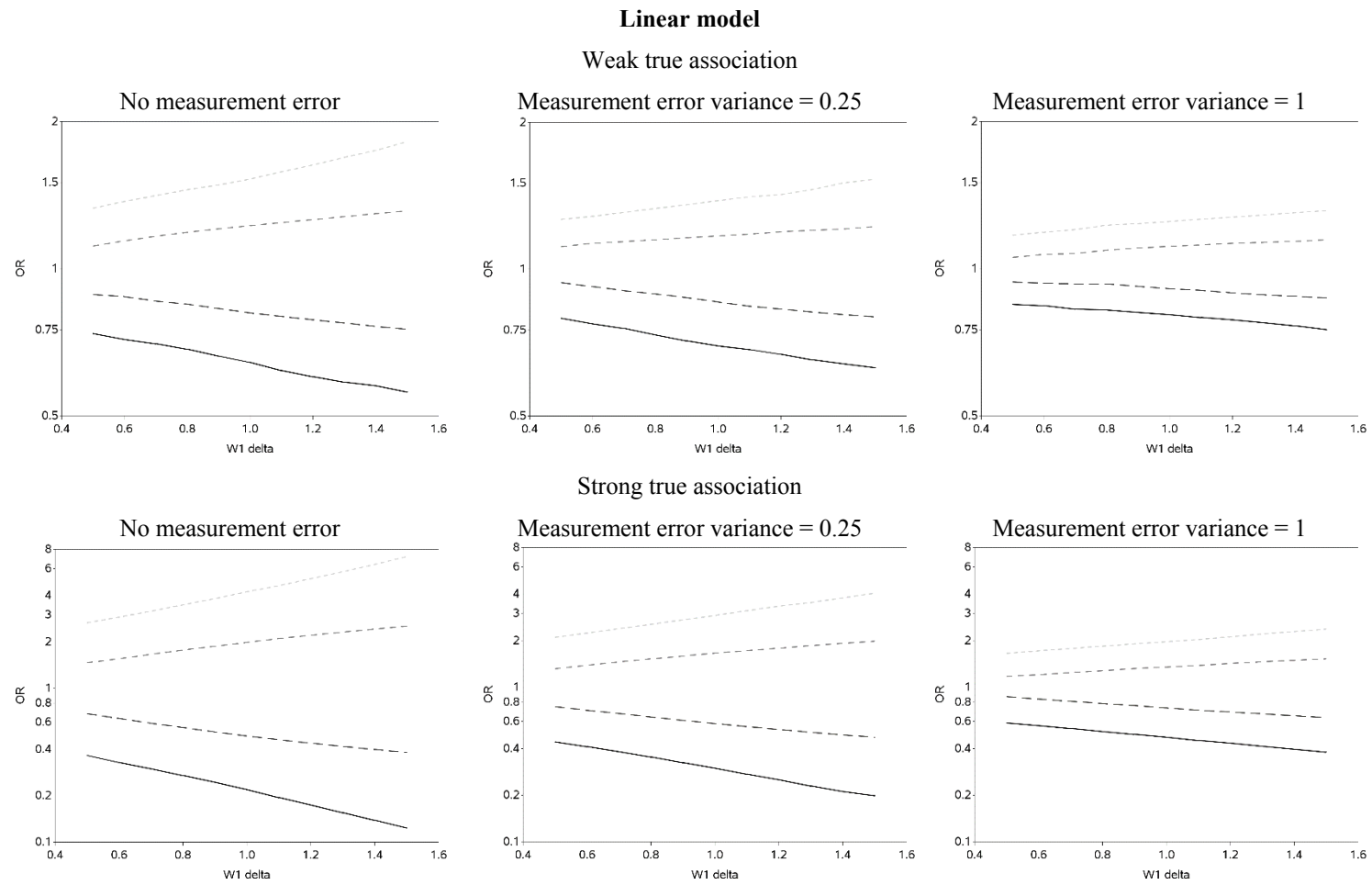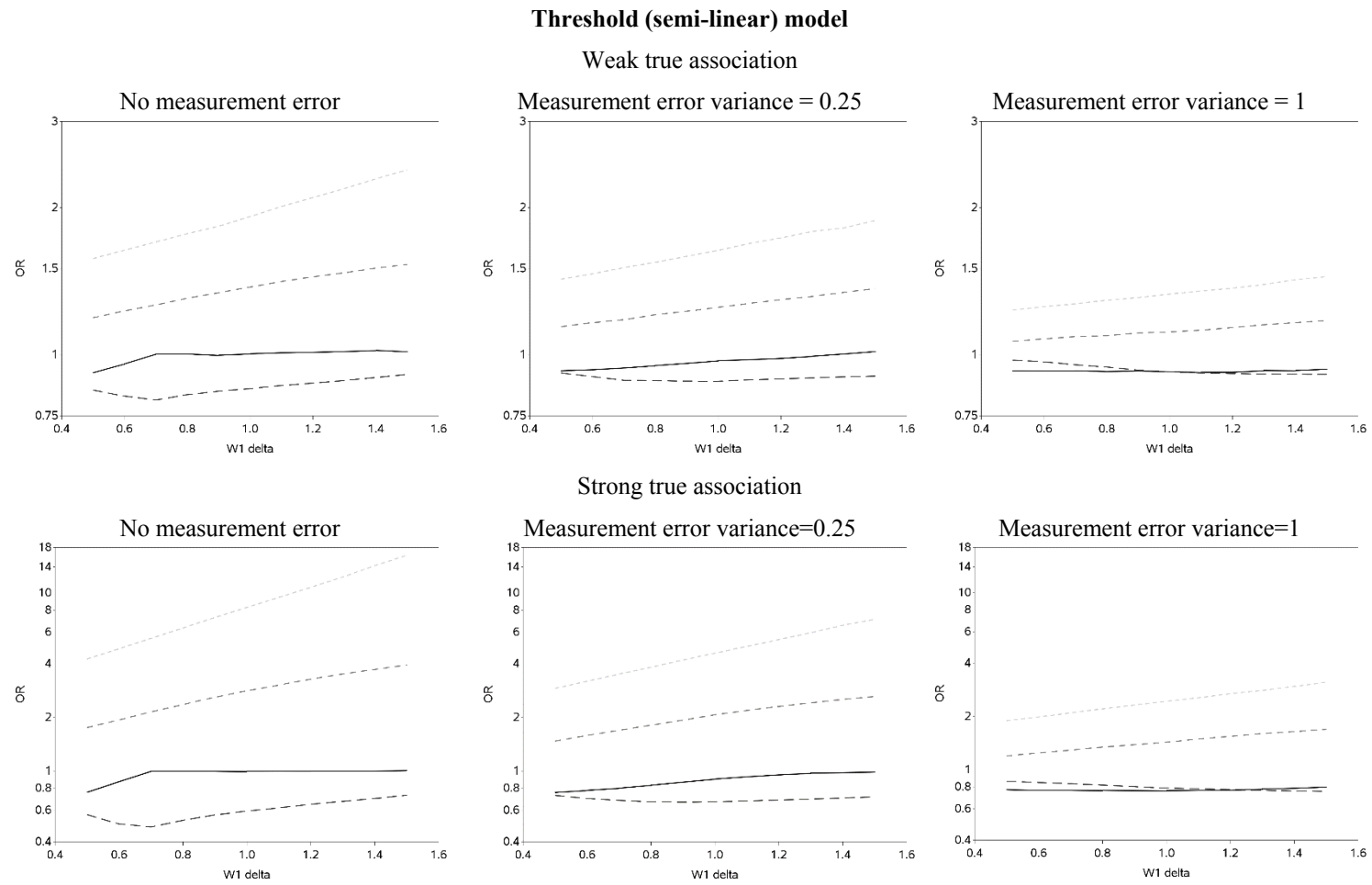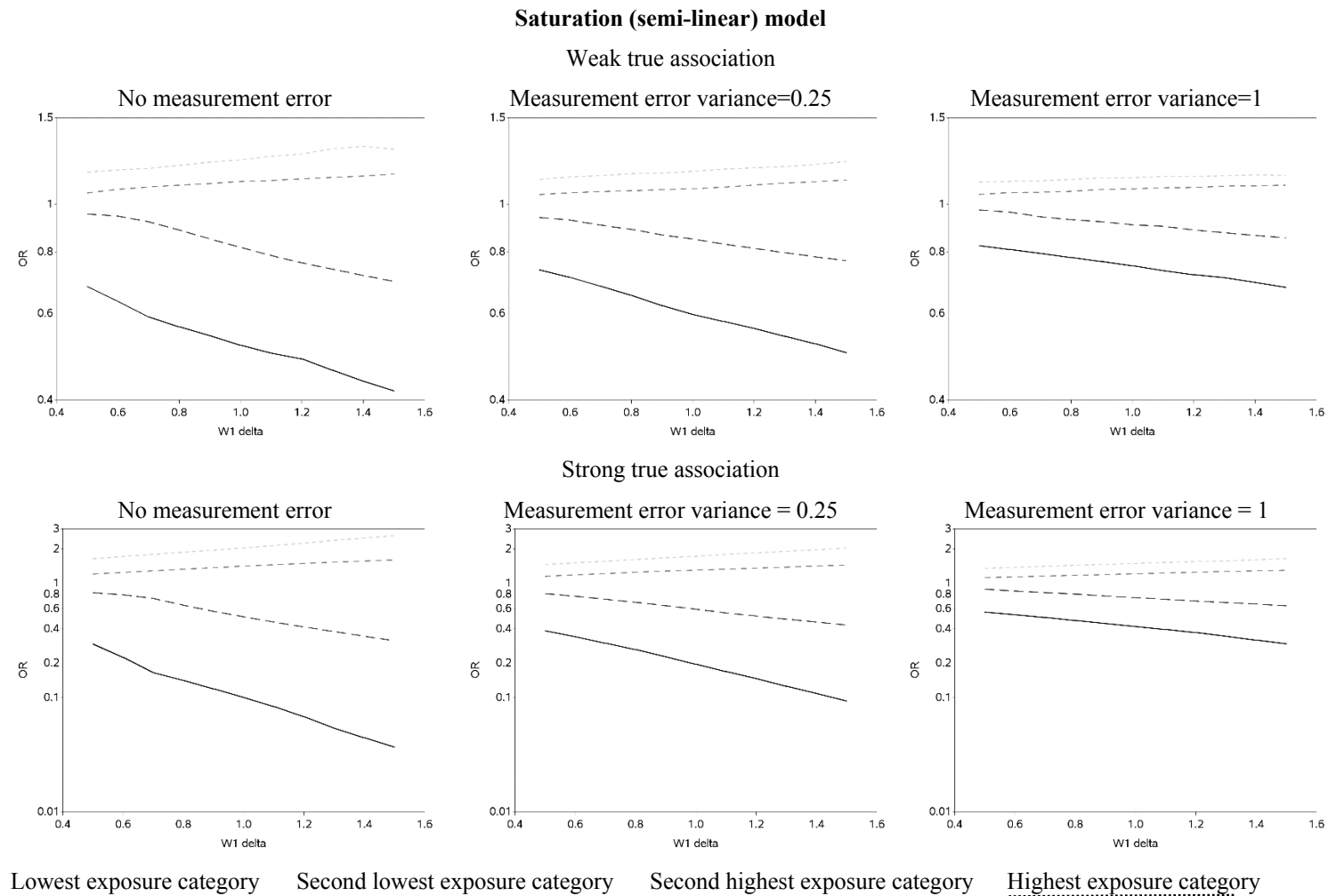
Linear model

Weak true association

Strong true association

Figure 3. *Cont.*

Threshold (semi-linear) model

Weak true association

Strong true association

Figure 3. *Cont.*

Saturation (semi-linear) model

Weak true association

Strong true association

Lowest exposure category   Second lowest exposure category   Second highest exposure category   Highest exposure category

Figure 3. Odds ratios (OR) for different cutoffs used to create 5 categories for the causal yet mismeasured exposure (W_1) (δ is identified as "W1 delta" in the figure). The 4 W1 cutoffs were centered at 0 with δ from 0.5 to 1.5 standard deviations, in increments of 0.1. Reference category for W1 is mean $+/-$ ($0.5 \times \delta \times$ standard deviation).

Weak true association

Strong true association

Measurement error variance = 0.25 **Measurement error variance = 1** **Measurement error variance = 0.25** **Measurement error variance = 1**

Linear model

Threshold model

Figure 4. *Cont.*

Weak true association

Strong true association

Measurement error variance = 0.25

Measurement error variance = 1

Measurement error variance = 0.25

Measurement error variance = 1

Saturation model

Cases—black

Controls—grey

Sensitivity—solid lines

Specificity—dotted lines
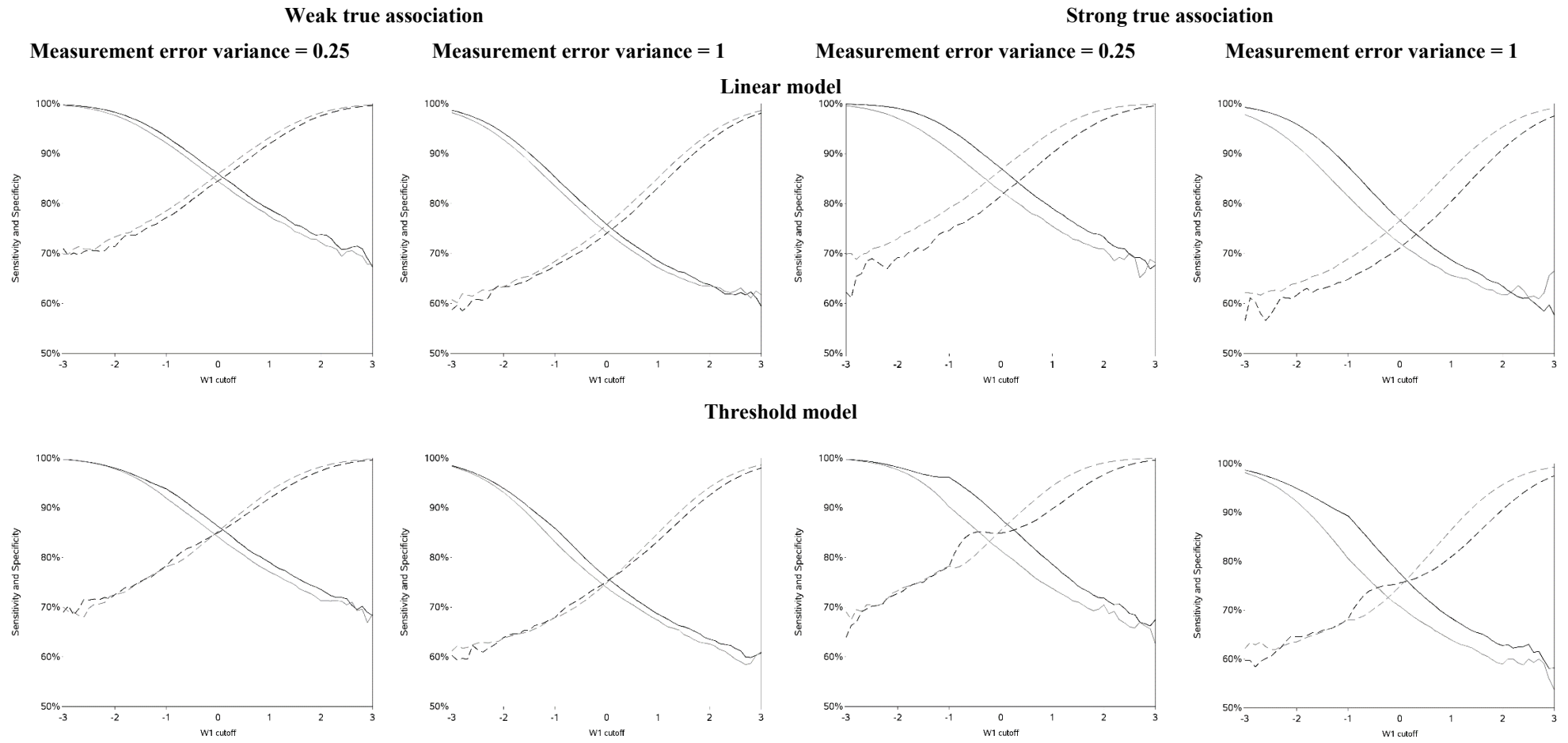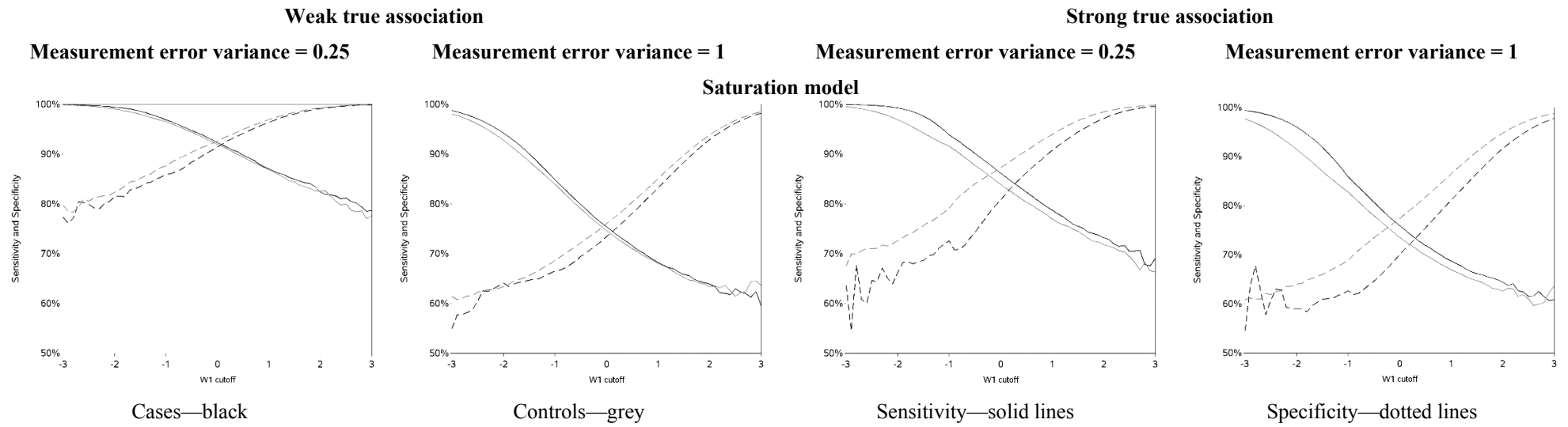
Figure 4. Sensitivity and specificity at each cutoff for the mismeasured exposure (W_1).

4. Discussion

As expected, the cutoff (s) chosen for categorizing continuous variables has a profound effect on the results of this simulation study. The two types of OR plots presented here allow us to gain insight into the shape of the underlying exposure-outcome relationship, particularly in scenarios where there is minimal measurement error. This builds on our previous work [1,4] when we did not know the true shape of the underlying relationship. A U-shaped curve when the measure of the exposure of interest was dichotomized indicated that the underlying relationship was linear. This may have particular implication for meta-analyses of studies with binary exposures and outcomes. Cutoffs at different exposure values measured with different degrees of accuracy across studies may manufacture U-shaped associations and other non-linearities where none may in fact exist. As measurement error increases, we lose the ability to infer the shape of the underlying exposure-response relationship if the exposure of interest is dichotomized. The plots used for five categories of W_1 may be a good way to infer shape, even in the presence of larger measurement error. In practice, it may be sensible for a *post hoc* correction for measurement error (e.g., [22,23]) to be applied to the data before construction of the OR curve.

It is particularly important to investigate a wide range of cutoffs if there is a suspected inflection point as failure to include the true inflection point in the OR curve would likely lead to incorrect assumptions about the underlying shape of the exposure-response relationship. (Methods, such as those used by Hinkley [24] or Muggeo [25], may be used to identify such inflection points. However, we do not know how these operate with multiple sources of measurement error (in exposure and outcome) or under the realist conditions in ASD research that we simulated. Thus, we maintain that exploring the extent to which our approach is useful in addressing such question was justified, especially since our method is more easily implemented by researchers who do not have the support needed to implement more advanced statistical methods.) The number of categories and cutoffs are also important considerations as analyzing too many categories may lead to instability in the curve and incorrect inferences about the exposure-response relationship. If many categories are needed or desirable it is better to have a greater distance between cutoffs to obtain a clearer picture of the exposure-response relationship by minimizing cross-contamination of adjacent exposure groups. This simulation does not take into account random sampling error, the effects of which are well known. The use of smaller samples will lead to increasing instability, particularly in the tails of the OR curve. Our previous work [1] illustrated that in an OR curve for a large administrative dataset (n ~ 19,000), a relatively wide range of ORs was possible for what is suspected to be, at most, a modest association. Although we simulated a realistic scenario in autism research, model misspecification is an additional potential source of bias in this simulation and similar observational studies.

Use of non-parametric methods to model non-linearity is an attractive alternative to categorization [7]. But there are problems for which one is interested in the change in risk due to a change in exposure below a certain level (as is the case in environmental epidemiology and is the aim of regulatory exposure limits). Furthermore, our method does not make any assumptions that are otherwise involved in modeling exposure-response shapes, e.g., splines assume no measurement error. We in fact show that under 'small' measurement error our approach can yield insights into the true shape of the exposure-response association without invoking unfamiliar or hard to test assumptions, or complications

involved in selection of the exact form of the spline to fit to the data (e.g., number of nodes and degree of smoothing are left to the investigator's judgment and can be a source of uncertainty about just how much a given result is an artifact of selective reporting). Yet it is clear that measurement error leads to loss of features in the data and it is therefore perilous to infer any specific shape of the exposure-outcome association based on our method or any other method that does not correct for measurement error. Thus, we again reinforce the need to understand the nature of measurement error present in any particular analysis.

5. Conclusions

We presented results for a simple scenario in which there are no confounders and the theoretically optimal cutoffs for the other variables included in the regression models are known. The scenario presented here is common in autism research as well as other areas of epidemiology, in which sex and (gestational) age are included in the regression model even if they are not true confounders. We do not know how the presence of confounding may influence the OR curve and the inferences about the underlying shape of the exposure-response relationship but it is likely that it will be obscured. Our SAS code is available upon request for other researchers to apply to other scenarios.

A discussion of the merits and limitations of categorization of continuous variables is outside the scope of this study and has been discussed in-depth elsewhere (e.g., Bennette and Vickers [6]). This study reinforced results of our previous work that demonstrated that although vastly different numerical estimates may be obtained with different cutoffs, presenting results for a wide range of cutoffs is feasible, provides valuable insight into the exposure-response relationship when measurement error is small, and may help prevent fishing for a desired study result [1,4]. We also demonstrate that noticeable differential exposure misclassification arises when measurement error is present in a dichotomized exposure estimate and the cutoff is at the extremes of the observed distribution. Thus, it is important to consider that in such settings, an investigator may incur bias of unpredictable direction despite attempting (as is often done) to reduce the impact of measurement error by separating truly and highly exposed subjects from the rest.

It is possible (and relatively straightforward) to present measures of association for a wide range of cutoffs used to dichotomize a continuous exposure. If only one such cutoff is presented, the choice of cutoff has profound effects on the study results. Except in the presence of large measurement error, the shape of the OR curve may provide insight into the true shape of the underlying exposure-disease relationship.

Acknowledgements

Author contributions

Karyn Heavner and Igor Burstyn conceptualized the study. Karyn Heavner conducted the simulations. Karyn Heavner and Igor Burstyn drafted the manuscript. Both authors read and approved the final manuscript.

Conflict of Interests

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

List of Abbreviations

AOSI—Autism Observation Scale in Infants
ASD—autism spectrum disorders
EARLI—Early Autism Risk Longitudinal Investigation
OR—odds ratio

Appendix

Figure A1. *Cont.*

Figure A1. Odds ratios for different cutoffs (between mean ±3 standard deviations) used to dichotomize a mismeasured exposure (W_1) (all scenarios for Figure 2). The X_1 cutoff was changed from −3.0 to 3.0 in increments of 0.1. Exposed $W_1 \geq$ cutoff, referent: $W_1 <$ cutoff. The vertical reference line for the semi-linear models is the inputted inflection point.

Linear model

A. Weak true association
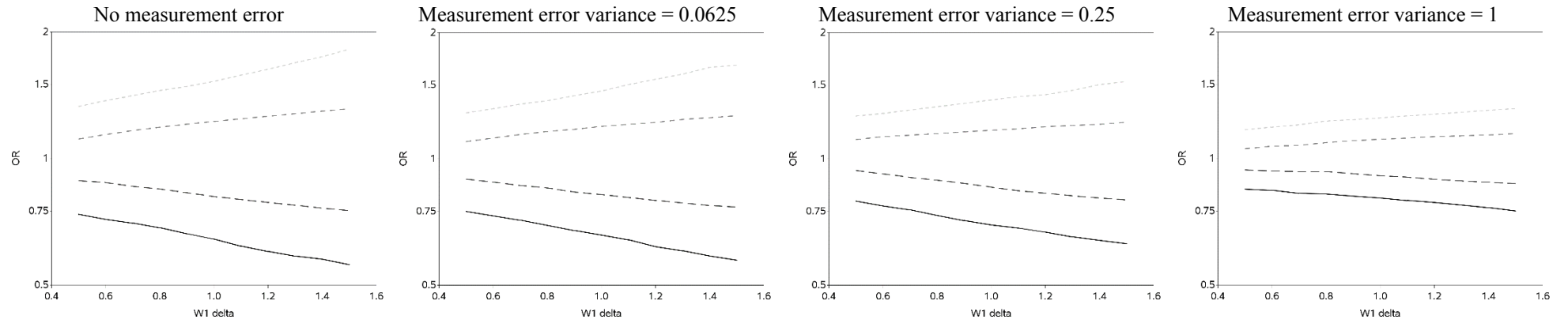
B. Moderate true association
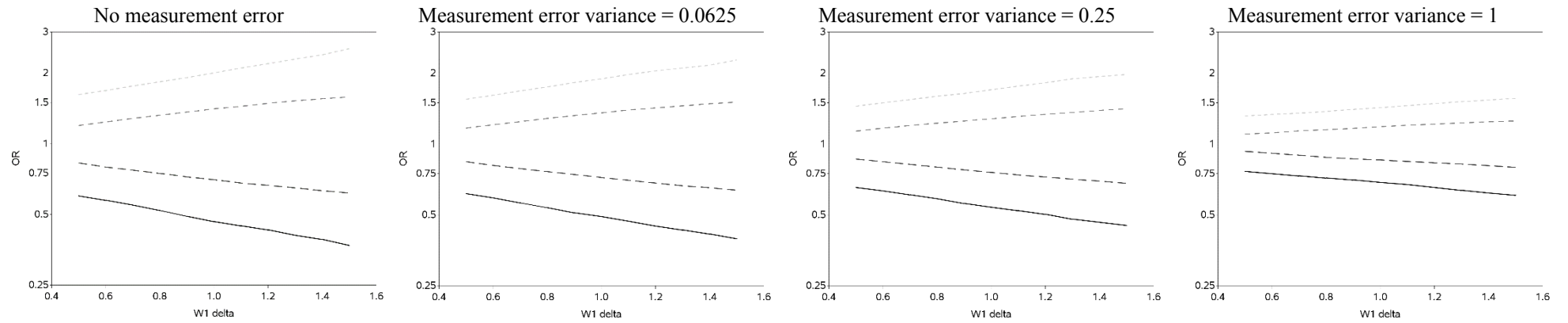
Figure A2. *Cont.*

C. Strong true association
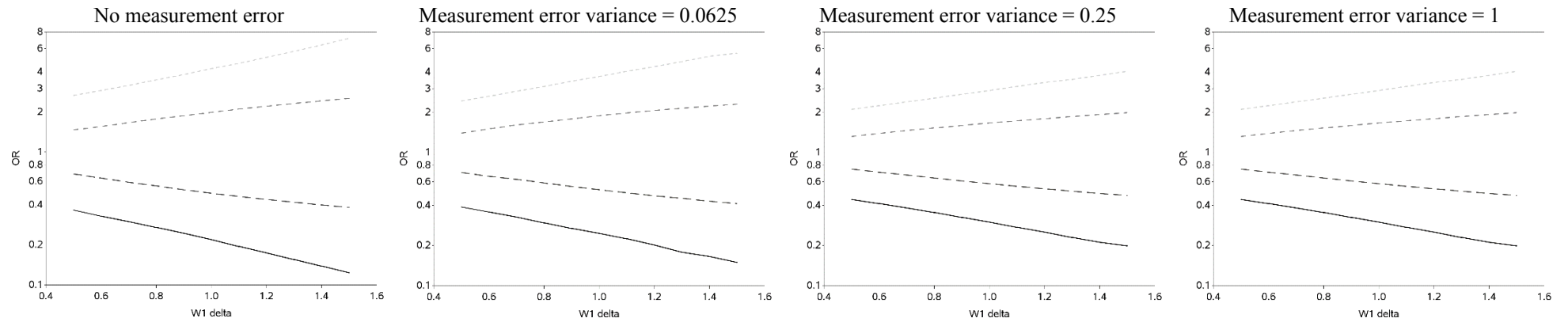
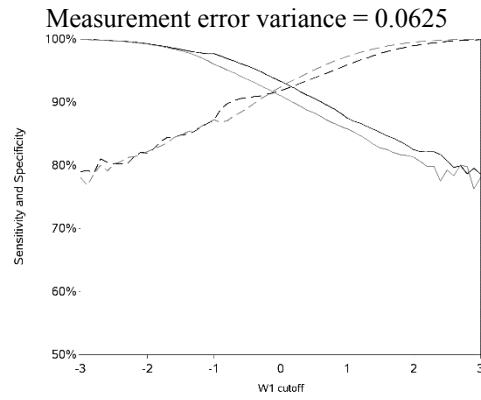Threshold (semi-linear) model

D. Weak true association
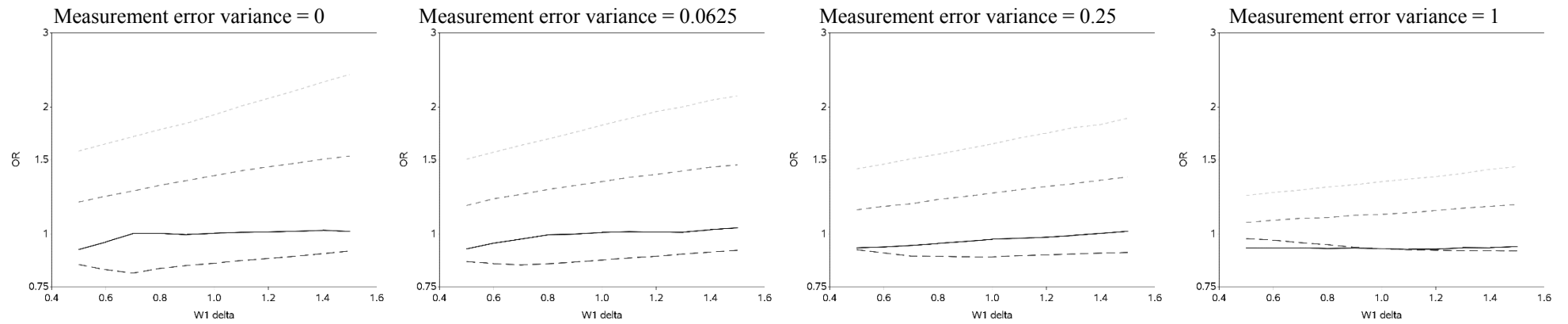
Figure A2. *Cont.*
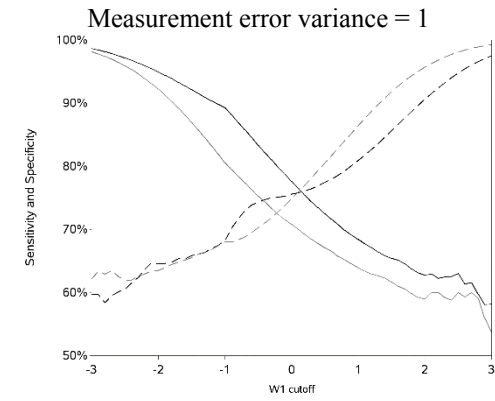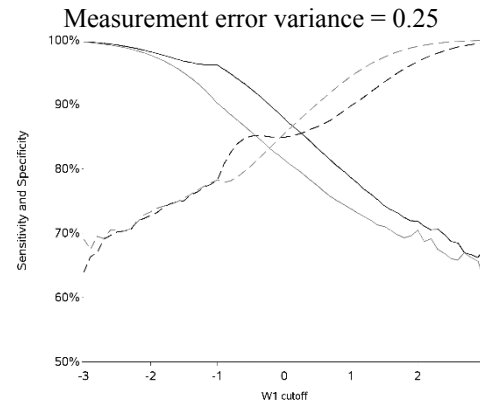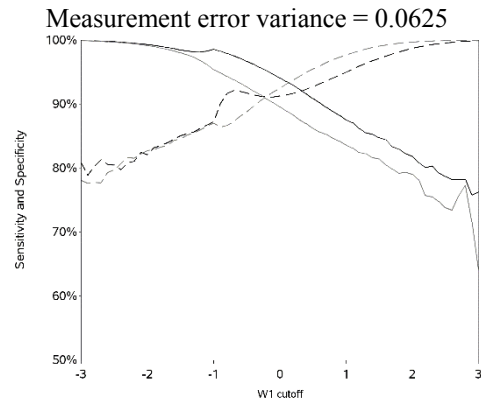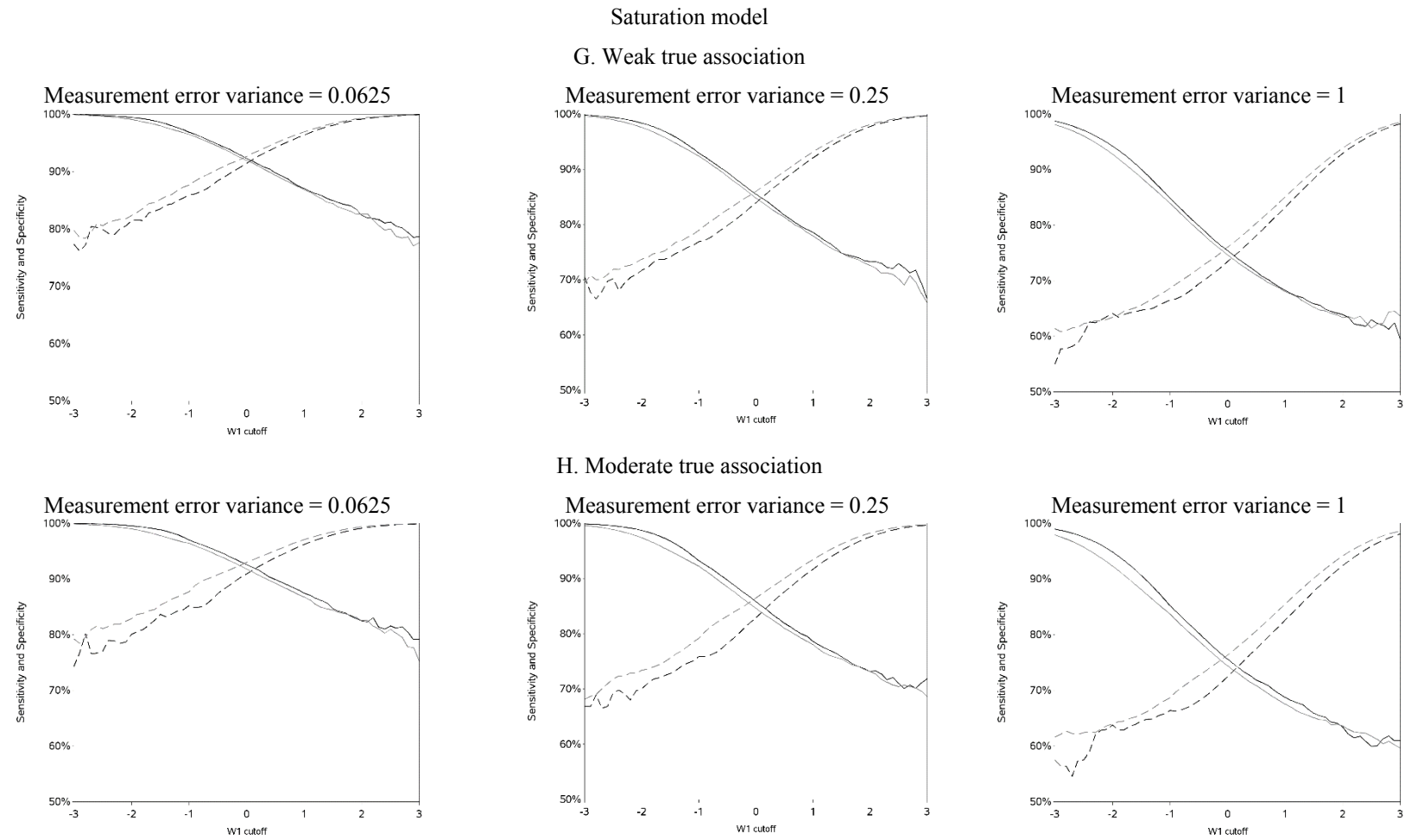
E. Moderate true association
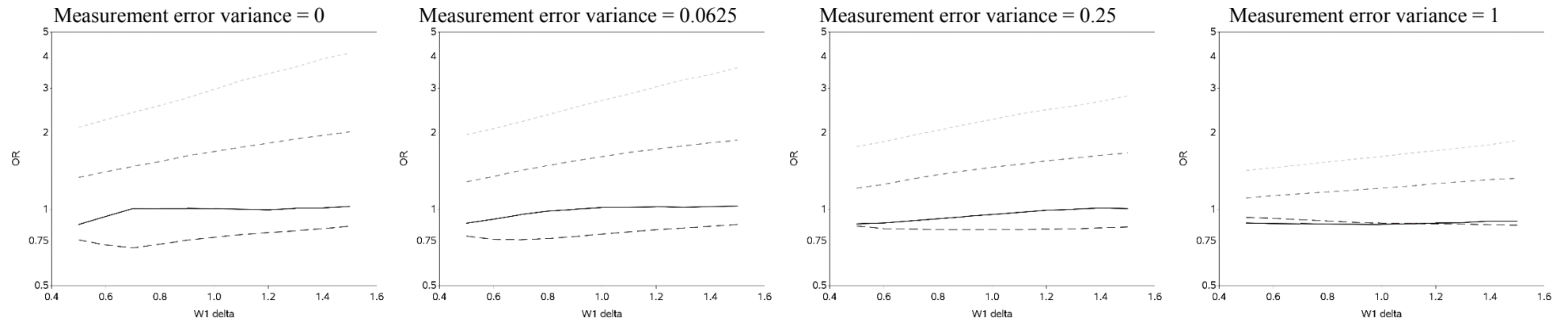
F. Strong true association
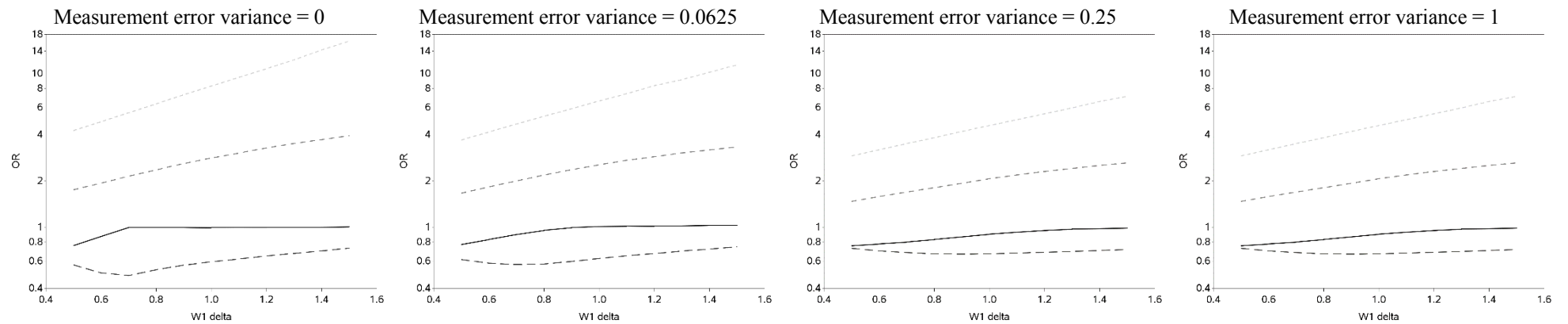
Figure A2. *Cont.*

Saturation (semi-linear) model

G. Weak true association
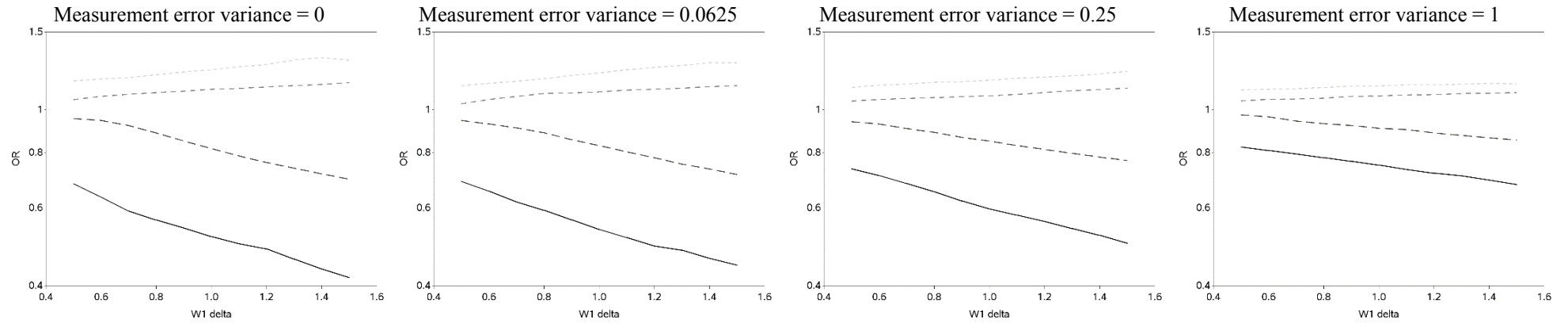
H. Moderate true association
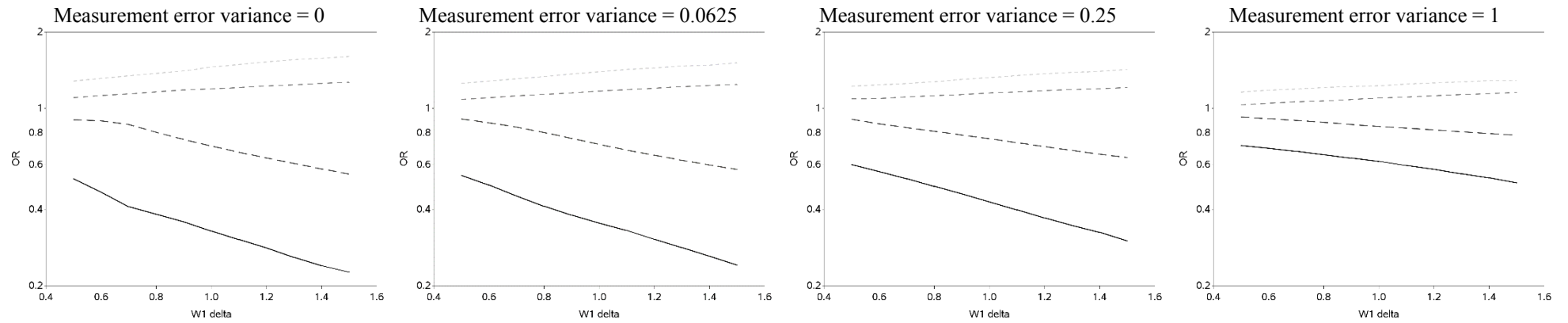
Figure A2. *Cont.*

I. Strong true association
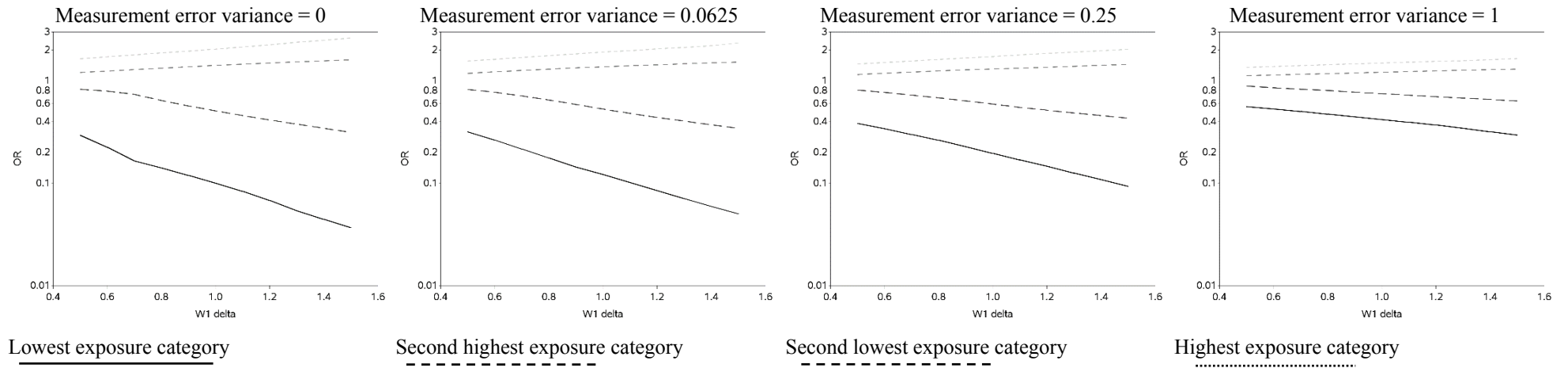
Figure A2. Odds ratios (OR) for different cutoffs used to create 5 categories for the causal yet mismeasuered exposure (W_1) (all scenarios for Figure 3) (δ is identified as "W1 delta" in the figure). The 4 W1 cutoffs were centered at 0 with δ from 0.5 to 1.5 standard deviations, in increments of 0.1. Reference category for W1 is mean $+/-$ ($0.5 \times \delta \times$ standard deviation).

A. Threshold (semi-linear) model

B. Saturation (semi-linear) model
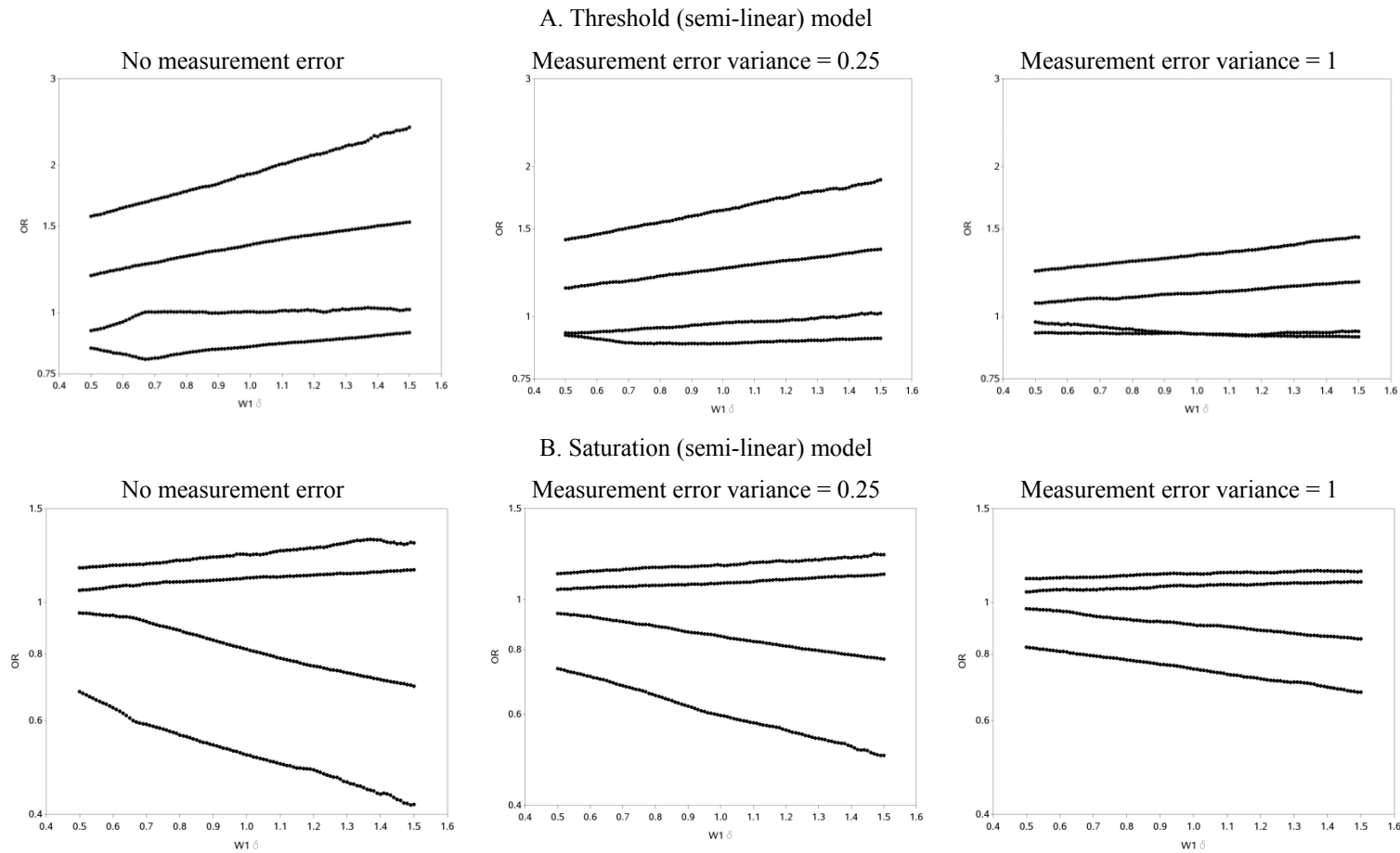
Figure A3. Odds ratios (OR) for different cutoffs used to create 5 categories for mismeasured exposure (W_1), 101 cutoffs, with the true "weak" association. The 4 W1 cutoffs were centered at 0 with δ from 0.5 to 1.5 standard deviations, in increments of 0.1. Reference category for W1 is mean +/−(0.5 × δ × standard deviation).

Linear model

A. Weak true association

B. Moderate true association

Figure A4. *Cont.*

C. Strong true association

Threshold model
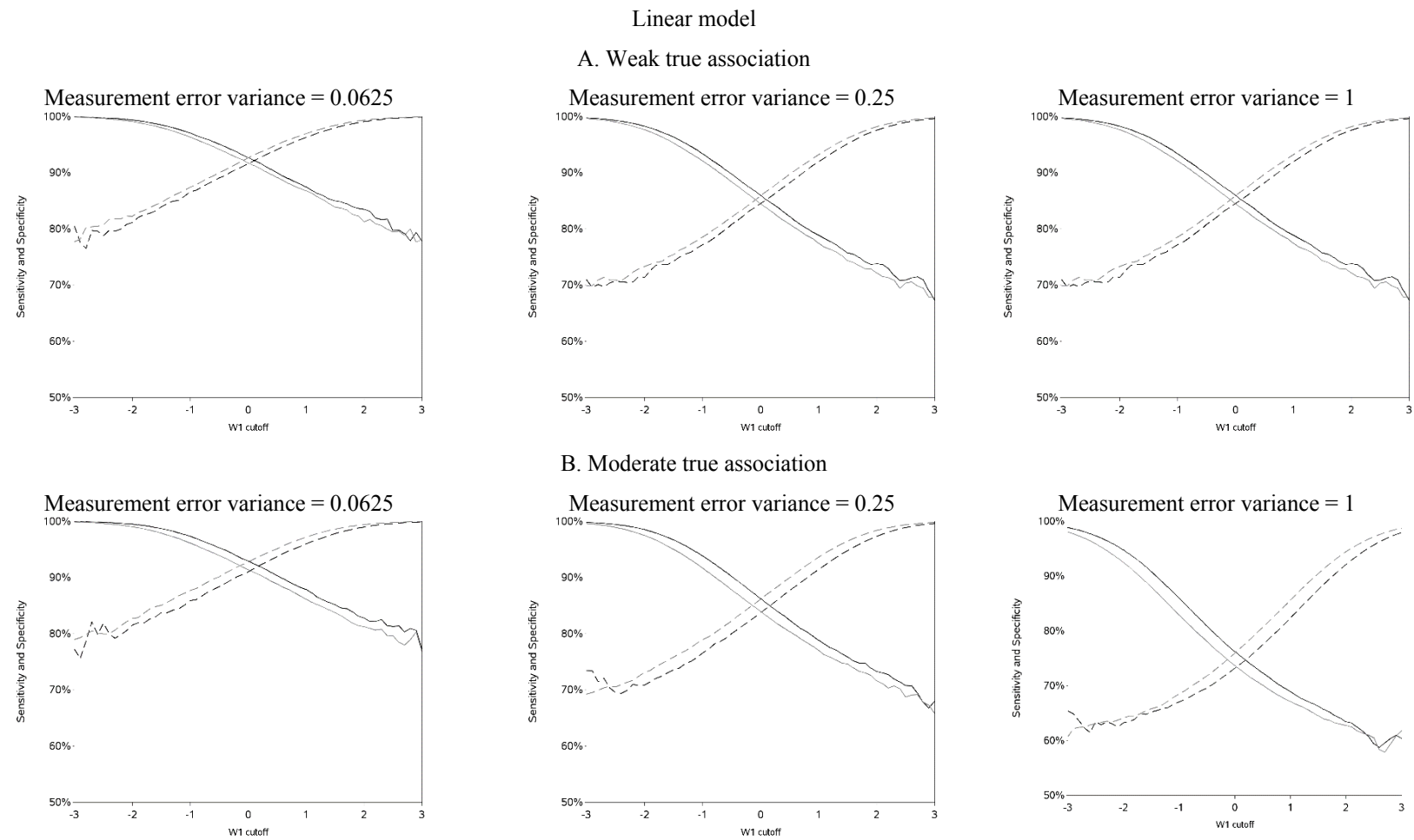
D. Weak true association
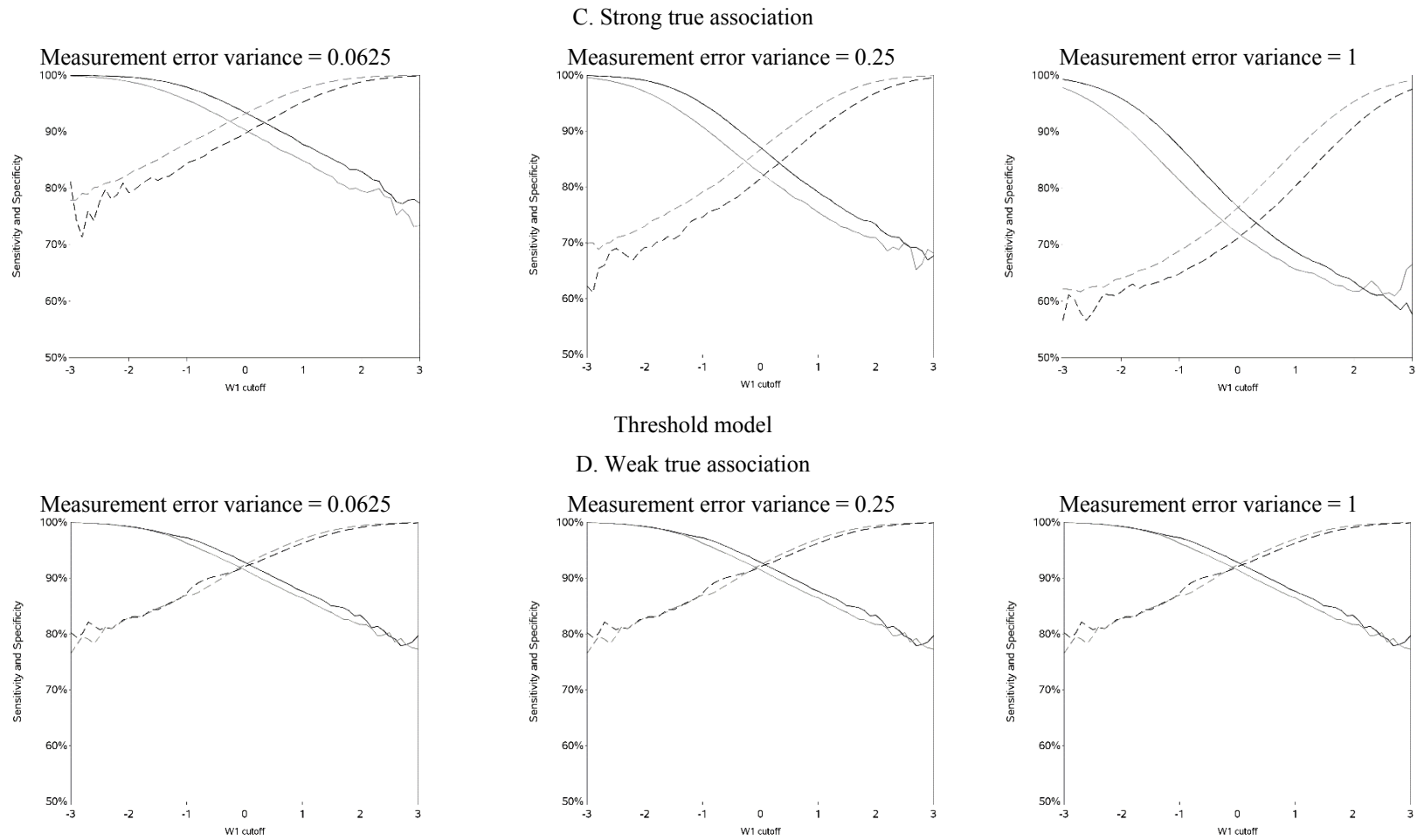
Figure A4. *Cont.*

E. Moderate true association

Measurement error variance = 0.0625   Measurement error variance = 0.25   Measurement error variance = 1

F. Strong true association

Measurement error variance = 0.0625   Measurement error variance = 0.25   Measurement error variance = 1

Figure A4. *Cont.*

Saturation model

G. Weak true association
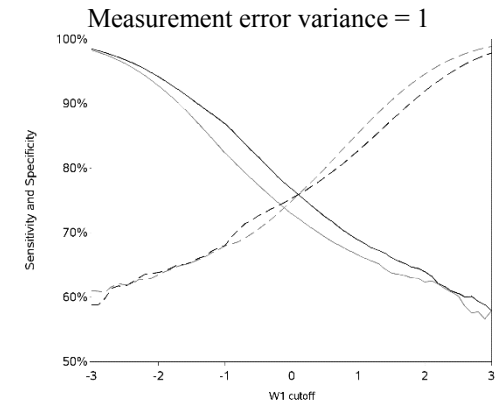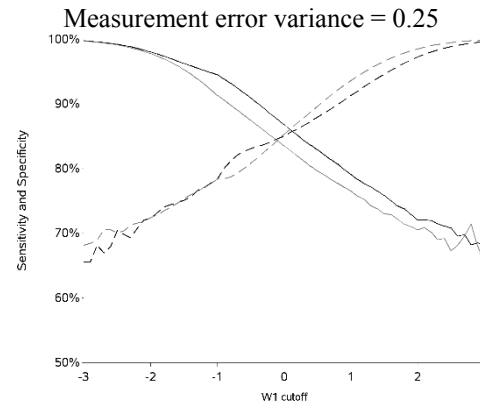
H. Moderate true association

Figure A4. *Cont.*

I. Strong true association

Figure A4. Sensitivity and specificity at each cutoff for W$_1$ (all scenarios for Figure 4). Cases—black; Controls—grey; Sensitivity—solid lines; Specificity—dotted lines.

Linear model

A. Weak true association
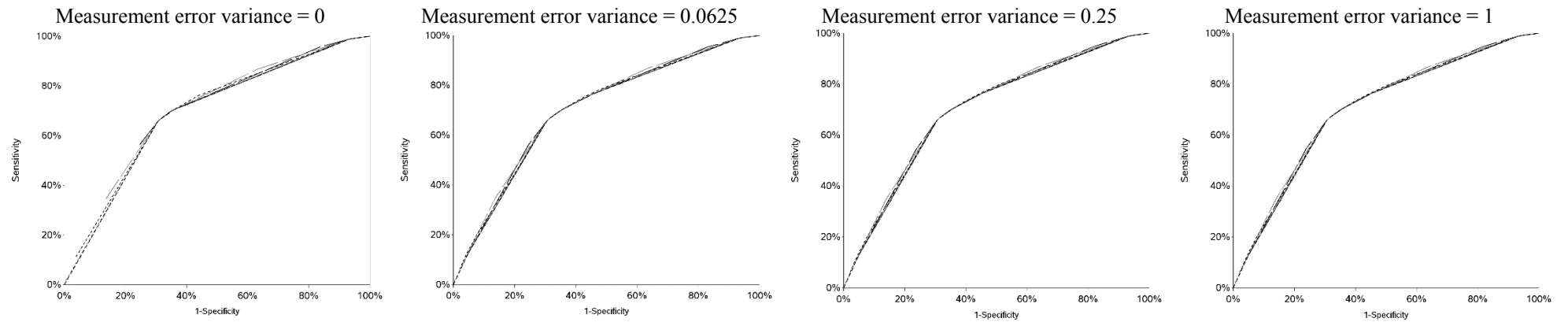
Figure A5. *Cont.*
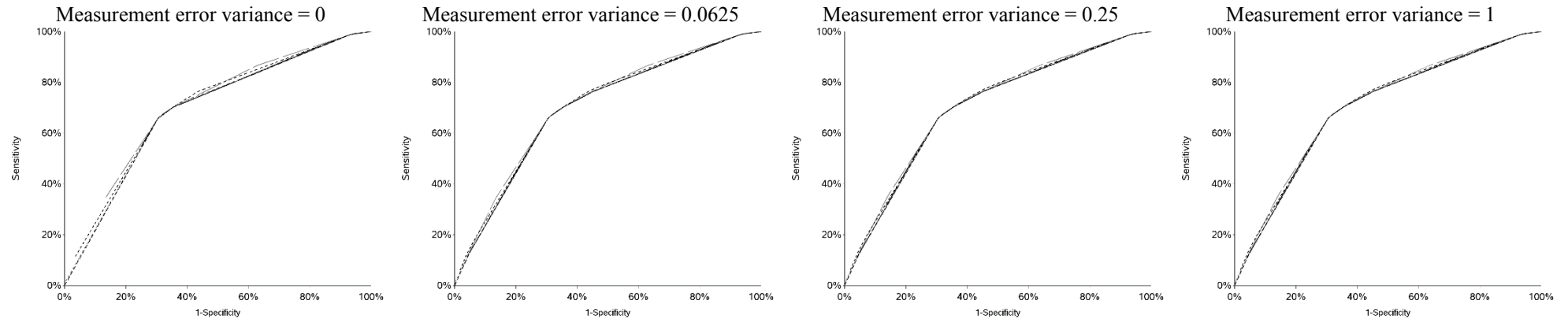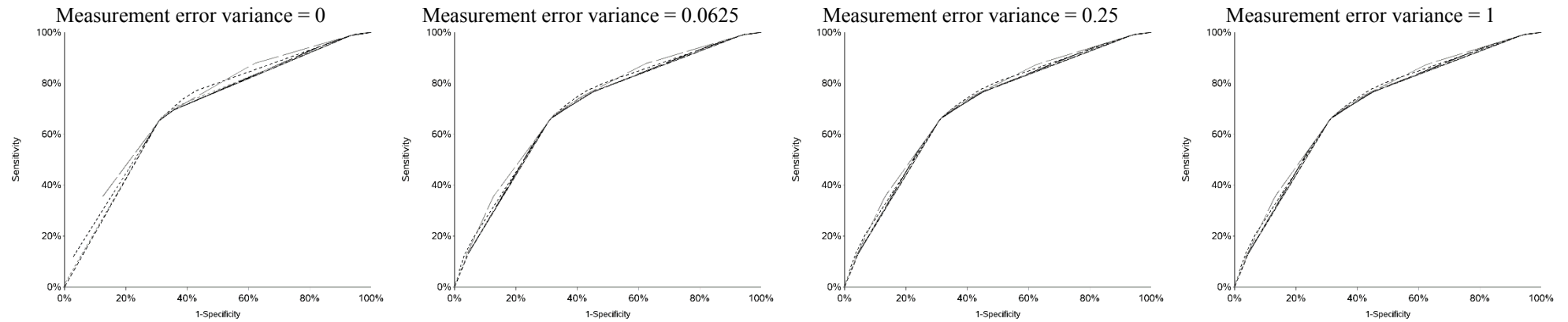
B. Moderate true association
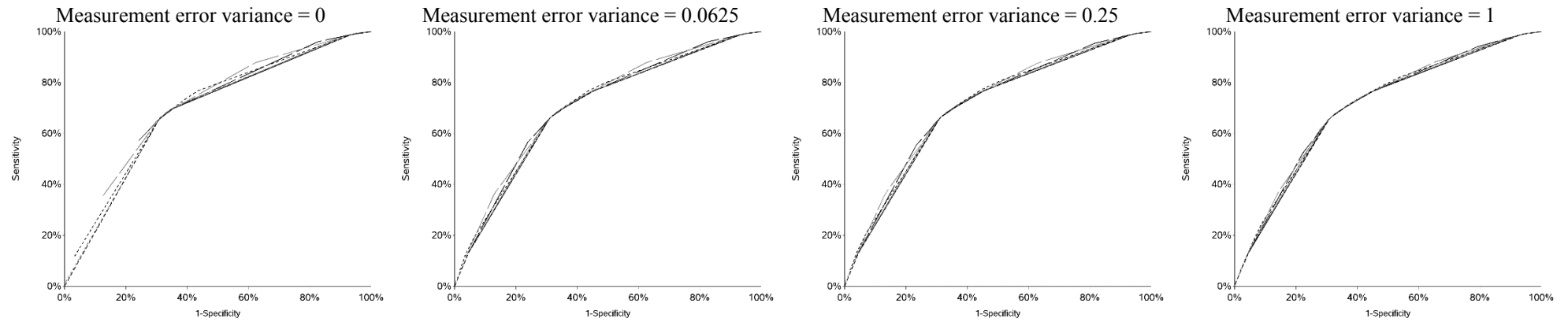
C. Strong true association
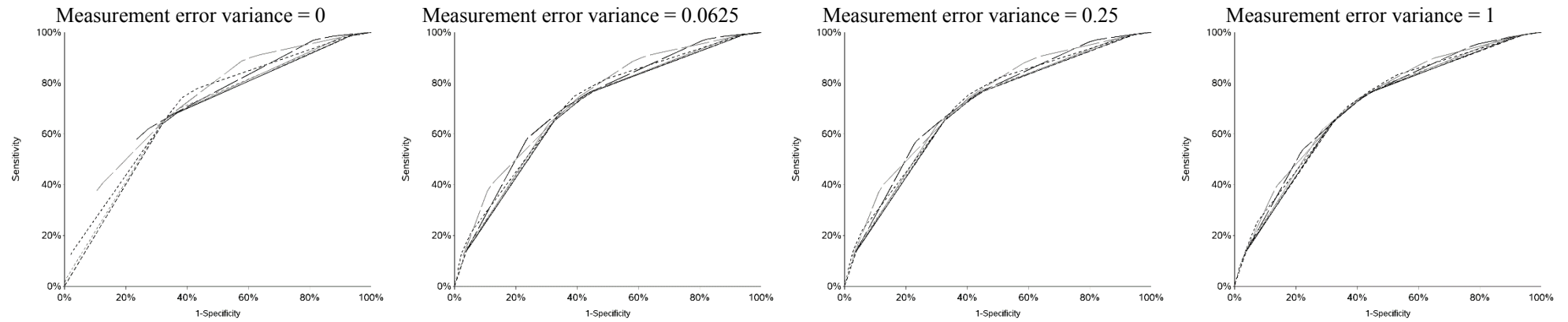
Figure A5. *Cont.*

Threshold model

A. Weak true association

B. Moderate true association

Figure A5. *Cont.*

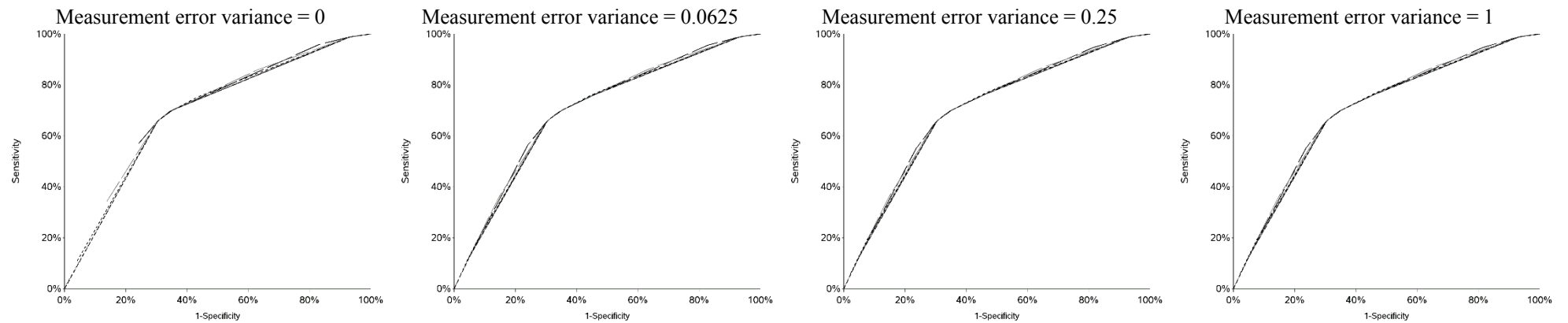C. Strong true association
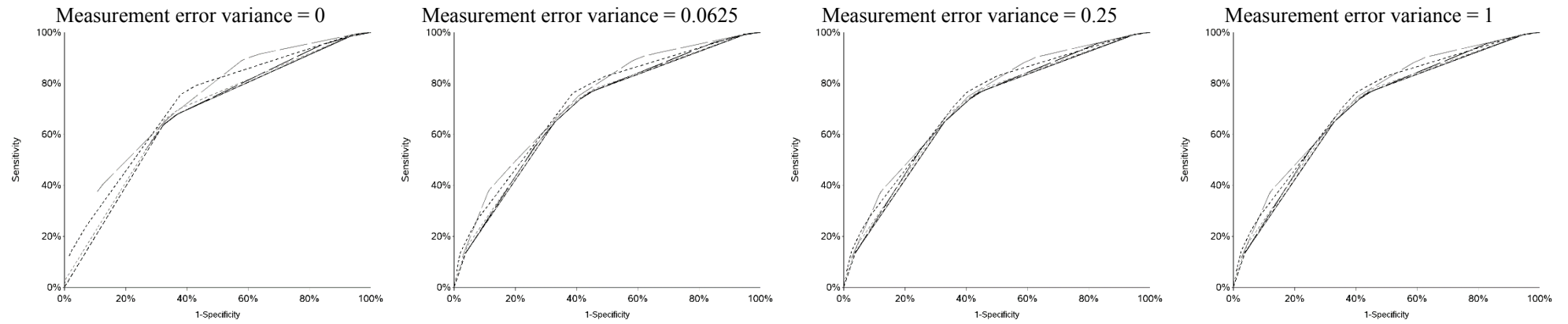
Saturation model

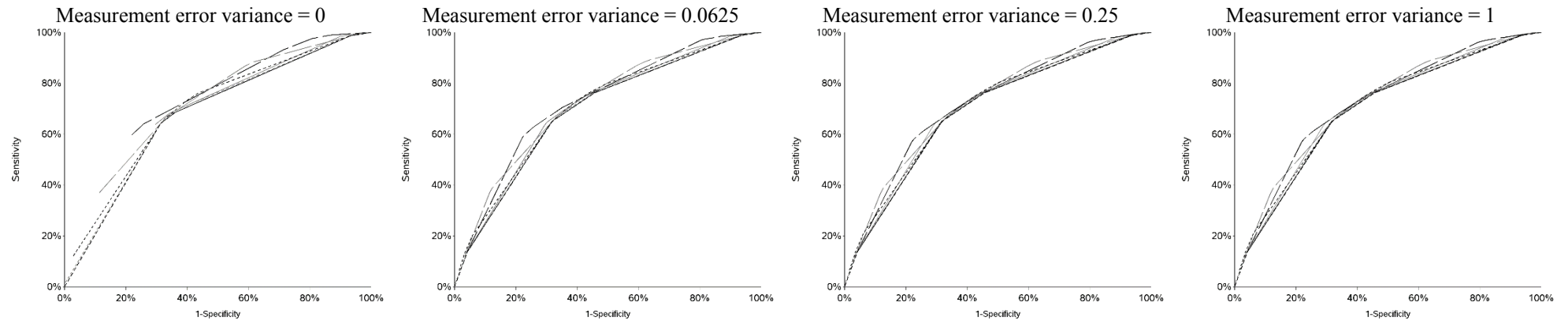A. Weak true association

Figure A5. *Cont.*
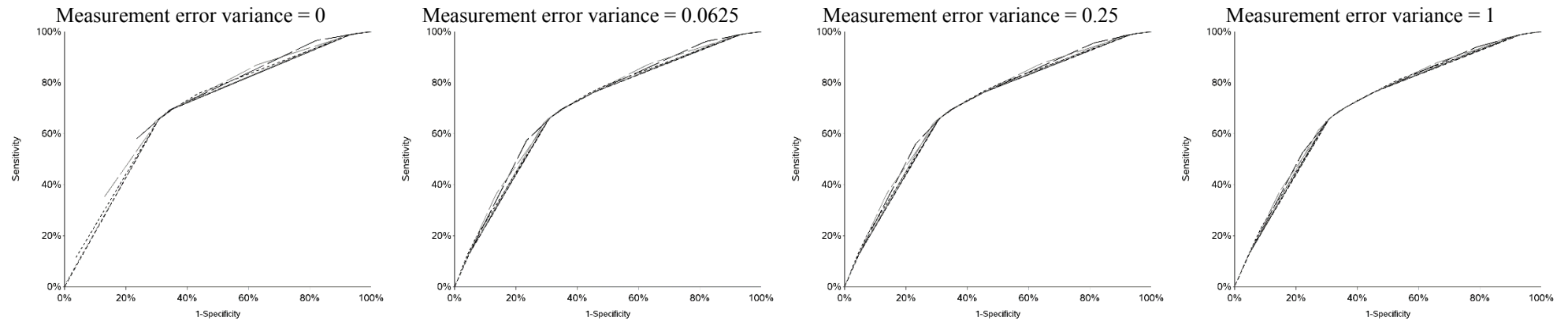
Figure A5. ROC curves for selected cutoffs for W_1.

References

1. Heavner, K.K.; Phillips, C.V.; Burstyn, I.; Hare, W. Dichotomization: 2×2 ($\times 2 \times 2 \times 2...$) categories: Infinite possibilities. *BMC Med. Res. Methodol.* **2010**, *10*, doi:10.1186/1471-2288-10-59.

2. Wartenberg, D.; Northridge, M. Defining exposure in case-control studies: A new approach. *Am. J. Epidemiol.* **1991**, *133*, 1058–1071.

3. Phillips, C.V. Publication bias *in situ*. *BMC Med. Res. Methodol.* **2004**, *4*, doi:10.1186/1471-2288-4-20.

4. Leavey, A.; Zwaigenbaum, L.; Heavner, K.; Burstyn, I. Gestational age at birth and risk of autism spectrum disorders in Alberta, Canada. *J. Pediatr.* **2013**, *162*, 361–368.

5. O'Brien, S.M. Cutpoint selection for categorizing a continuous predictor. *Biometrics* **2004**, *60*, 504–509.

6. Bennette, C.; Vickers, A. Against quantiles: Categorization of continuous variables in epidemiologic research, and its discontents. *BMC Med. Res. Methodol.* **2012**, *12*, doi:10.1186/1471-2288-12-21.

7. Greenland, S. Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. *Epidemiology* **1995**, *6*, 450–454.

8. ACOG Committee Opinion No. 579: Definition of term pregnancy. *Obstet. Gynecol.* **2013**, *122*, 1139–1140.

9. Spong, C.Y. Defining "term" pregnancy: Recommendations from the defining "term" pregnancy workgroup. *JAMA* **2013**, *309*, 2445–2446.

10. Carroll, R.J.; Ruppert, D.; Stefanski, L.A. *Measurement Error in Nonlinear Models*; Chapman and Hall Ltd.: London, UK, 1995.

11. Lovasi, G.S.; Underhill, L.J.; Jack, D.; Richards, C.; Weiss, C.; Rundle, A. At odds: Concerns raised by using odds ratios for continuous or common dichotomous outcomes in research on physical activity and obesity. *Open Epidemiol. J.* **2012**, *5*, 13–17.

12. Gustafson, P. *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*; Chapman and Hall/CRC Press: Boca Raton, FL, USA, 2003.

13. Flegal, K.M.; Keyl, P.M.; Nieto, F.J. Differential misclassification arising from nondifferential errors in exposure measurement. *Am. J. Epidemiol.* **1991**, *134*, 1233–1244.

14. Newschaffer, C.J.; Croen, L.A.; Fallin, M.D.; Hertz-Picciotto, I.; Nguyen, D.V.; Lee, N.L.; Berry, C.A.; Farzadegan, H.; Hess, H.N.; Landa, R.J.; *et al.* Infant siblings and the investigation of autism risk factors. *J. Neurodev. Disord.* **2012**, *4*, doi:10.1186/1866-1955-4-7.

15. Heavner, K.; Newschaffer, C.; Hertz-Picciotto, I.; Bennett, D.; Burstyn, I. Quantifying the potential impact of measurement error in an investigation of autism spectrum disorder (ASD). *J. Epidemiol. Comm. Health* **2014**, *68*, 438–445.

16. Aibar, L.; Puertas, A.; Valverde, M.; Carrillo, M.P.; Montoya, F. Fetal sex and perinatal outcomes. *J. Perinat. Med.* **2012**, *40*, 271–276.

17. Werling, D.M.; Geschwind, D.H. Sex differences in autism spectrum disorders. *Curr. Opin. Neurol.* **2013**, *26*, 146–153.

18. Gardener, H.; Spiegelman, D.; Buka, S.L. Perinatal and neonatal risk factors for autism: A comprehensive meta-analysis. *Pediatrics* **2011**, *128*, 344–355.

19. Kuzniewicz, M.W.; Wi, S.; Qian, Y.; Walsh, E.M.; Armstrong, M.A.; Croen, L.A. Prevalence and neonatal factors associated with autism spectrum disorders in preterm infants. *J. Pediatr.* **2014**, *164*, 20–25.

20. Tilghman, M.W.; Guerena, D.D.; Licea, A.; Perez-Santiago, J.; Richman, D.D.; May, S.; Smith, D.M. Pooled nucleic acid testing to detect antiretroviral treatment failure in Mexico. *J. Acquir. Immune. Defic. Syndr.* **2011**, *56*, e70–e74.

21. Bryson, S.E.; Zwaigenbaum, L.; McDermott, C.; Rombough, V.; Brian, J. The autism observation scale for infants: Scale development and reliability data. *J Autism Dev. Disord.* **2008**, *38*, 731–738.

22. Allodji, R.S.; Thiebaut, A.C.; Leuraud, K.; Rage, E.; Henry, S.; Laurier, D.; Benichou, J. The performance of functional methods for correcting non-gaussian measurement error within poisson regression: Corrected excess risk of lung cancer mortality in relation to radon exposure among French uranium miners. *Stat. Med.* **2012**, *31*, 4428–4443.

23. Phillips, C.V. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* **2003**, *14*, 459–466.

24. Hinkley, D. Inference about the intersection in two-phase regression. *Biometrika* **1969**, *56*, 495–504.

25. Muggeo, V.M. Estimating regression models with unknown break-points. *Stat. Med.* **2003**, *22*, 3055–3071.