Risk prediction models for discrete ordinal outcomes: calibration and the impact of the proportional odds assumption

Michael EDLINGER, Maarten VAN SMEDEN, Hannes F ALBER,

Maria WANITSCHEK, Ben VAN CALSTER [*] [†]

SUPPLEMENTARY MATERIAL

[*] Correspondence to: Ben Van Calster, Department of Development and Regeneration, KU Leuven, Herestraat 49 Box 805, 3000 Leuven, Belgium; ben.vancalster@kuleuven.be

TABLE OF CONTENTS

1. Explanation for the sample sizes used in the simulation study

We used large datasets of 200,000 simulated cases to approach true model coefficients and model performance. When evaluating at all 20 simulation scenarios, this sample size ensures a minimum of 10,000 cases in each outcome category and scenario. Knowing that, across all scenarios, the maximum number of parameters in a model is 18 (including intercepts). This implies that there events per parameter (EPP) in the smallest outcome category was at least 1,111 (Table A2). Note that EPP is calculated without taking intercepts into account.[34] This is very high, even for models with modest true discriminatory ability.

Table A2. Events per parameter (EPP) in the smallest outcome category by N, scenario, and model.

	EPP in smallest outcome category								
	N=200,000			N=500			N=100		
Scenario	**MLR**	**CL-PO/ AC-PO**	**SLM**	**MLR**	**CL-PO/ AC-PO**	**SLM**	**MLR**	**CL-PO/ AC-PO**	**SLM**
MLR 1	8,333	16,667	13,333	20.8	41.7	33.3	4.2	8.3	6.7
MLR 2	3,750	7,500	6,000	9.4	18.8	15.0	1.9	3.8	3.0
MLR 3	8,333	16,667	13,333	20.8	41.7	33.3	4.2	8.3	6.7
MLR 4	3,750	7,500	6,000	9.4	18.8	15.0	1.9	3.8	3.0
MLR 5	3,750	7,500	6,000	nd	nd	nd	nd	nd	nd
MLR 6	3,333	10,000	6,000	nd	nd	nd	nd	nd	nd
MLR 7	3,333	10,000	6,000	nd	nd	nd	nd	nd	nd
MLR 8	1,111	3,333	2,000	nd	nd	nd	nd	nd	nd
MLR 9	3,750	7,500	6,000	nd	nd	nd	nd	nd	nd
MLR 10	3,333	10,000	6,000	nd	nd	nd	nd	nd	nd
MLR 11	1,875	3,750	3,333	nd	nd	nd	nd	nd	nd
CL-PO 1	8,333	16,667	13,333	20.8	41.7	33.3	4.2	8.3	6.7
CL-PO 2	3,750	7,500	6,000	9.4	18.8	15.0	1.9	3.8	3.0
CL-PO 3	1,250	2,500	2,000	3.1	6.3	5.0	0.6	1.3	1.0
CL-PO 4	3,333	10,000	6,000	nd	nd	nd	nd	nd	nd
CL-PO 5	3,333	10,000	6,000	nd	nd	nd	nd	nd	nd
CL-PO 6	1,111	3,333	2,000	nd	nd	nd	nd	nd	nd
CL-PO 7	3,750	7,500	6,000	nd	nd	nd	nd	nd	nd
CL-PO 8	3,333	10,000	6,000	nd	nd	nd	nd	nd	nd
CL-PO 9	4,167	8,333	7,407	nd	nd	nd	nd	nd	nd

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model; nd, not done.

To investigate overfitting in the 7 main scenarios, we need datasets where the events per parameter is too low. We sampled datasets of size 100 and 500. When sample size is 100, the EPP in the smallest outcome category varied between 0.6 and 8.3 depending on outcome category and scenario. These values are low or very low by all standards. When the sample size is 500, the EPP in the smallest outcome category varied between 3.1 and 41.7. These values are low to acceptable.

2. Example R code

```
library(VGAM)

# CODE TO FIT DIFFERENT MODELS

mlr <- vglm(y ~ x1 + x2, family=multinomial(refLevel = "1"), data=dataset)
 # Multinomial logistic regression; alternative: multinom (nnet package)

clpo <- vglm(y ~ x1 + x2, family=cumulative(parallel=T, reverse=T), data=dataset)
 # Cumulative logit model with proportional odds; alternatives: polr (MASS package), clm (ordinal
   package), orm (rms package)
 # If you do not specify reverse = T, it focuses on Y ≤ k instead of Y ≥ k

acp <- vglm(y ~ x1 + x2, family=acat(parallel=T), data=dataset)
 # Adjacent category logit model with proportional odds

crp <- vglm(y ~ x1 + x2, family=cratio(parallel=T), data=dataset)
 # Continuation ratio logit model with proportional odds

crnp <- vglm(y ~ x1 + x2, family=cratio(parallel=F), data=dataset)
 # Continuation ratio logit model without proportional odds

slm=rrvglm(y ~ x1 + x2, multinomial(refLevel = "1"), data = dataset)
 # Stereotype logistic model


# CALCULATE ESTIMATED PROBABILITIES AND LINEAR PREDICTORS (MLR ONLY)

mlrpred <- predictvglm(mlr,newdata=cad,type="response")
mlrlpred <- predictvglm(mlr,newdata=cad,type="link")


# CALIBRATION INTERCEPT AND SLOPE FOR EACH OUTCOME CATEGORY

calout <- function(out,preds,k){
  cores = matrix(0,k,2)
  for (i in (1:k)){
    cores[i,1] = glm(out==i ~ 1, offset=logit(preds[,i]), family=binomial)$coefficients
    cores[i,2] = glm(out==i ~ logit(preds[,i]), family=binomial)$coefficients[2]
  }
  return(cores)
}
mlrcalout = calout(out=dataset$y,preds=mlrpred,5) # MLR, 5 CATEGORIES


# CALIBRATION INTERCEPT AND SLOPE FOR EACH OUTCOME DICHOTOMY

caldout <- function(out,preds,k){
  cores = matrix(0,k-1,2)
  for (i in (2:(k-1))){
    cores[i-1,1] = glm(out>=i ~ 1, offset=logit(rowSums(preds[,i:k])),
                       family=binomial)$coefficients
    cores[i-1,2] = glm(out>=i ~ logit(rowSums(preds[,i:k])), family=binomial)$coefficients[2]
  }
  cores[k-1,1] = glm(out>=k ~ 1, offset=logit(preds[,k]), family=binomial)$coefficients
  cores[k-1,2] = glm(out>=k ~ logit(preds[,k]),
                     family=binomial)$coefficients[2]
  return(cores)
}
mlrcaldout = caldout(out=dataset$y,preds=mlrpred,k=5) # MLR, 5 CATEGORIES


# MODEL SPECIFIC CALIBRATION INTERCEPTS AND SLOPES (MLR ONLY, 5 OUTCOME CATEGORIES)

mlrrecali <- coefficients(vglm(dataset$y ~ 1, offset = mlrlpred[,1:4],
                               family=multinomial(refLevel = "1")))[c(1:4)]
mlrrecals <- coefficients(vglm(dataset$y ~ mlrlpred[,1] + mlrlpred[,2] +
                                           mlrlpred[,3] + mlrlpred[,4],
```

```
                                constraints=list("(Intercept)"=diag(4),
                                                 "mlrlpred[, 1]"=rbind(1,0,0,0),
                                                 "mlrlpred[, 2]"=rbind(0,1,0,0),
                                                 "mlrlpred[, 3]"=rbind(0,0,1,0),
                                                 "mlrlpred[, 4]"=rbind(0,0,0,1)),
                                family=multinomial(refLevel = "1")))[c(5:8)]


# FLEXIBLE RECALIBRATION MODEL (MLR ONLY, 5 OUTCOME CATEGORIES)

mlrlp1=log(mlrpred[,2]/mlrpred[,1])
mlrlp2=log(mlrpred[,3]/mlrpred[,1])
mlrlp3=log(mlrpred[,4]/mlrpred[,1])
mlrlp4=log(mlrpred[,5]/mlrpred[,1])
mlrvgamsmps4 = vgam(dataset$y ~ sm.ps(mlrlp1,df=4) + sm.ps(c(mlrlp2),df=4) +
                                 sm.ps(c(mlrlp3),df=4) + sm.ps(c(mlrlp4),df=4),
                    family=multinomial(refLevel = "1"))


# ECI, ORIGINAL FORMULA FROM VAN HOORDE ET AL (J BIOMED INFORM 2015)

eci_bvc <- function(calout,preds,k){
  (mean((preds-fitted(calout))*(preds-fitted(calout))))*(100*k/2)
}
mlrECI = eci_bvc(calout=mlrvgamsmps4,preds=mlrpred,k=5) # MLR, 5 CATEGORIES


# ECI, ADAPTED FORMULA TO COMPARE WITH RANDOM MODEL – VERSION USED IN THIS PAPER

eci_rel <- function(calout,preds,k,outc){
  prevm=matrix((table(outc)/length(outc))[1:k],nrow=dim(preds)[1],ncol=k,byrow=T)
  ecir=mean((preds-prevm)*(preds-prevm))
  ecim=mean((preds-fitted(calout))*(preds-fitted(calout)))
  return(ecim/ecir)
}
mlrECIr = eci_rel(calout=mlrvgamsmps4,preds=mlrpred,k=5,outc=cad$o5) # MLR, 5 CATEGORIES


# CALIBRATION SCATTER PLOT PER OUTCOME CATEGORY BASED ON FLEXIBLE RECALIBRATION MODEL
# (MLR, 5 CATEGORIES)

plot(preds[,1],fitted(obs)[,1],type="p",pch=1,col="green",lwd=1,
     ylab="Observed proportion",xlab="Estimated probability", xlim=0:1 ,ylim=0:1)
points(preds[,2],fitted(obs)[,2],type="p",pch=1,col="orange")
points(preds[,3],fitted(obs)[,3],type="p",pch=1,col="red")
points(preds[,4],fitted(obs)[,4],type="p",pch=1,col="brown")
points(preds[,5],fitted(obs)[,5],type="p",pch=1,col="black")
lines(c(0,1),c(0,1),type="l",col="gray",lty=3) # plot the ideal diagonal line


# CALIBRATION CURVES BASED PER OUTCOME CATEGORY ON FLEXIBLE RECALIBRATION MODEL
# (MLR, 5 CATEGORIES)

wa1=smooth.spline(preds[,1],fitted(obs)[,1])
wa2=smooth.spline(preds[,2],fitted(obs)[,2])
wa3=smooth.spline(preds[,3],fitted(obs)[,3])
wa4=smooth.spline(preds[,4],fitted(obs)[,4])
wa5=smooth.spline(preds[,5],fitted(obs)[,5])
plot(wa1$x, wa1$y,type="l",col="green",ylab="Observed proportion",
     xlab="Estimated probability",xlim=0:1,ylim=0:1)
lines(wa2$x, wa2$y,col="orange")
lines(wa3$x, wa3$y,col="red")
lines(wa4$x, wa4$y,col="brown")
lines(wa5$x, wa5$y,col="black")
lines(c(0,1),c(0,1),type="l",col="gray",lty=3) # plot the ideal diagonal line


# CALIBRATION SCATTER PLOT PER OUTCOME DICHOTOMY BASED ON FLEXIBLE RECALIBRATION MODEL
# (MLR, 5 CATEGORIES)

plot(preds[,2]+preds[,3]+preds[,4]+preds[,5],
```

```
     fitted(obs)[,2]+fitted(obs)[,3]+fitted(obs)[,4]+fitted(obs)[,5],
     type="p",pch=1,col="orange",lwd=1,ylab="Observed proportion",
     xlab="Estimated probability",xlim=0:1,ylim=0:1)
points(preds[,3]+preds[,4]+preds[,5],fitted(obs)[,3]+fitted(obs)[,4]+fitted(obs)[,5],
       type="p",pch=1,col="red")
points(preds[,4]+preds[,5],fitted(obs)[,4]+fitted(obs)[,5],type="p",pch=1,col="brown")
points(preds[,5],fitted(obs)[,5],type="p",pch=1,col="black")
lines(c(0,1),c(0,1),type="l",col="gray",lty=3) # plot the ideal diagonal line


# CALIBRATION CURVES BASED PER OUTCOME CATEGORY ON FLEXIBLE RECALIBRATION MODEL
# (MLR, 5 CATEGORIES)

wa2=smooth.spline(preds[,2]+preds[,3]+preds[,4]+preds[,5],
                    fitted(obs)[,2]+fitted(obs)[,3]+fitted(obs)[,4]+fitted(obs)[,5])
wa3=smooth.spline(preds[,3]+preds[,4]+preds[,5],fitted(obs)[,3]+fitted(obs)[,4]+fitted(obs)[,5])
wa4=smooth.spline(preds[,4]+preds[,5],fitted(obs)[,4]+fitted(obs)[,5])
wa5=smooth.spline(preds[,5],fitted(obs)[,5])
plot(wa2$x, wa2$y,type="l",col="orange", ylab="Observed proportion",xlab="Estimated probability",
     xlim=0:1,ylim=0:1)
lines(wa3$x, wa3$y,col="red")
lines(wa4$x, wa4$y,col="brown")
lines(wa5$x, wa5$y,col="black")
lines(ref,ref,type="l",col="gray",lty=3) # plot the ideal diagonal line


# ORDINAL C-STATISTIC (ORC)

orc <- function(out,preds,k){
  library(DescTools) # Cstat
  Ec=preds
  for (i in (1:k)){
    Ec[,i]=i*Ec[,i]
  }
  E=rowSums(Ec)
  pwc = rep(NA,k*(k-1)*0.5)
  for (i in (2:k)){
    for (j in (1:(i-1))){
      pwc[((i-1)*(i-2)*0.5)+j]=Cstat(x=E[out==(j) | out==i],resp=out[out==(j) | out==i]==i) # c
statistic 1 vs 2
    }
  }
  mean(pwc)
}
mlrc = orc(dataset$y,mlrpred,5) # MLR ONLY, 5 CATEGORIES
```

3. Comparison of flexible recalibration models

The flexible recalibration model that we used in this work was based on the MLR framework (Equation 18). This may disadvantage the resulting calibration plots, and derived measures such as the ECI, for prediction models that were based on another type of model. It is perhaps impossible to propose a recalibration model that is fully model agnostic, but we can compare different setups to evaluate their impact on the results. The recalibration model should in any case not make a proportional odds assumption. We evaluated 6 alternative flexible recalibration models, by varying whether the model has an MLR or CR-NP setup and whether the $K - 1$ linear predictors compare every category with a reference category, compare every dichotomy, or compare a category with their complement (i.e. all other categories combined):

MLR, reference LP: $\log\left(\frac{P(Y=k)}{P(Y=1)}\right) = a_{flex,k} + \sum_{j=2}^{K} s_{k,j}(\hat{Z}_j)$, $k = 2, \dots, K$, $\hat{Z}_j = \log(\hat{P}_j/\hat{P}_1)$

CR-NP, reference LP: $\log\left(\frac{P(Y>k)}{P(Y\geq k)}\right) = a_{flex,k}^{\sim} + \sum_{j=2}^{K} s_{k,j}^{\sim}(\hat{Z}_j)$, $k = 1, \dots, K-1$, $\hat{Z}_j = \log(\hat{P}_j/\hat{P}_1)$

MLR, dichotomy LP: $\log\left(\frac{P(Y=k)}{P(Y=1)}\right) = a_{flex,k}' + \sum_{j=2}^{K} s_{k,j}'(\widehat{D}_j)$, $k = 2, \dots, K$, $\widehat{D}_j = \text{logit}(\hat{V}_j)$

CR-NP, dichotomy LP: $\log\left(\frac{P(Y>k)}{P(Y\geq k)}\right) = a_{flex,k}'' + \sum_{j=2}^{K} s_{k,j}''(\widehat{D}_j)$, $k = 1, \dots, K-1$, $\widehat{D}_j = \text{logit}(\hat{V}_j)$

MLR, category LP: $\log\left(\frac{P(Y=k)}{P(Y=1)}\right) = a_{flex,k}^{*} + \sum_{j=1}^{K-1} s_{k,j}^{*}(\hat{C}_j)$, $k = 2, \dots, K$, $\hat{C}_j = \log\left(\hat{P}_j/(1-\hat{P}_j)\right)$

CR-NP, category LP: $\log\left(\frac{P(Y>k)}{P(Y\geq k)}\right) = a_{flex,k}^{**} + \sum_{j=1}^{K-1} s_{k,j}^{**}(\hat{C}_j)$, $k = 1, \dots, K-1$, $\hat{C}_j = \log\left(\hat{P}_j/(1-\hat{P}_j)\right)$

Th first option is equal to Equation 18 from the main paper. Note that MLR equals the adjacent category approach without proportional odds (AC-NP). We did not include CL-NP models, because these may lead to invalid models. Indeed, when we tried to fit flexible recalibration models of the CL-NP type, we nearly always received error messages (for each of the three types of linear predictors). We applied these models to all simulation scenarios (using the large sample datasets) and the case study. We summarized calibration using the ECI (Tables A3.1-3). Differences between results for the six recalibration models were small, although linear predictors for a category vs its complement appear less appealing. Further, we constructed calibration scatter plots. This yielded 516 plots (480 for the simulation study and 36 for the case study). We only show plots for the MLR and CL-PO models from the case study. These plots confirm that differences between the six recalibration models were small. If the approach used in the main paper favors MLR models, the advantage is marginal.

Table A3.1. ECI values on large sample simulated datasets (n=200,000) using MLR truth.

Simulation scenario	**Model**	**MLR Ref LP**	**CR-NP Ref LP**	**MLR Dich LP**	**CR-NP Dich LP**	**MLR Cat LP**	**CR-NP Cat LP**
MLR 1	MLR	0.0000	0.0001	0.0000	0.0001	0.0001	0.0001
	CL-PO	0.0057	0.0052	0.0060	0.0055	0.0052	0.0056
	AC-PO	0.0000	0.0001	0.0000	0.0001	0.0001	0.0001
	SLM	0.0000	0.0001	0.0000	0.0001	0.0001	0.0001
MLR 2	MLR	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
	CL-PO	0.0098	0.0090	0.0099	0.0096	0.0097	0.0098
	AC-PO	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000
	SLM	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
MLR 3	MLR	0.0000	0.0009	0.0008	0.0007	0.0012	0.0004
	CL-PO	0.0493	0.0486	0.0490	0.0484	0.0483	0.0485
	AC-PO	0.0457	0.0455	0.0462	0.0459	0.0455	0.0460
	SLM	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000
MLR 4	MLR	0.0000	0.0004	0.0008	0.0004	0.0008	0.0003
	CL-PO	0.0324	0.0325	0.0326	0.0322	0.0326	0.0321
	AC-PO	0.0587	0.0583	0.0588	0.0584	0.0585	0.0584
	SLM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MLR 5	MLR	0.0000	0.0019	0.0101	0.0006	0.0042	0.0001
	CL-PO	0.0128	0.0126	0.0130	0.0125	0.0127	0.0124
	AC-PO	0.0178	0.0175	0.0182	0.0176	0.0175	0.0176
	SLM	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000
MLR 6	MLR	0.0000	0.0032	0.0165	0.0004	0.0061	0.0002
	CL-PO	0.1185	0.1154	0.1184	0.1184	0.1188	0.1187
	AC-PO	0.0961	0.0957	0.0926	0.0963	0.0960	0.0963
	SLM	0.0000	0.0000	0.0001	0.0001	0.0000	0.0001
MLR 7	MLR	0.0000	0.0018	0.0074	0.0009	0.0042	0.0002
	CL-PO	0.1560	0.1502	0.1560	0.1560	0.1556	0.1561
	AC-PO	0.1163	0.1164	0.1117	0.1169	0.1166	0.1169
	SLM	0.0000	0.0001	0.0003	0.0001	0.0001	0.0001
MLR 8	MLR	0.0000	0.0017	0.0096	0.0011	0.0043	0.0002
	CL-PO	0.1845	0.1772	0.1841	0.1845	0.1838	0.1846
	AC-PO	0.1584	0.1598	0.1551	0.1601	0.1600	0.1601
	SLM	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001
MLR 9	MLR	0.0000	0.0003	0.0002	0.0002	0.0003	0.0001
	CL-PO	0.1060	0.1073	0.0986	0.1062	0.1082	0.1063
	AC-PO	0.1448	0.1408	0.1435	0.1423	0.1407	0.1426
	SLM	0.0480	0.0587	0.0558	0.0416	0.0588	0.0464
MLR 10	MLR	0.0000	0.0001	0.0001	0.0002	0.0003	0.0002
	CL-PO	0.2362	0.2531	0.2275	0.2267	0.2526	0.2226
	AC-PO	0.2178	0.2298	0.2145	0.2023	0.2318	0.1714
	SLM	0.1482	0.1512	0.1478	0.1094	0.1475	0.1307
MLR 11	MLR	0.0000	0.0004	0.0008	0.0004	0.0008	0.0002
	CL-PO	0.0324	0.0325	0.0326	0.0322	0.0326	0.0321
	AC-PO	0.0587	0.0583	0.0588	0.0584	0.0585	0.0584
	SLM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mean	MLR	0.0000	0.0010	0.0042	0.0005	0.0020	0.0002
	CL-PO	0.0858	0.0858	0.0843	0.0847	0.0873	0.0844
	AC-PO	0.0831	0.0838	0.0818	0.0817	0.0841	0.0789
	SLM	0.0178	0.0191	0.0186	0.0138	0.0188	0.0161

ECI, estimated calibration index; MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model; CR-NP, continuation ration logit model without proportional odds; Ref LP, linear predictors using a reference category; Dich LP, linear predictors based on dichotomize the outcome; Cat LP, linear predictors based on one category vs all other categories.

Table A3.2. ECI values on large sample simulated datasets (n=200,000) using CL-PO truth.

Simulation scenario	Model	MLR Ref LP	CR-NP Ref LP	MLR Dich LP	CR-NP Dich LP	MLR Cat LP	CR-NP Cat LP
CLPO 1	MLR	0.0057	0.0068	0.0055	0.0073	0.0072	0.0070
	CL-PO	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001
	AC-PO	0.0058	0.0068	0.0055	0.0072	0.0071	0.0069
	SLM	0.0058	0.0068	0.0055	0.0072	0.0070	0.0069
CLPO 2	MLR	0.0054	0.0064	0.0057	0.0067	0.0068	0.0065
	CL-PO	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	AC-PO	0.0122	0.0135	0.0128	0.0137	0.0132	0.0135
	SLM	0.0057	0.0063	0.0057	0.0066	0.0064	0.0064
CLPO 3	MLR	0.0038	0.0042	0.0039	0.0045	0.0042	0.0044
	CL-PO	0.0001	0.0000	0.0001	0.0002	0.0001	0.0003
	AC-PO	0.0153	0.0156	0.0160	0.0159	0.0155	0.0158
	SLM	0.0038	0.0042	0.0024	0.0045	0.0042	0.0044
CLPO 4	MLR	0.0069	0.0088	0.0067	0.0091	0.0090	0.0091
	CL-PO	0.0000	0.0001	0.0000	0.0000	0.0002	0.0000
	AC-PO	0.0172	0.0190	0.0169	0.0189	0.0187	0.0190
	SLM	0.0076	0.0088	0.0069	0.0090	0.0087	0.0090
CLPO 5	MLR	0.0036	0.0055	0.0041	0.0056	0.0055	0.0056
	CL-PO	0.0001	0.0002	0.0001	0.0001	0.0003	0.0001
	AC-PO	0.0153	0.0181	0.0169	0.0183	0.0183	0.0182
	SLM	0.0039	0.0055	0.0038	0.0056	0.0055	0.0055
CLPO 6	MLR	0.0030	0.0043	0.0027	0.0046	0.0044	0.0045
	CL-PO	0.0000	0.0002	0.0000	0.0001	0.0003	0.0001
	AC-PO	0.0244	0.0259	0.0251	0.0258	0.0259	0.0258
	SLM	0.0031	0.0045	0.0029	0.0045	0.0045	0.0044
CLPO 7	MLR	0.0046	0.0050	0.0043	0.0050	0.0050	0.0051
	CL-PO	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	AC-PO	0.0126	0.0130	0.0123	0.0130	0.0127	0.0130
	SLM	0.0047	0.0052	0.0042	0.0050	0.0051	0.0050
CLPO 8	MLR	0.0075	0.0081	0.0073	0.0079	0.0082	0.0078
	CL-PO	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	AC-PO	0.0179	0.0181	0.0171	0.0181	0.0180	0.0181
	SLM	0.0080	0.0082	0.0078	0.0083	0.0082	0.0082
CLPO 9	MLR	0.0057	0.0068	0.0054	0.0073	0.0073	0.0070
	CL-PO	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001
	AC-PO	0.0058	0.0068	0.0055	0.0072	0.0071	0.0069
	SLM	0.0058	0.0068	0.0055	0.0072	0.0070	0.0069
Mean	MLR	0.0051	0.0062	0.0051	0.0064	0.0064	0.0063
	CL-PO	0.0001	0.0001	0.0001	0.0001	0.0002	0.0001
	AC-PO	0.0141	0.0152	0.0142	0.0153	0.0152	0.0152
	SLM	0.0054	0.0063	0.0050	0.0064	0.0063	0.0063

ECI, estimated calibration index; MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model; CR-NP, continuation ration logit model without proportional odds; Ref LP, linear predictors using a reference category; Dich LP, linear predictors based on dichotomize the outcome; Cat LP, linear predictors based on one category vs all other categories.

Table A3.3. ECI values for the case study (n=4884, apparent validation).

Model	**MLR Ref LP**	**CR-NP Ref LP**	**MLR Dich LP**	**CR-NP Dich LP**	**MLR Cat LP**	**CR-NP Cat LP**
MLR	0.0052	0.0078	0.0047	0.0068	0.0152	0.0124
CL-PO	0.0295	0.0328	0.0286	0.0314	0.0319	0.0335
AC-PO	0.1412	0.1490	0.1462	0.1479	0.1474	0.1513
CR-PO	0.1939	0.2008	0.1844	0.1826	0.1868	0.1896
CR-NP	0.0052	0.0056	0.0095	0.0067	18.3*	0.0154
SLM	0.0044	0.0061	0.0090	0.0069	0.0084	0.0060

* Warning that “fitted probabilities numerically 0 or 1 occurred”.

ECI, estimated calibration index; MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model; CR-PO, continuation ration logit model with proportional odds; CR-NP, continuation ration logit model without proportional odds; Ref LP, linear predictors using a reference category; Dich LP, linear predictors based on dichotomize the outcome; Cat LP, linear predictors based on one category vs all other categories.

Figure A3.1. Calibration scatter plots based on six different flexible recalibration models for the MLR model in the case study. Green: category 1 (no coronary artery disease); orange: category 2 (non-obstructive stenosis); red: category 3 (1-vessel disease); brown: category 4 (2-vessel disease); black: category 5 (3-vessel disease).

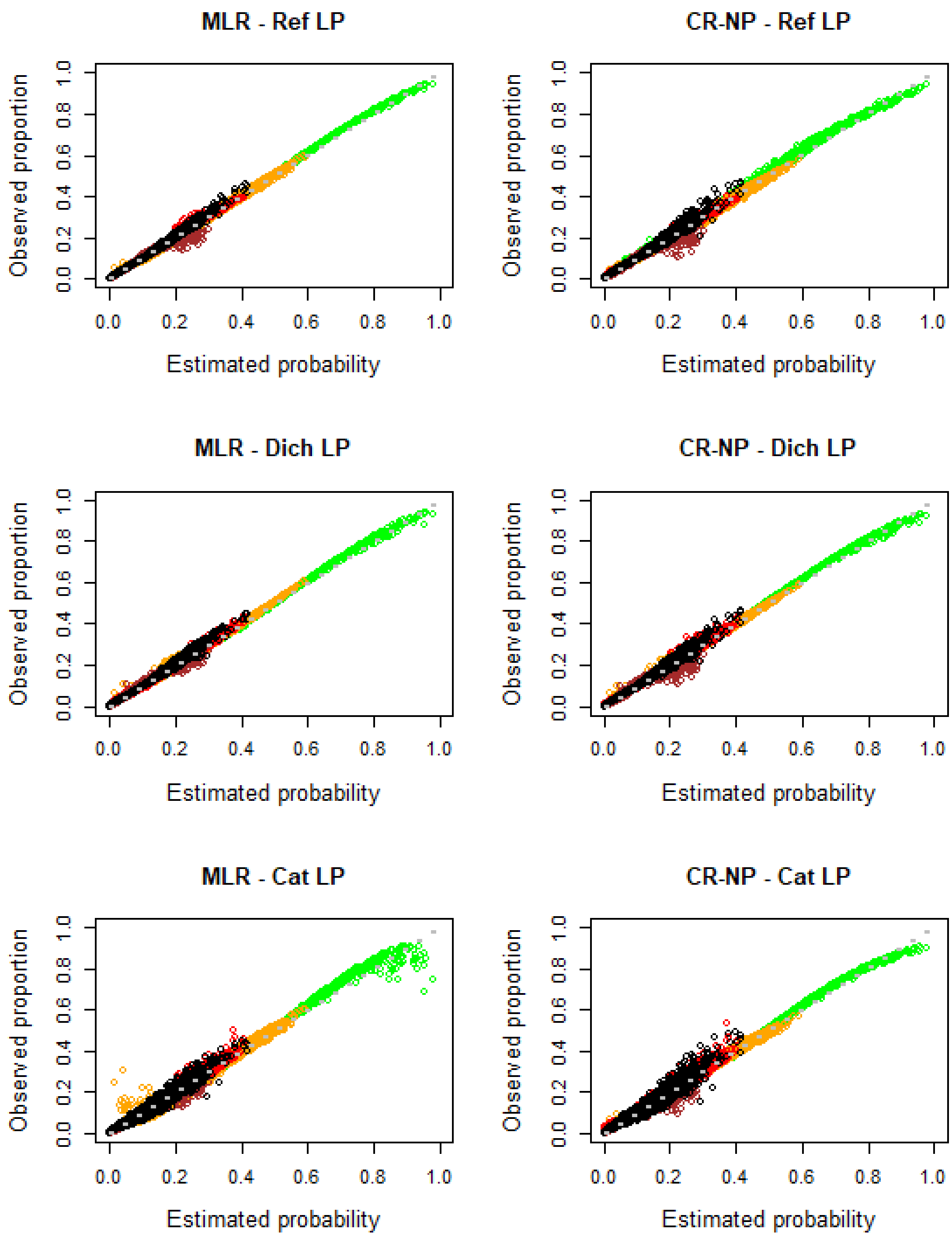
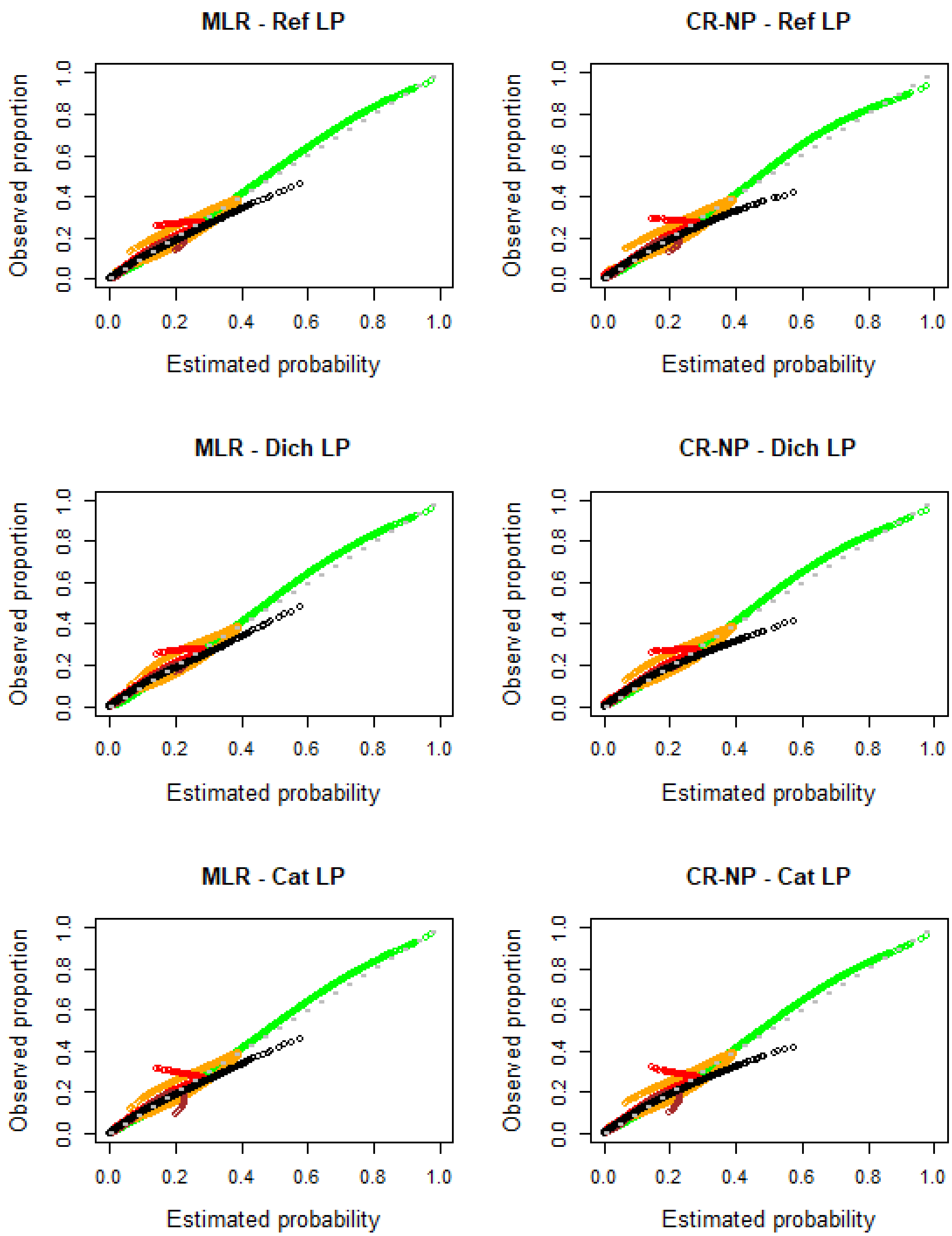


Figure A3.2. Calibration scatter plots based on six different flexible recalibration models for the CL-PO model in the case study. Green: category 1 (no coronary artery disease); orange: category 2 (non-obstructive stenosis); red: category 3 (1-vessel disease); brown: category 4 (2-vessel disease); black: category 5 (3-vessel disease).



4. Supplementary tables

Table S1. Details of the simulation scenarios when the true model has an MLR form.

Scenario	***Q***	***K***	**ORC**	**Outcome distribution**	**Means of $X_{p,k}$***
1	4 continuous	3	0.74	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	$\boldsymbol{\mu}_{\mathbf{1k}}=(0.0,0.4,0.8)$ $\boldsymbol{\mu}_{\mathbf{2k}}=(0.0,0.3,0.6)$ $\boldsymbol{\mu}_{\mathbf{3k}}=(0.0,0.4,0.8)$ $\boldsymbol{\mu}_{\mathbf{4k}}=(0.0,0.3,0.6)$
2	4 continuous	3	0.74	$(0.55,0.30,0.15)$	$\boldsymbol{\mu}_{\mathbf{1k}}=(0.0,0.4,0.8)$ $\boldsymbol{\mu}_{\mathbf{2k}}=(0.0,0.3,0.6)$ $\boldsymbol{\mu}_{\mathbf{3k}}=(0.0,0.4,0.8)$ $\boldsymbol{\mu}_{\mathbf{4k}}=(0.0,0.3,0.6)$
3	4 continuous	3	0.74	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	$\boldsymbol{\mu_{1k}}=(0.0,0.7,0.8)$ $\boldsymbol{\mu_{2k}}=(0.0,0.6,0.6)$ $\boldsymbol{\mu_{3k}}=(0.0,0.5,0.8)$ $\boldsymbol{\mu_{4k}}=(0.0,0.1,0.6)$
4	4 continuous	3	0.74	$(0.55,0.30,0.15)$	$\boldsymbol{\mu_{1k}}=(0.0,0.7,0.8)$ $\boldsymbol{\mu_{2k}}=(0.0,0.6,0.6)$ $\boldsymbol{\mu_{3k}}=(0.0,0.5,0.8)$ $\boldsymbol{\mu_{4k}}=(0.0,0.1,0.6)$
5	4 continuous	3	0.74	$(0.55,0.30,0.15)$	$\boldsymbol{\mu_{1k}}=(0.0,0.7,0.8)$ $\boldsymbol{\mu_{2k}}=(0.0,0.7,0.6)$ $\boldsymbol{\mu_{3k}}=(0.0,0.0,1.0)$ $\boldsymbol{\mu_{4k}}=(0.3,0.0,0.3)$
6	3 continuous	4	0.74	$(0.40,0.25,0.20,0.15)$	$\boldsymbol{\mu_{1k}}=(0.0,0.0,1.0,1.0)$ $\boldsymbol{\mu_{2k}}=(0.0,0.8,0.8,0.9)$ $\boldsymbol{\mu_{3k}}=(0.2,0.0,0.9,1.0)$
7	3 continuous	4	0.66	$(0.40,0.25,0.20,0.15)$	$\boldsymbol{\mu_{1k}}=(0.0,0.0,0.6,0.6)$ $\boldsymbol{\mu_{2k}}=(0.0,0.4,0.4,0.5)$ $\boldsymbol{\mu_{3k}}=(0.1,0.0,0.6,0.7)$
8	3 continuous	4	0.66	$(0.45,0.30,0.20,0.05)$	$\boldsymbol{\mu_{1k}}=(0.0,0.0,0.6,0.6)$ $\boldsymbol{\mu_{2k}}=(0.0,0.4,0.4,0.5)$ $\boldsymbol{\mu_{3k}}=(0.1,0.0,0.6,0.7)$
9	4 binary	3	0.74	$(0.55,0.30,0.15)$	$\boldsymbol{\mu_{1k}}=(0.20,0.55,0.58)$ $\boldsymbol{\mu_{2k}}=(0.20,0.50,0.50)$ $\boldsymbol{\mu_{3k}}=(0.20,0.45,0.58)$ $\boldsymbol{\mu_{4k}}=(0.20,0.25,0.50)$
10	3 binary	4	0.74	$(0.40,0.25,0.20,0.15)$	$\boldsymbol{\mu_{1k}}=(0.20,0.20,0.65,0.65)$ $\boldsymbol{\mu_{2k}}=(0.20,0.40,0.40,0.60)$ $\boldsymbol{\mu_{3k}}=(0.25,0.20,0.60,0.70)$
11	8 continuous (4 true + 4 noise)	3	0.74	$(0.55,0.30,0.15)$	$\boldsymbol{\mu_{1k}}=(0.0,0.7,0.8)$ $\boldsymbol{\mu_{2k}}=(0.0,0.6,0.6)$ $\boldsymbol{\mu_{3k}}=(0.0,0.5,0.8)$ $\boldsymbol{\mu_{4k}}=(0.0,0.1,0.6)$ $\boldsymbol{\mu_{5k}}=(0.0,0.0,0.0)$ $\boldsymbol{\mu_{6k}}=(0.0,0.0,0.0)$ $\boldsymbol{\mu_{7k}}=(0.0,0.0,0.0)$ $\boldsymbol{\mu_{8k}}=(0.0,0.0,0.0)$

* For binary predictors, the means refer to the prevalences of the predictor for each outcome category.
MLR, multinomial logistic regression; ORC, ordinal C statistic.

Table S2. Details of the simulation scenarios when the true model has a CL-PO form.

Scenario	Q	K	ORC	Outcome distribution	True model parameters
1	4 continuous	3	0.74	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	$\boldsymbol{\alpha_k} = [-0.18,1.55]^T$ $\boldsymbol{\beta} = [-0.55,-0.41,-0.55,-0.41]^T$
2	4 continuous	3	0.74	$(0.55, 0.30, 0.15)$	$\boldsymbol{\alpha_k} = [0.92,2.80]^T$ $\boldsymbol{\beta} = [-0.53,-0.39,-0.53,-0.39]^T$
3	4 continuous	3	0.74	$(0.70, 0.25, 0.05)$	$\boldsymbol{\alpha_k} = [1.73,4.15]^T$ $\boldsymbol{\beta} = [-0.53,-0.39,-0.53,-0.39]^T$
4	3 continuous	4	0.74	$(0.55, 0.30, 0.15)$	$\boldsymbol{\alpha_k} = [-0.12,1.22,2.62]^T$ $\boldsymbol{\beta} = [-0.54,-0.47,-0.51]^T$
5	3 continuous	4	0.66	$(0.55, 0.30, 0.15)$	$\boldsymbol{\alpha_k} = [-0.05,1.10,2.35]^T$ $\boldsymbol{\beta} = [-0.54,-0.47,-0.51]^T$
6	3 continuous	4	0.66	$(0.70, 0.25, 0.05)$	$\boldsymbol{\alpha_k} = [-0.18,1.55]^T$ $\boldsymbol{\beta} = [-0.55,-0.41,-0.55,-0.41]^T$
7	4 binary	3	0.74	$(0.55, 0.30, 0.15)$	$\boldsymbol{\alpha_k} = [-0.18,1.55]^T$ $\boldsymbol{\beta} = [-0.55,-0.41,-0.55,-0.41]^T$
8	3 binary	4	0.74	$(0.55, 0.30, 0.15)$	$\boldsymbol{\alpha_k} = [-0.18,1.55]^T$ $\boldsymbol{\beta} = [-0.55,-0.41,-0.55,-0.41]^T$
9	8 continuous (4 true + 4 noise)	3	0.74	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	$\boldsymbol{\alpha_k} = [-0.18,1.55]^T$ $\boldsymbol{\beta} = [-0.55,-0.41,-0.55,-0.41, 0,0,0,0]^T$

* For binary predictors, the means refer to the prevalences of the predictor for each outcome category.
CL-PO, cumulative logit model with proportional odds; ORC, ordinal C statistic.

Table S3. Large sample estimates of the model coefficients for scenarios 1-11 under MLR truth.

Model Parameter	Scen. 1	Scen. 2	Scen. 3	Scen. 4	Scen. 5	Scen. 6	Scen. 7	Scen. 8	Scen. 9	Scen. 10	Scen. 11[a]
MLR											
Int, k=2 vs k=1	-0.25	-0.86	-0.56	-1.17	-1.06	-0.77	-0.55	-0.48	-2.09	-0.70	-1.17
Int, k=3 vs k=1	-1.00	-2.30	-1.00	-2.30	-2.30	-1.91	-1.13	-1.26	-3.55	-2.44	-2.30
Int, k=4 vs k=1	na	na	na	na	na	-2.36	-1.53	-2.73	na	-3.45	na
X_1, k=2 vs k=1	0.40	0.40	0.70	0.70	0.70	0.00	0.00	0.00	1.59	0.03	0.70
X_1, k=3 vs k=1	0.80	0.79	0.80	0.79	0.79	1.00	0.60	0.59	1.74	2.03	0.79
X_1, k=4 vs k=1	na	na	na	na	na	0.99	0.59	0.61	na	2.06	na
X_2, k=2 vs k=1	0.29	0.30	0.59	0.60	0.70	0.80	0.40	0.40	1.38	0.97	0.60
X_2, k=3 vs k=1	0.59	0.59	0.59	0.59	0.59	0.80	0.40	0.40	1.38	0.96	0.59
X_2, k=4 vs k=1	na	na	na	na	na	0.89	0.49	0.46	na	1.77	na
X_3, k=2 vs k=1	0.40	0.40	0.50	0.50	0.00	-0.20	-0.10	-0.10	1.20	-0.28	0.50
X_3, k=3 vs k=1	0.80	0.80	0.80	0.79	1.00	0.70	0.50	0.50	1.71	1.50	0.79
X_3, k=4 vs k=1	na	na	na	na	na	0.79	0.59	0.59	na	1.95	na
X_4, k=2 vs k=1	0.30	0.30	0.10	0.10	-0.30	na	na	na	0.28	na	0.10
X_4, k=3 vs k=1	0.60	0.59	0.60	0.59	-0.01	na	na	na	1.38	na	0.59
CL-PO											
Int, k≥2 vs k=1	0.18	-0.63	0.07	-0.81	-0.68	-0.14	0.20	0.05	-1.73	-0.69	-0.81
Int, k≥3 vs k≤2	-1.55	-2.47	-1.65	-2.72	-2.55	-1.47	-0.94	-1.36	-3.72	-2.02	-2.72
Int, k≥4 vs k≤3	na	na	na	na	na	-2.87	-2.18	-3.31	na	-3.49	na
X_1	0.55	0.53	0.53	0.62	0.63	0.59	0.40	0.36	1.36	1.43	0.62
X_2	0.41	0.39	0.39	0.49	0.54	0.60	0.34	0.34	1.09	1.00	0.49
X_3	0.55	0.53	0.54	0.55	0.48	0.43	0.37	0.29	1.20	1.17	0.55
X_4	0.41	0.39	0.42	0.32	-0.11	na	na	na	0.84	na	0.32
AC-PO											
Int, k=2 vs k=1	-0.26	-0.86	-0.31	-0.95	-0.87	-0.68	-0.54	-0.48	-1.62	-1.01	-0.95
Int, k=3 vs k=2	-0.75	-1.44	-0.80	-1.62	-1.48	-0.76	-0.42	-0.62	-2.39	-1.08	-1.62
Int, k=4 vs k=3	na	na	na	na	na	-1.16	-0.63	-1.73	na	-1.52	na
X_1	0.40	0.40	0.39	0.45	0.46	0.36	0.23	0.24	1.00	0.81	0.45
X_2	0.30	0.30	0.28	0.35	0.38	0.33	0.17	0.20	0.80	0.57	0.35
X_3	0.40	0.40	0.39	0.42	0.40	0.28	0.21	0.20	0.91	0.70	0.42
X_4	0.30	0.30	0.31	0.26	-0.07	na	na	na	0.67	na	0.26
SLM											
Int, k=2 vs k=1	-0.25	-0.86	-0.51	-1.15	-1.00	-0.53	-0.48	-0.41	-2.13	-0.57	-1.15
Int, k=3 vs k=1	-1.00	-2.30	-1.00	-2.26	-2.11	-1.91	-1.13	-1.25	-3.50	-2.45	-2.26
Int, k=4 vs k=1	na	na	na	na	na	-2.35	-1.51	-2.72	na	-3.38	na
X_1	0.40	0.40	0.59	0.64	0.62	0.17	0.05	0.05	1.48	0.18	0.64
X_2	0.30	0.30	0.45	0.52	0.54	0.11	0.02	0.02	1.23	0.09	0.52
X_3	0.40	0.40	0.55	0.55	0.40	0.13	0.05	0.04	1.27	0.16	0.55
X_4	0.30	0.30	0.37	0.28	-0.14	na	na	na	0.75	na	0.28
□, k=2 vs k=1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
□, k=3 vs k=1	2.01	1.98	1.41	1.33	1.41	5.98	12.20	13.16	1.28	10.67	1.33
□, k=4 vs k=1	na	na	na	na	na	6.35	13.43	14.50	na	13.06	na

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model; na, not applicable.

[a] Coefficients for the noise variables were all 0.00 for CL-PO, AC-PO, and SLM. For MLR, 6 out of 8 coefficients were 0.00, one was -0.01, and one was 0.01.

Table S4. Apparent performance based on a large dataset of n=200,000 for simulation scenarios 5 to 11 under MLR truth.

	CALIBRATION INTERCEPTS AND SLOPES										SINGLE NUMBER METRICS		
	Per outcome category				Per outcome dichotomy			Model-specific					
MODEL	**Y=1**	**Y=2**	**Y=3**	**Y=4**	**Y>1**	**Y>2**	**Y>3**	**LP1**	**LP2**	**LP3**	**ECI**	**rMSPE**	**ORC**
MLR truth scenario 5: K=3, Q=4, imbalanced outcome, highly non-equidistant means, ORC 0.74													
MLR	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	na	0.00 / 1.00	0.00 / 1.00	na	0.00 / 1.00	0.00 / 1.00	na	0.000	0.002	0.740
CL-PO	-0.01 / 1.06	0.00 / 1.02	0.01 / 0.94	na	0.01 / 1.06	0.01 / 0.94	na	0.00 / 1.00	0.00 / 1.00	na	0.013	0.100	0.730
AC-PO	0.00 / 1.09	0.00 / 1.26	0.00 / 0.87	na	0.00 / 1.09	0.00 / 0.87	na	0.00 / 1.00	0.00 / 1.00	na	0.018	0.103	0.734
SLM	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	na	0.00 / 1.00	0.00 / 1.00	na	0.00 / 1.00	0.00 / 1.00	na	0.000	0.095	0.724
MLR truth scenario 6: K=4, Q=3, imbalanced outcome, highly non-equidistant means, ORC 0.74													
MLR	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.000	0.002	0.741
CL-PO	0.03 / 0.94	-0.04 / 0.66	-0.03 / 1.66	0.04 / 0.89	-0.03 / 0.94	0.00 / 1.37	0.04 / 0.89	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.119	0.098	0.735
AC-PO	0.00 / 0.94	0.00 / 1.17	0.00 / 1.73	0.00 / 0.79	0.00 / 0.94	0.00 / 1.26	0.00 / 0.79	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.096	0.096	0.737
SLM	0.00 / 1.00	0.00 / 1.01	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.000	0.088	0.737
MLR truth scenario 7: K=4, Q=3, imbalanced outcome, highly non-equidistant means, ORC 0.66													
MLR	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.000	0.002	0.663
CL-PO	0.02 / 0.86	-0.01 / 0.30	-0.03 / 1.63	0.01 / 1.01	-0.02 / 0.86	-0.01 / 1.33	0.01 / 1.01	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.156	0.058	0.662
AC-PO	0.00 / 0.87	0.00 / 1.48	0.00 / 1.68	0.00 / 0.88	0.00 / 0.87	0.00 / 1.23	0.00 / 0.88	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.116	0.055	0.663
SLM	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.000	0.046	0.663
MLR truth scenario 8: K=4, Q=3, highly imbalanced outcome, highly non-equidistant means, ORC 0.66													
MLR	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.000	0.002	0.661
CL-PO	0.02 / 0.81	-0.01 / -0.08	-0.03 / 1.46	0.01 / 1.12	-0.02 / 0.81	-0.02 / 1.39	0.01 / 1.12	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.185	0.064	0.659
AC-PO	0.00 / 0.83	0.00 / -0.01	0.00 / 1.47	0.00 / 0.85	0.00 / 0.83	0.00 / 1.28	0.00 / 0.85	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.158	0.061	0.661
SLM	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.000	0.052	0.662
MLR truth scenario 9: K=3, Q=4 binary, imbalanced outcome, non-equidistant means, ORC 0.74													
MLR	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	na	0.00 / 1.00	0.00 / 1.00	na	0.00 / 1.00	0.00 / 1.00	na	0.000	0.002	0.745
CL-PO	-0.03 / 1.15	0.01 / 1.20	0.04 / 0.81	na	0.03 / 1.15	0.04 / 0.81	na	0.00 / 1.00	0.00 / 1.00	na	0.106	0.066	0.742
AC-PO	0.00 / 1.22	0.00 / 1.58	0.00 / 0.74	na	0.00 / 1.22	0.00 / 0.74	na	0.00 / 1.00	0.00 / 1.00	na	0.145	0.074	0.742
SLM	0.00 / 1.00	0.00 / 1.01	0.00 / 0.99	na	0.00 / 1.00	0.00 / 0.99	na	0.00 / 1.00	0.00 / 1.00	na	0.048	0.051	0.742
MLR truth scenario 10: K=4, Q=3 binary, imbalanced outcome, highly non-equidistant means, ORC 0.74													
MLR	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.000	0.002	0.742
CL-PO	0.06 / 0.81	-0.02 / 0.90	-0.06 / 1.51	0.02 / 1.03	-0.06 / 0.81	-0.04 / 1.34	0.02 / 1.03	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.236	0.077	0.742
AC-PO	0.00 / 0.83	0.00 / 1.60	0.00 / 1.53	0.00 / 0.93	0.00 / 0.83	0.00 / 1.24	0.00 / 0.93	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.218	0.071	0.742
SLM	0.00 / 0.99	0.00 / 1.05	0.00 / 1.01	0.00 / 0.99	0.00 / 0.99	0.00 / 1.00	0.00 / 0.99	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.148	0.058	0.742
MLR truth scenario 11: K=3, Q=8 continuous (4 true + 4 noise), imbalanced outcome, non-equidistant means, ORC 0.74													
MLR	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	na	0.00 / 1.00	0.00 / 1.00	na	0.00 / 1.00	0.00 / 1.00	na	0.000	0.003	0.737
CL-PO	-0.02 / 1.11	0.01 / 1.13	0.03 / 0.85	na	0.02 / 1.11	0.03 / 0.85	na	0.00 / 1.00	0.00 / 1.00	na	0.032	0.058	0.735
AC-PO	0.00 / 1.17	0.00 / 1.47	0.00 / 0.77	na	0.00 / 1.17	0.00 / 0.77	na	0.00 / 1.00	0.00 / 1.00	na	0.059	0.064	0.736
SLM	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	na	0.00 / 1.00	0.00 / 1.00	na	0.00 / 1.00	0.00 / 1.00	na	0.000	0.047	0.733

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model; LP, linear predictor; ECI, estimated calibration index; rMSPE, root mean squared prediction error; ORC, ordinal C statistic; CAD, coronary artery disease; na, not applicable.

Table S5. Apparent performance based on a large dataset of n=200,000 for simulation scenarios 4 to 9 under CL-PO truth.

	CALIBRATION INTERCEPTS AND SLOPES										SINGLE NUMBER METRICS		
	Per outcome category				Per outcome dichotomy			Model-specific					
MODEL	**Y=1**	**Y=2**	**Y=3**	**Y=4**	**Y>1**	**Y>2**	**Y>3**	**LP1**	**LP2**	**LP3**	**ECI**	**rMSPE**	**ORC**
CL-PO truth scenario 4: K=4, Q=3, imbalanced outcome, ORC 0.74													
MLR	0.00 / 0.99	0.00 / 1.34	0.00 / 1.05	0.00 / 0.98	0.00 / 0.99	0.00 / 1.00	0.00 / 0.98	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.007	0.014	0.741
CL-PO	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.000	0.001	0.741
AC-PO	0.00 / 1.09	0.00 / 1.28	0.00 / 1.03	0.00 / 0.92	0.00 / 1.09	0.00 / 0.95	0.00 / 0.92	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.017	0.020	0.741
SLM	0.00 / 0.99	0.00 / 1.34	0.00 / 1.05	0.00 / 0.98	0.00 / 0.99	0.00 / 1.00	0.00 / 0.98	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.008	0.014	0.741
CL-PO truth scenario 5: K=4, Q=3, imbalanced outcome, ORC 0.66													
MLR	0.00 / 1.00	0.00 / 1.39	0.00 / 1.02	0.00 / 0.99	0.00 / 1.00	0.00 / 1.00	0.00 / 0.99	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.004	0.008	0.661
CL-PO	0.00 / 1.00	0.00 / 0.98	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.000	0.001	0.661
AC-PO	0.00 / 1.09	0.00 / 1.20	0.00 / 1.06	0.00 / 0.91	0.00 / 1.09	0.00 / 0.97	0.00 / 0.91	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.015	0.014	0.661
SLM	0.00 / 1.00	0.00 / 1.39	0.00 / 1.02	0.00 / 0.99	0.00 / 1.00	0.00 / 1.00	0.00 / 0.99	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.004	0.008	0.661
CL-PO truth scenario 6: K=4, Q=3, highly imbalanced outcome, ORC 0.66													
MLR	0.00 / 1.00	0.00 / 1.15	0.00 / 1.00	0.00 / 0.98	0.00 / 1.00	0.00 / 1.00	0.00 / 0.98	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.003	0.007	0.660
CL-PO	0.00 / 1.00	0.00 / 1.01	0.00 / 1.00	0.00 / 0.99	0.00 / 1.00	0.00 / 1.00	0.00 / 0.99	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.000	0.001	0.660
AC-PO	0.00 / 1.11	0.00 / 1.65	0.00 / 1.02	0.00 / 0.77	0.00 / 1.11	0.00 / 0.95	0.00 / 0.77	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.024	0.014	0.660
SLM	0.00 / 1.00	0.00 / 1.15	0.00 / 1.00	0.00 / 0.98	0.00 / 1.00	0.00 / 1.00	0.00 / 0.98	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.003	0.007	0.660
CL-PO truth scenario 7: K=3, Q=4 binary, imbalanced outcome, ORC 0.74													
MLR	0.00 / 1.00	0.00 / 1.07	0.00 / 0.99	na	0.00 / 1.00	0.00 / 0.99	na	0.00 / 1.00	0.00 / 1.00	na	0.005	0.013	0.742
CL-PO	0.00 / 1.00	0.00 / 1.01	0.00 / 0.99	na	0.00 / 1.00	0.00 / 0.99	na	0.00 / 1.00	0.00 / 1.00	na	0.000	0.002	0.742
AC-PO	0.00 / 1.07	0.00 / 1.32	0.00 / 0.88	na	0.00 / 1.07	0.00 / 0.88	na	0.00 / 1.00	0.00 / 1.00	na	0.013	0.018	0.742
SLM	0.00 / 1.00	0.00 / 1.07	0.00 / 0.99	na	0.00 / 1.00	0.00 / 0.99	na	0.00 / 1.00	0.00 / 1.00	na	0.005	0.013	0.742
CL-PO truth scenario 8: K=4, Q=3 binary, imbalanced outcome, ORC 0.74													
MLR	0.00 / 0.99	0.00 / 1.35	0.00 / 1.03	0.00 / 0.98	0.00 / 0.99	0.00 / 1.00	0.00 / 0.98	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.007	0.014	0.740
CL-PO	0.00 / 1.00	0.00 / 1.02	0.00 / 0.99	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.000	0.001	0.740
AC-PO	0.00 / 1.09	0.00 / 1.30	0.00 / 1.02	0.00 / 0.91	0.00 / 1.09	0.00 / 0.95	0.00 / 0.91	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.018	0.020	0.740
SLM	0.00 / 0.99	0.00 / 1.38	0.00 / 1.03	0.00 / 0.98	0.00 / 0.99	0.00 / 1.00	0.00 / 0.98	0.00 / 1.00	0.00 / 1.00	0.00 / 1.00	0.008	0.014	0.740
CL-PO truth scenario 9: K=3, Q=8 continuous (4 true + 4 noise), balanced outcome, ORC 0.74													
MLR	0.00 / 0.99	0.00 / 1.38	0.00 / 0.99	na	0.00 / 0.99	0.00 / 0.99	na	0.00 / 1.00	0.00 / 1.00	na	0.006	0.014	0.740
CL-PO	0.00 / 1.00	0.00 / 1.02	0.00 / 0.99	na	0.00 / 1.00	0.00 / 0.99	na	0.00 / 1.00	0.00 / 1.00	na	0.000	0.003	0.740
AC-PO	0.00 / 1.00	0.00 / 1.38	0.00 / 0.99	na	0.00 / 1.00	0.00 / 0.99	na	0.00 / 1.00	0.00 / 1.00	na	0.006	0.014	0.740
SLM	0.00 / 0.99	0.00 / 1.38	0.00 / 0.99	na	0.00 / 0.99	0.00 / 0.99	na	0.00 / 1.00	0.00 / 1.00	na	0.006	0.014	0.740

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model; LP, linear predictor; ECI, estimated calibration index; rMSPE, root mean squared prediction error; ORC, ordinal C statistic; CAD, coronary artery disease; na, not applicable.

Table S6. Large sample estimates of the model coefficients for scenarios 1-8 under CL-PO truth.

Model Parameter	**Scen. 1**	**Scen. 2**	**Scen. 3**	**Scen. 4**	**Scen. 5**	**Scen. 6**	**Scen. 7**	**Scen. 8**	**Scen. 9[a]**
MLR									
Int, k=2 vs k=1	-0.27	-1.14	-1.81	-0.51	-0.65	-0.64	-2.07	-1.66	-0.27
Int, k=3 vs k=1	-0.99	-2.58	-4.18	-1.07	-1.10	-1.38	-4.19	-2.98	-0.99
Int, k=4 vs k=1	na	na	na	-2.00	-1.74	-3.03	na	-4.74	na
X_1, k=2 vs k=1	0.41	0.45	0.50	0.37	0.36	0.40	1.19	1.18	0.41
X_1, k=3 vs k=1	0.79	0.75	0.75	0.62	0.60	0.70	1.95	2.03	0.79
X_1, k=4 vs k=1	na	na	na	0.90	0.83	0.87	na	2.89	na
X_2, k=2 vs k=1	0.29	0.32	0.37	0.33	0.31	0.36	0.93	0.82	0.29
X_2, k=3 vs k=1	0.58	0.55	0.56	0.54	0.52	0.60	1.56	1.38	0.58
X_2, k=4 vs k=1	na	na	na	0.78	0.72	0.76	na	2.02	na
X_3, k=2 vs k=1	0.39	0.45	0.50	0.36	0.34	0.39	1.03	1.06	0.39
X_3, k=3 vs k=1	0.78	0.76	0.75	0.59	0.57	0.64	1.73	1.71	0.78
X_3, k=4 vs k=1	na	na	na	0.85	0.79	0.81	na	2.42	na
X_4, k=2 vs k=1	0.29	0.32	0.35	na	na	na	0.69	na	0.29
X_4, k=3 vs k=1	0.58	0.54	0.55	na	na	na	1.17	Na	0.57
CL-PO									
Int, k≥2 vs k=1	0.17	-0.92	-1.73	0.13	0.05	-0.18	-2.03	-1.52	0.17
Int, k≥3 vs k≤2	-1.56	-2.79	-4.16	-1.22	-1.10	-1.63	-3.94	-2.88	-1.56
Int, k≥4 vs k≤3	na	na	na	-2.62	-2.35	-3.58	na	-4.30	na
X_1	0.55	0.53	0.53	0.54	0.54	0.54	1.39	1.75	0.55
X_2	0.41	0.39	0.40	0.47	0.47	0.47	1.09	1.20	0.41
X_3	0.55	0.54	0.53	0.51	0.51	0.50	1.22	1.46	0.55
X_4	0.41	0.38	0.38	na	na	na	0.82	na	0.41
AC-PO									
Int, k=2 vs k=1	-0.26	-1.06	-1.69	-0.48	-0.61	-0.58	-1.87	-1.41	-0.26
Int, k=3 vs k=2	-0.73	-1.66	-2.96	-0.61	-0.48	-0.77	-2.54	-1.52	-0.73
Int, k=4 vs k=3	na	na	na	-1.01	-0.72	-1.91	na	-1.96	na
X_1	0.39	0.39	0.44	0.30	0.28	0.33	1.03	0.97	0.39
X_2	0.29	0.28	0.33	0.26	0.24	0.28	0.81	0.68	0.29
X_3	0.39	0.40	0.44	0.29	0.27	0.31	0.90	0.83	0.39
X_4	0.29	0.28	0.32	na	na	na	0.61	na	0.29
SLM									
Int, k=2 vs k=1	-0.27	-1.14	-1.81	-0.51	-0.65	-0.64	-2.07	-1.65	-0.27
Int, k=3 vs k=1	-0.99	-2.58	-4.18	-1.07	-1.10	-1.38	-4.19	-2.99	-0.99
Int, k=4 vs k=1	na	na	na	-2.00	-1.74	-3.03	na	-4.75	na
X_1	0.40	0.45	0.50	0.38	0.36	0.41	1.17	1.21	0.40
X_2	0.29	0.33	0.37	0.33	0.31	0.36	0.93	0.84	0.29
X_3	0.40	0.45	0.50	0.36	0.34	0.38	1.03	1.02	0.40
X_4	0.29	0.32	0.36	na	na	na	0.70	na	0.29
ϕ, k=2 vs k=1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ϕ, k=3 vs k=1	1.98	1.67	1.52	1.66	1.67	1.69	1.67	1.68	1.98
ϕ, k=4 vs k=1	na	na	na	2.39	2.33	2.12	na	2.40	na

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model; na, not applicable.

[a] For MLR, 5 out of 8 coefficients for the noise variables were 0.00, three were (-)0.01. For CL-PO, AC-PO, and SLM, 3 out of 4 coefficients for the noise variables were 0.00, one was -0.01.

Table S7. Descriptive statistics for the CARDIIGAN cohort (n=4888). The statistics are based on the imputed dataset. The amount of missing values that had to be imputed is stated as well.

Predictor	**All patients (n=4888)**	**No CAD (n=1381, 28%)**	**Non-obstructive stenosis (n=1606, 33%)**	**1-vessel disease (n=997, 20%)**	**2-vessel disease (n=475, 10%)**	**3-vessel disease 429 (9%)**	**Missing, n (%)**
Age, years	64 (10.8), range 18-89	59 (11.2)	66 (9.7)	65 (10.3)	67 (9.9)	67 (10.7)	0
HDL cholesterol	57 (17.4), range 15-188	61 (19.0)	57 (17.3)	54 (15.5)	52 (15.4)	52 (15.9)	312
LDL cholesterol	128 (37.7), range 21-341	126 (36.2)	126 (35.7)	130 (39.5)	129 (38.2)	134 (40.1)	310
Fibrinogen	380 (120), range 97-1414	355 (104)	379 (112)	393 (133)	396 (129)	413 (139)	119
Male sex	3028 (62%)	621 (45%)	968 (60%)	729 (73%)	373 (79%)	337 (79%)	0
Chest pain	2987 (61%)	749 (54%)	927 (58%)	663 (66%)	334 (70%)	314 (73%)	0
Diabetes mellitus	757 (15%)	124 (9%)	264 (16%)	167 (17%)	94 (20%)	108 (25%)	0
Hypertension	4090 (84%)	1073 (78%)	1376 (86%)	833 (84%)	428 (90%)	380 (89%)	0
Dyslipidemia	3591 (73%)	947 (69%)	1163 (72%)	743 (75%)	396 (83%)	342 (80)	0
Ever smoked	2317 (47%)	600 (43%)	746 (46%)	501 (50%)	244 (51%)	226 (53%)	640
CRP > 1 mg/dl	676 (14%)	142 (10%)	201 (13%)	178 (18%)	76 (16%)	79 (18%)	96

Results are presented as mean (standard deviation) or n (%).
CAD, coronary artery disease; CRP, c-reactive protein.

Table S8. P-values for the likelihood ratio test of the proportional odds assumption in the CL-PO model.

Predictor	p-value
Age	<0.0001
HDL cholesterol	0.60
LDL cholesterol	0.010
Log(Fibrinogen)	0.22
Male sex	0.46
Chest pain	0.13
Diabetes mellitus	0.017
Hypertension	0.0024
Dyslipidemia	0.016
Ever smoked	0.99
C-reactive protein > 1 mg/dl	0.13

Table S9. Coefficients for the models in the case study.

<table>
<tr><th>Model</th><th>Int</th><th>Age</th><th>Sex</th><th>Ch.pain</th><th>Diab</th><th>Hypert</th><th>Dyslip</th><th>Smok</th><th>HDL</th><th>LDL</th><th>Log(fib)</th><th>CRP>1</th><th>Phi</th></tr>
<tr><td colspan="14">MLR</td></tr>
<tr><td>k=2 vs k=1</td><td>-7.60</td><td>0.070</td><td>0.85</td><td>0.19</td><td>0.50</td><td>0.26</td><td>0.26</td><td>0.32</td><td>-0.012</td><td>0.0019</td><td>0.46</td><td>-0.053</td><td>na</td></tr>
<tr><td>k=3 vs k=1</td><td>-10.78</td><td>0.072</td><td>1.46</td><td>0.64</td><td>0.56</td><td>0.083</td><td>0.38</td><td>0.31</td><td>-0.019</td><td>0.0057</td><td>0.76</td><td>0.32</td><td>na</td></tr>
<tr><td>k=4 vs k=1</td><td>-14.04</td><td>0.10</td><td>1.83</td><td>0.79</td><td>0.68</td><td>0.53</td><td>1.03</td><td>0.34</td><td>-0.030</td><td>0.0044</td><td>0.81</td><td>0.12</td><td>na</td></tr>
<tr><td>k=5 vs k=1</td><td>-17.85</td><td>0.10</td><td>1.90</td><td>0.95</td><td>1.07</td><td>0.34</td><td>0.64</td><td>0.42</td><td>-0.028</td><td>0.0098</td><td>1.27</td><td>0.13</td><td>na</td></tr>
<tr><td colspan="14">CL-PO</td></tr>
<tr><td>k≥2 vs k=1</td><td>-7.23</td><td rowspan="4">0.058</td><td rowspan="4">1.12</td><td rowspan="4">0.54</td><td rowspan="4">0.46</td><td rowspan="4">0.18</td><td rowspan="4">0.43</td><td rowspan="4">0.19</td><td rowspan="4">-0.017</td><td rowspan="4">0.0046</td><td rowspan="4">0.56</td><td rowspan="4">0.15</td><td rowspan="4">na</td></tr>
<tr><td>k≥3 vs k≤2</td><td>-8.90</td></tr>
<tr><td>k≥4 vs k≤3</td><td>-10.08</td></tr>
<tr><td>k=5 vs k≤4</td><td>-11.02</td></tr>
<tr><td colspan="14">AC-PO</td></tr>
<tr><td>k=2 vs k=1</td><td>-3.78</td><td rowspan="4">0.026</td><td rowspan="4">0.54</td><td rowspan="4">0.27</td><td rowspan="4">0.23</td><td rowspan="4">0.10</td><td rowspan="4">0.22</td><td rowspan="4">0.087</td><td rowspan="4">-0.0083</td><td rowspan="4">0.0024</td><td rowspan="4">0.29</td><td rowspan="4">0.053</td><td rowspan="4">na</td></tr>
<tr><td>k=3 vs k=2</td><td>-4.60</td></tr>
<tr><td>k=4 vs k=3</td><td>-5.03</td></tr>
<tr><td>k=5 vs k=4</td><td>-4.54</td></tr>
<tr><td colspan="14">CR-PO</td></tr>
<tr><td>k>2 vs k≥1</td><td>-5.56</td><td rowspan="4">0.041</td><td rowspan="4">0.84</td><td rowspan="4">0.41</td><td rowspan="4">0.38</td><td rowspan="4">0.15</td><td rowspan="4">0.34</td><td rowspan="4">0.15</td><td rowspan="4">-0.013</td><td rowspan="4">0.0039</td><td rowspan="4">0.49</td><td rowspan="4">0.076</td><td rowspan="4">na</td></tr>
<tr><td>k>3 vs k≥2</td><td>-6.62</td></tr>
<tr><td>k>4 vs k≥3</td><td>-7.06</td></tr>
<tr><td>k>5 vs k≥4</td><td>-7.20</td></tr>
<tr><td colspan="14">CR-NP</td></tr>
<tr><td>k>2 vs k≥1</td><td>-8.85</td><td>0.076</td><td>1.21</td><td>0.44</td><td>0.59</td><td>0.24</td><td>0.40</td><td>0.33</td><td>-0.017</td><td>0.0039</td><td>0.65</td><td>0.10</td><td>na</td></tr>
<tr><td>k>3 vs k≥2</td><td>-4.40</td><td>0.015</td><td>0.78</td><td>0.55</td><td>0.21</td><td>-0.030</td><td>0.33</td><td>0.032</td><td>-0.011</td><td>0.0040</td><td>0.42</td><td>0.29</td><td>na</td></tr>
<tr><td>k>4 vs k≥3</td><td>-4.47</td><td>0.028</td><td>0.39</td><td>0.21</td><td>0.32</td><td>0.37</td><td>0.50</td><td>0.078</td><td>-0.011</td><td>0.0013</td><td>0.29</td><td>-0.20</td><td>na</td></tr>
<tr><td>k>5 vs k≥4</td><td>-4.04</td><td>0.0054</td><td>0.11</td><td>0.14</td><td>0.43</td><td>-0.16</td><td>-0.37</td><td>0.080</td><td>0.0026</td><td>0.0059</td><td>0.47</td><td>0.056</td><td>na</td></tr>
<tr><td colspan="14">SLM</td></tr>
<tr><td>k=2 vs k=1</td><td>-7.31</td><td rowspan="4">0.053</td><td rowspan="4">0.98</td><td rowspan="4">0.45</td><td rowspan="4">0.44</td><td rowspan="4">0.18</td><td rowspan="4">0.38</td><td rowspan="4">0.20</td><td rowspan="4">-0.014</td><td rowspan="4">0.0038</td><td rowspan="4">0.54</td><td rowspan="4">0.11</td><td>1.00</td></tr>
<tr><td>k=3 vs k=1</td><td>-10.83</td><td>1.39</td></tr>
<tr><td>k=4 vs k=1</td><td>-15.33</td><td>1.85</td></tr>
<tr><td>k=5 vs k=1</td><td>-16.83</td><td>2.02</td></tr>
</table>

Int, intercept; ch.pain, chest pain; diab, diabetes; hypert, hypertension; dyslip, dyslipidemia; smok, ever smoked; HDL, HDL cholesterol; LDL, LDL cholesterol; log(fib), log of fibrinogen; CRP, c-reactive protein.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; CR-PO, continuation ratio logit model with proportional odds; CR-NP, continuation ration logit model without proportional odds; SLM, stereotype logit model; na, not applicable.

Table S10. Mean and range of predicted probabilities for the case study.

Model	No CAD	Non-obstructive stenosis	One-vessel disease	Two-vessel disease	Three-vessel disease
	Outcome category				
MLR	0.28 (0.01-0.98)	0.33 (0.02-0.59)	0.20 (<.01-0.41)	0.10 (<.01-0.30)	0.09 (<.01-0.42)
CL-PO	0.29 (0.02-0.98)	0.33 (0.02-0.39)	0.20 (<.01-0.29)	0.10 (<.01-0.23)	0.09 (<.01-0.57)
AC-PO	0.28 (0.01-0.87)	0.33 (0.05-0.38)	0.20 (0.01-0.26)	0.10 (<.01-0.23)	0.09 (<.01-0.58)
CR-PO	0.28 (0.04-0.93)	0.34 (0.07-0.40)	0.20 (<.01-0.25)	0.09 (<.01-0.17)	0.09 (<.01-0.63)
CR-NP	0.28 (0.01-0.99)	0.33 (0.01-0.58)	0.20 (<.01-0.40)	0.10 (<.01-0.30)	0.09 (<.01-0.40)
SLM	0.28 (0.01-0.98)	0.33 (0.02-0.38)	0.20 (<.01-0.27)	0.10 (<.01-0.29)	0.09 (<.01-0.37)

CAD, coronary artery disease; MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; CR-PO, continuation ratio logit model with proportional odds; CR-NP, continuation ration logit model without proportional odds; SLM, stereotype logit model

5. Supplementary figures

Figure S1. Flexible smoothed calibration curves per outcome category for simulation scenario 1 when the true model has the form of a multinomial logistic regression (green for category 1, orange for category 2, red for category 3). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

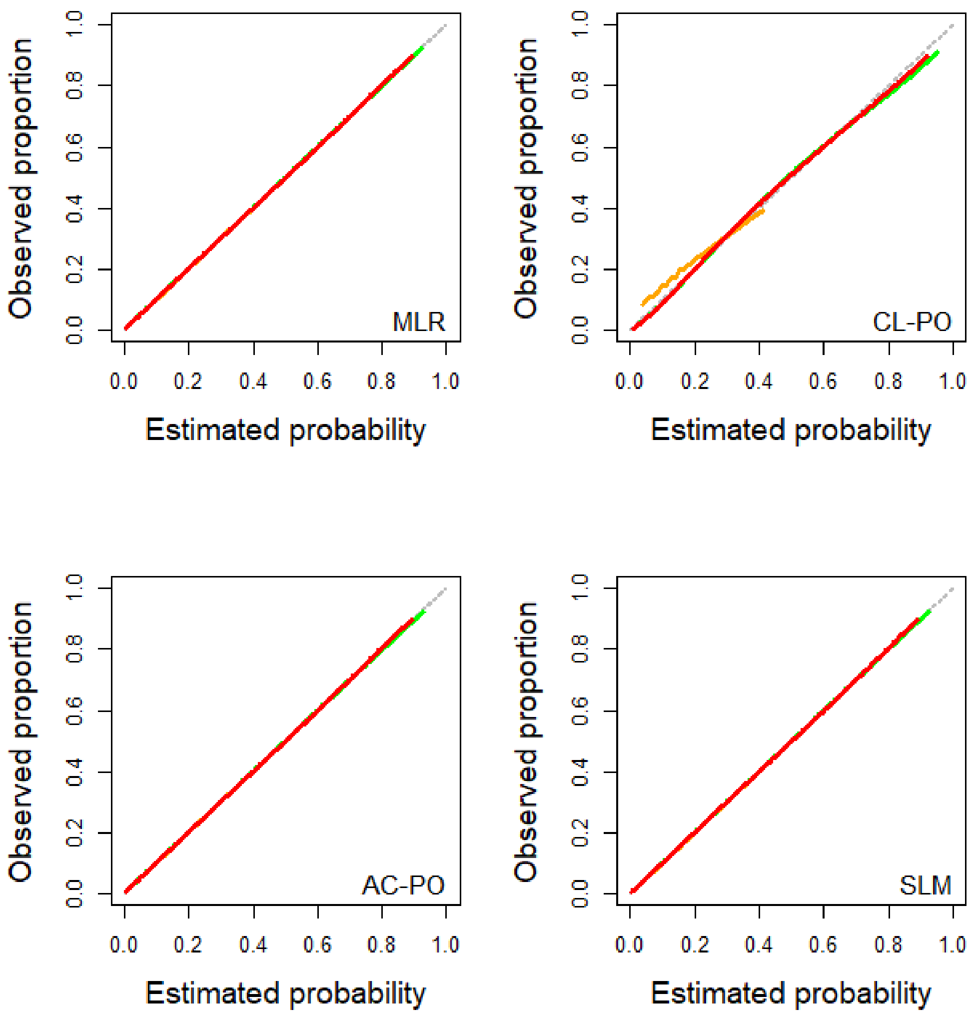


Figure S2. Flexible smoothed calibration curves per outcome category for simulation scenario 2 when the true model has the form of a multinomial logistic regression (green for category 1, orange for category 2, red for category 3). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

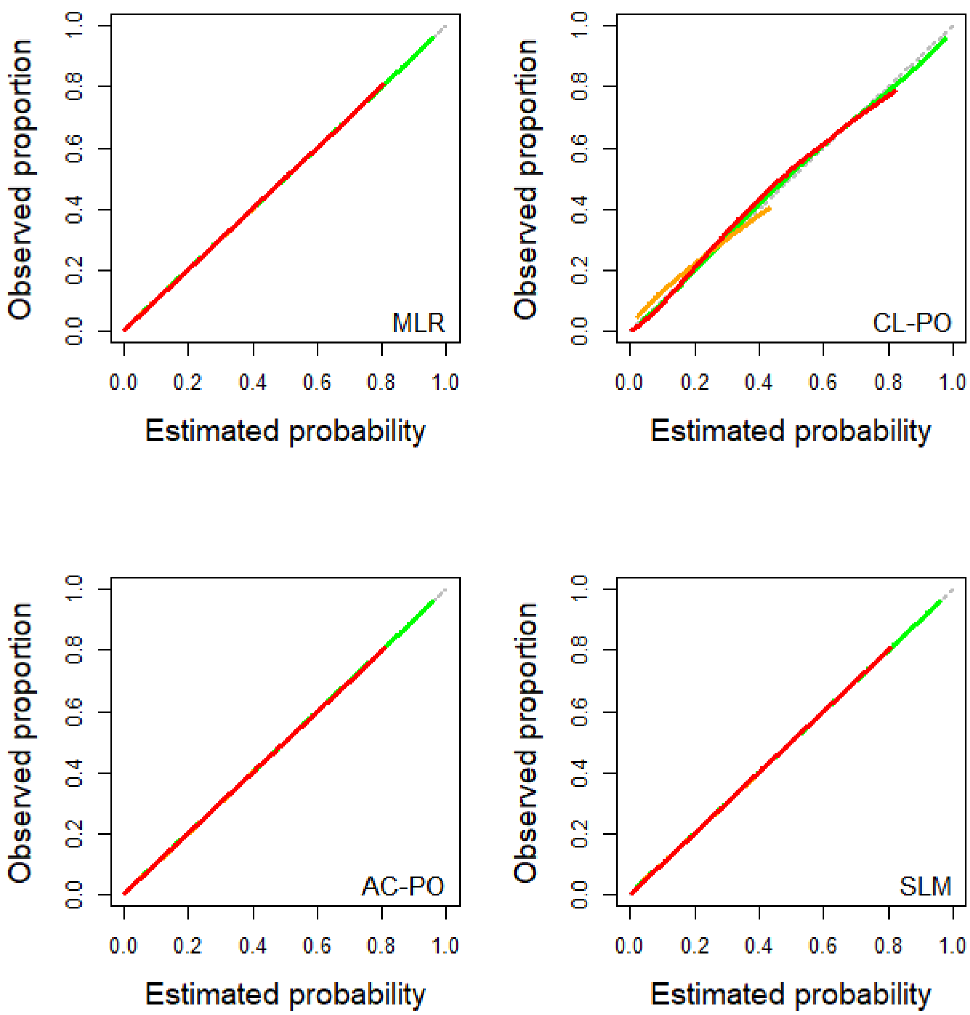


Figure S3. Flexible smoothed calibration curves per outcome category for simulation scenario 3 when the true model has the form of a multinomial logistic regression (green for category 1, orange for category 2, red for category 3). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

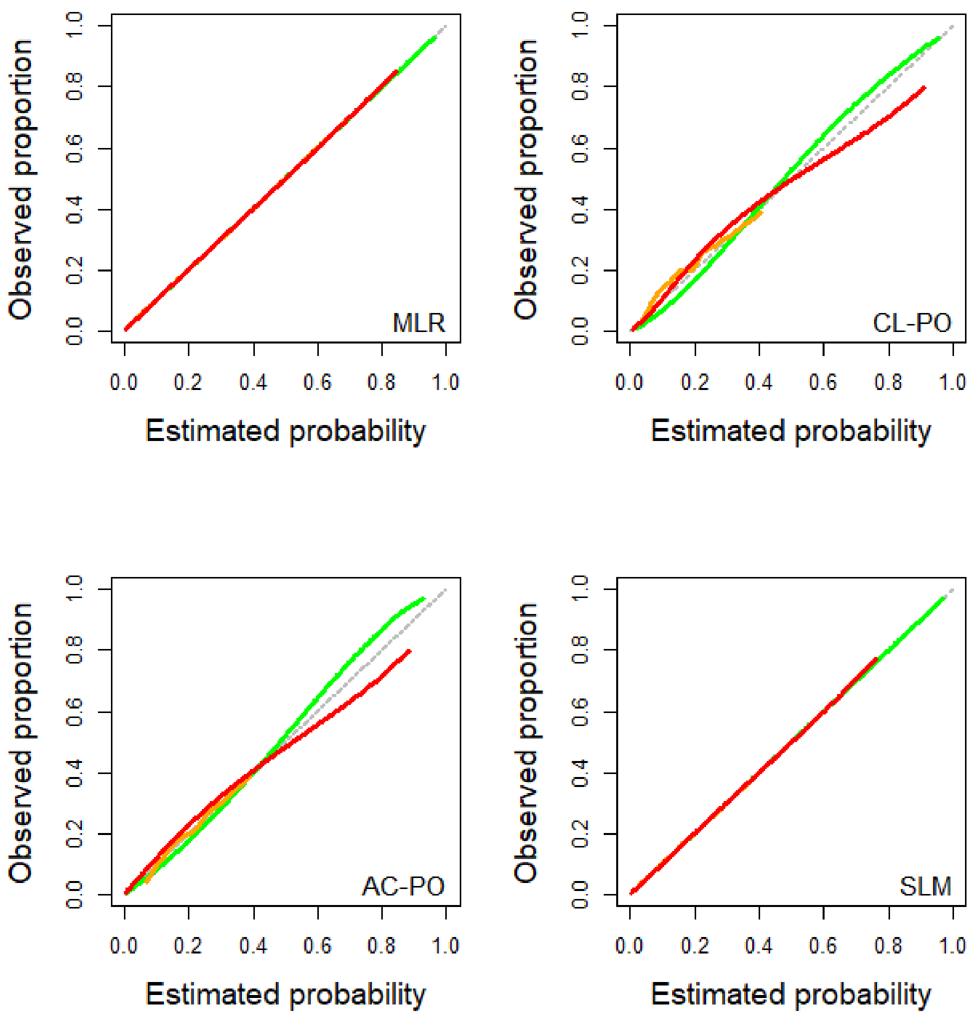


Figure S4. Flexible smoothed calibration curves per outcome category for simulation scenario 4 when the true model has the form of a multinomial logistic regression (green for category 1, orange for category 2, red for category 3). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

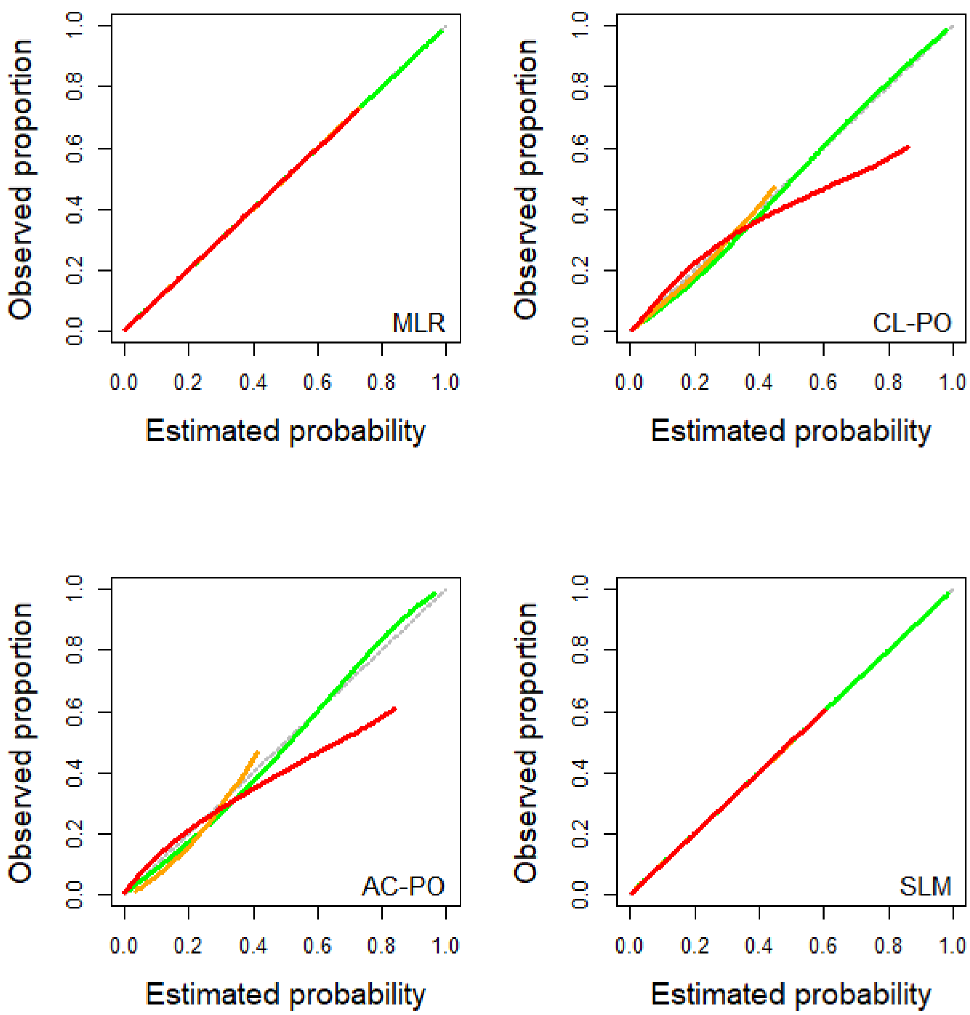


Figure S5. Scatter plots of true risks versus estimated risks for simulation scenario 5 when the true model has the form of a multinomial logistic regression. The plots are based on a random subset of 1,000 cases from all 200,000 cases.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

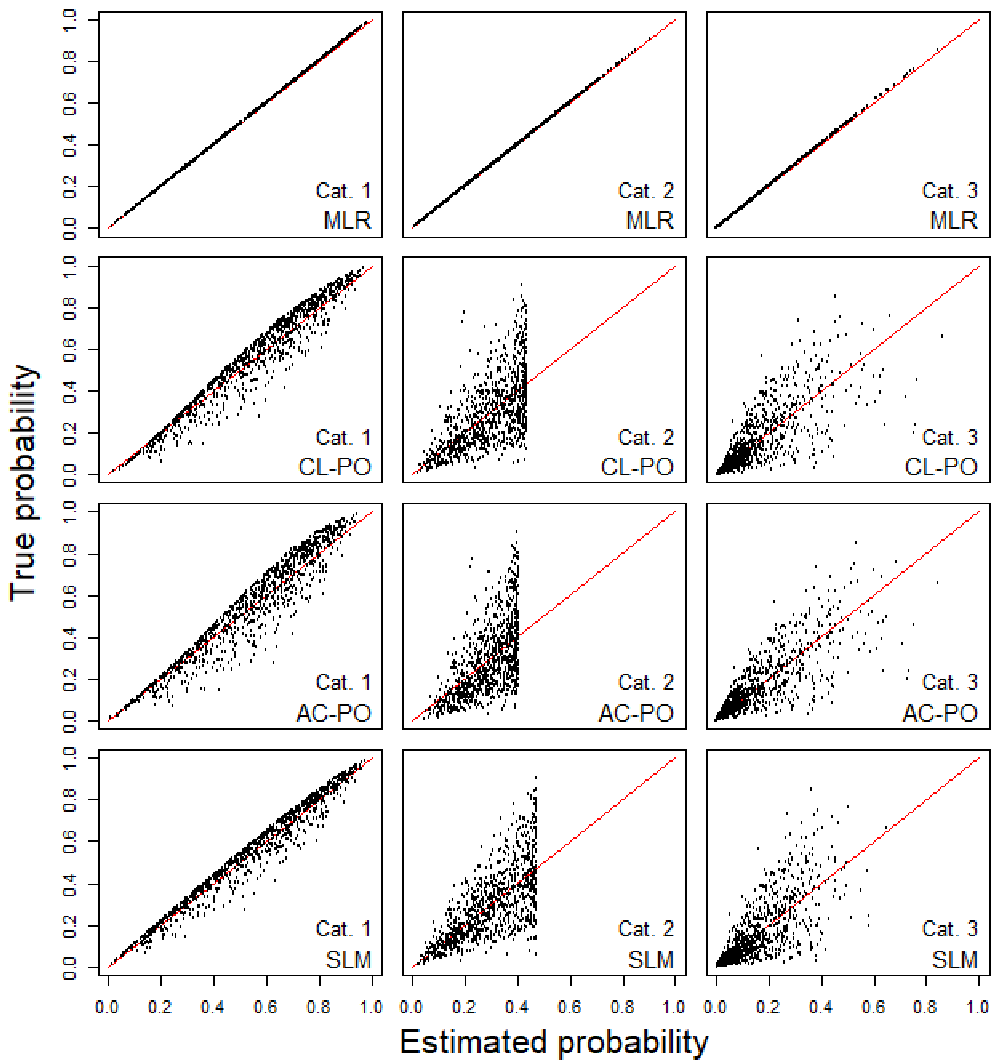


Figure S6. Scatter plots of true risks versus estimated risks for simulation scenario 6 when the true model has the form of a multinomial logistic regression. The plots are based on a random subset of 1,000 cases from all 200,000 cases.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

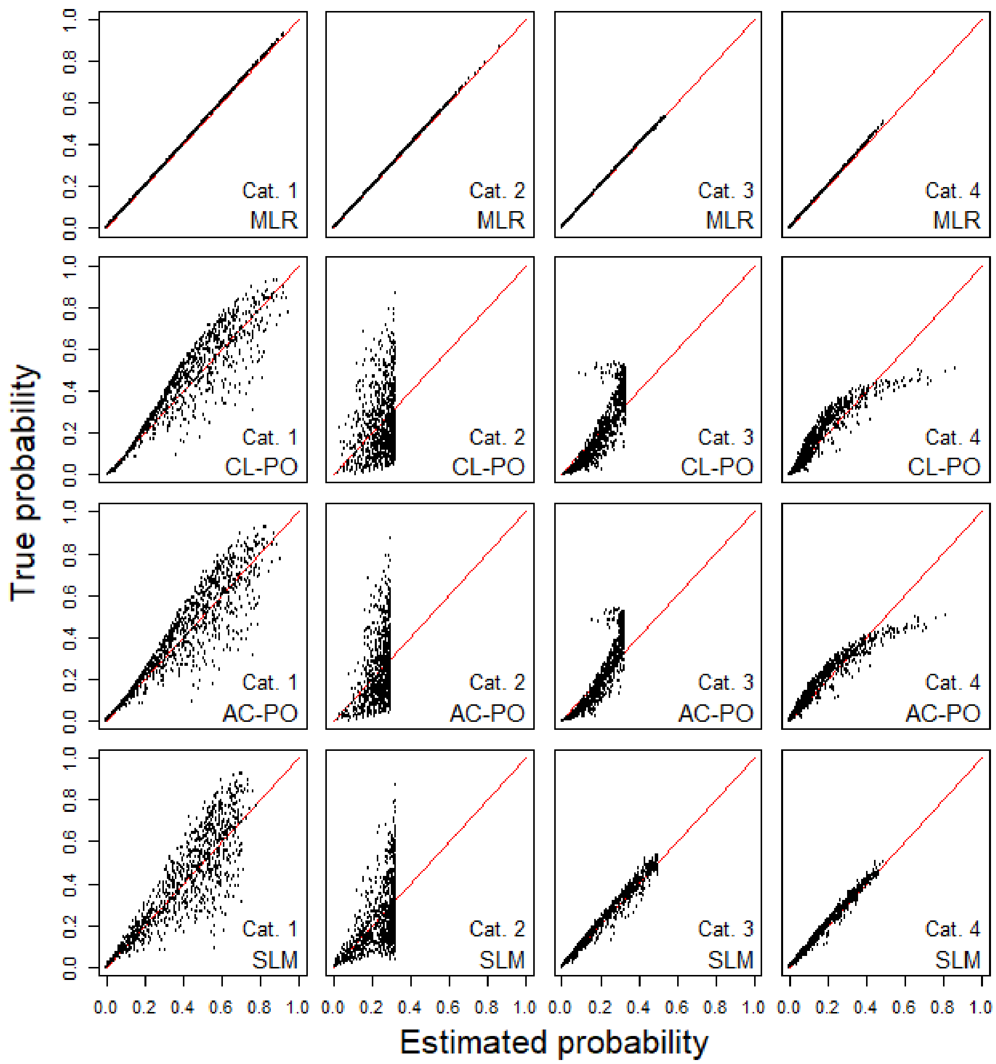


Figure S7. Scatter plots of true risks versus estimated risks for simulation scenario 7 when the true model has the form of a multinomial logistic regression. The plots are based on a random subset of 1,000 cases from all 200,000 cases.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

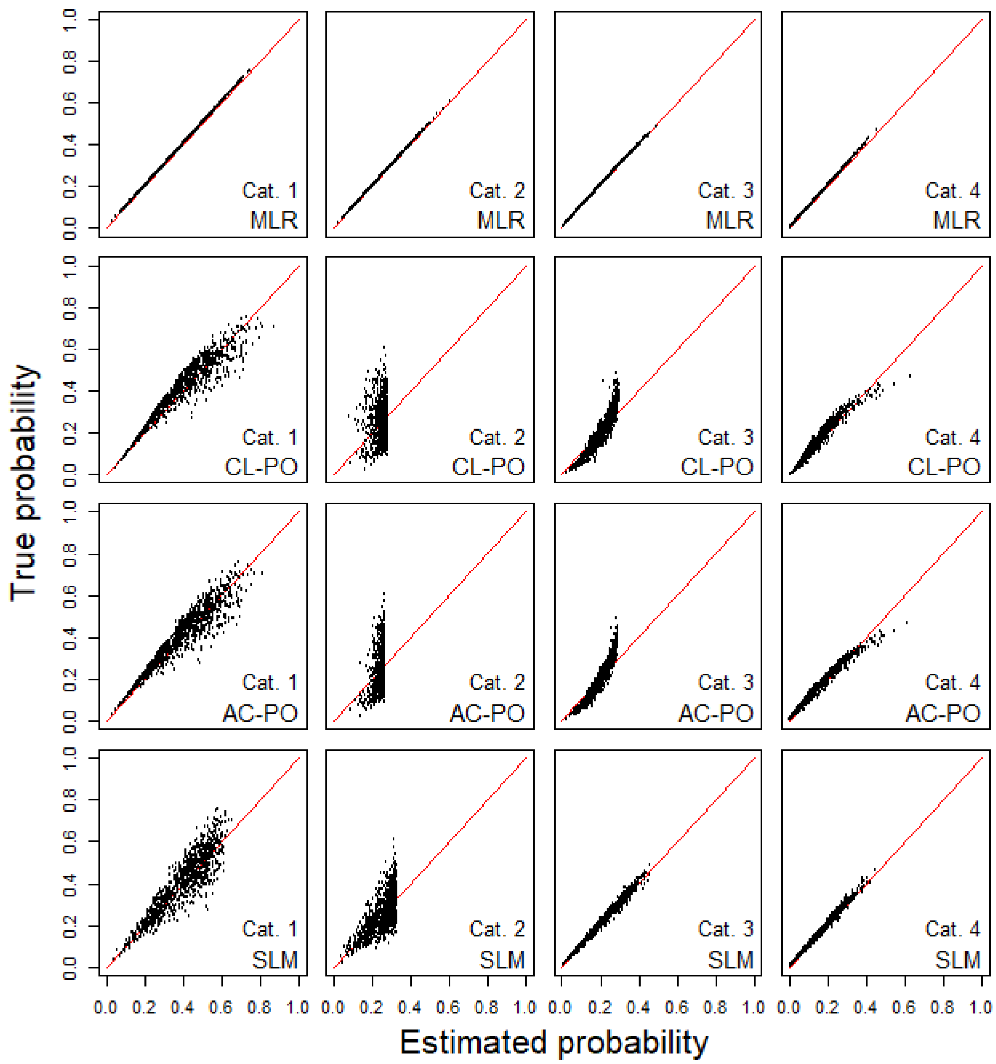


Figure S8. Scatter plots of true risks versus estimated risks for simulation scenario 8 when the true model has the form of a multinomial logistic regression. The plots are based on a random subset of 1,000 cases from all 200,000 cases.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

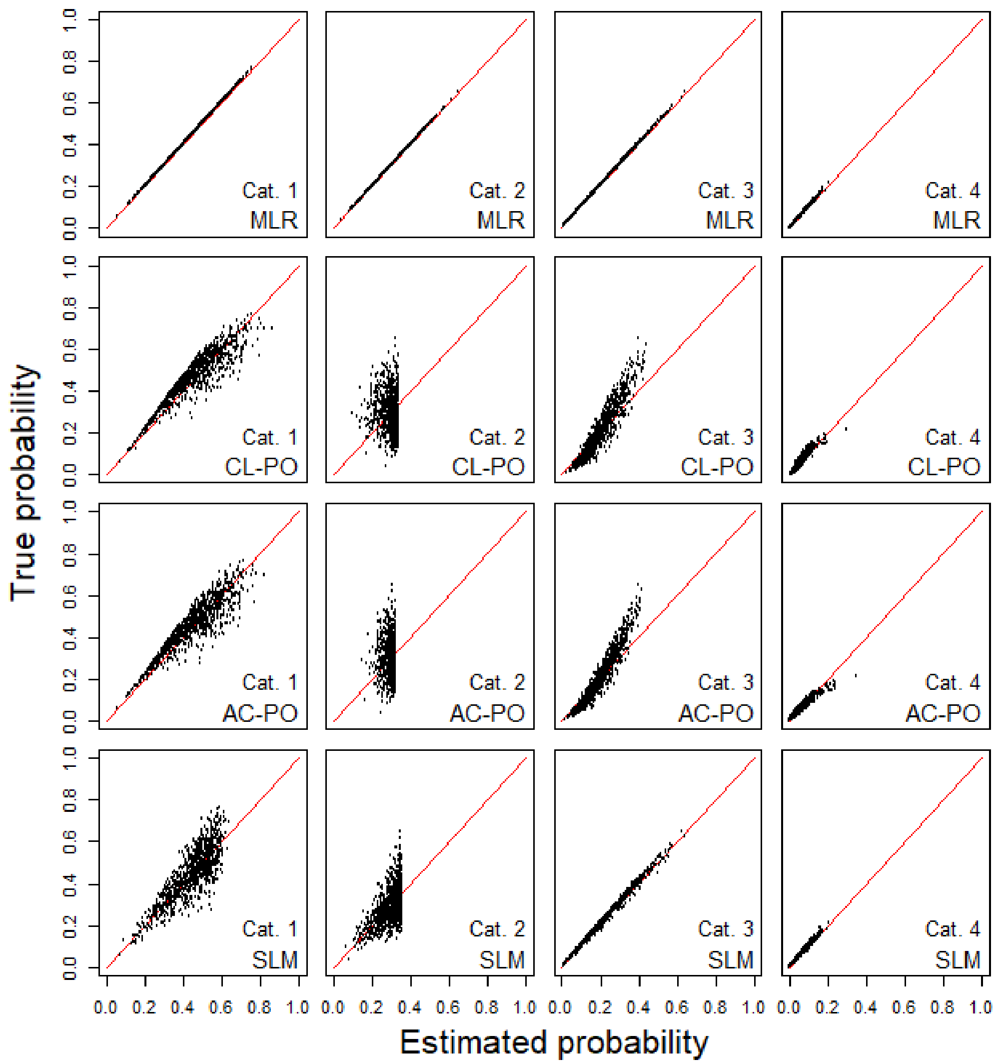


Figure S9. Scatter plots of true risks versus estimated risks for simulation scenario 9 when the true model has the form of a multinomial logistic regression.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

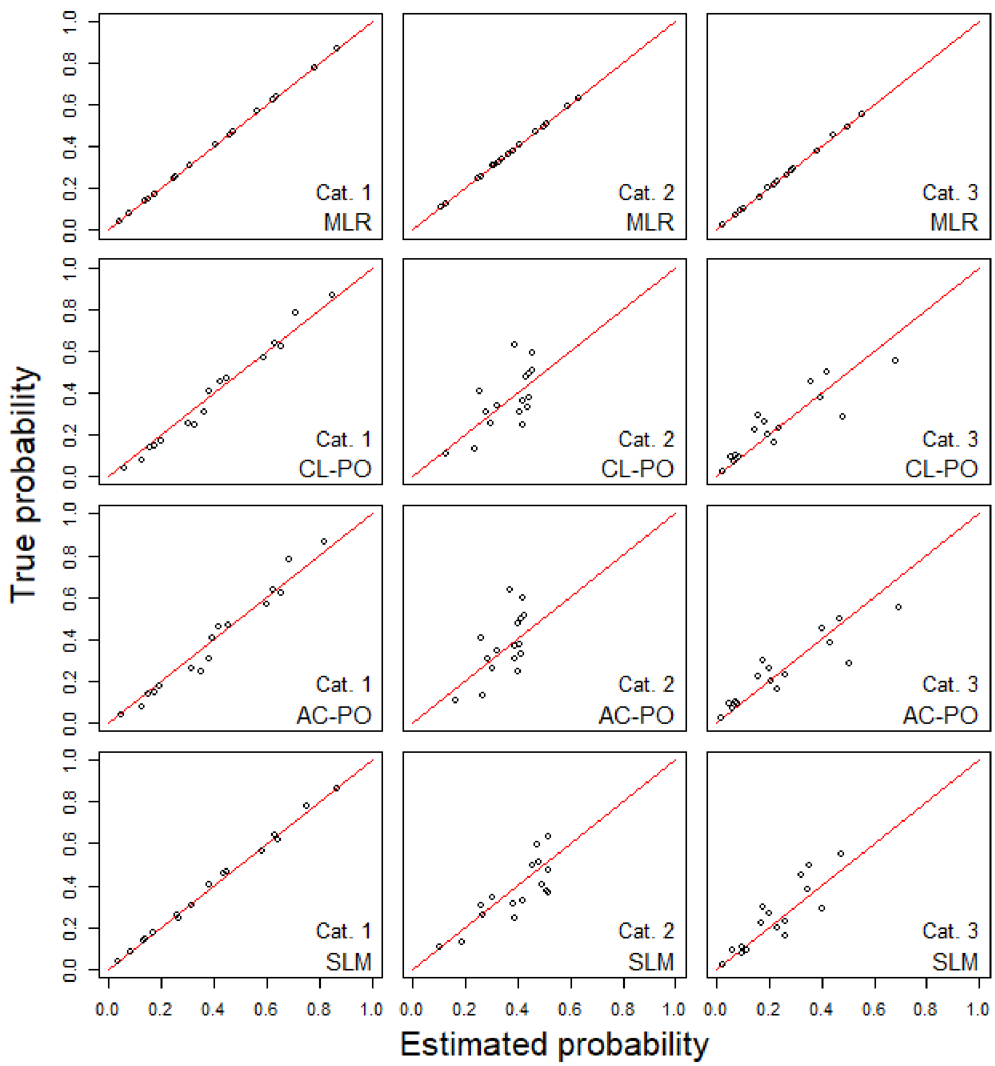


Figure S10. Scatter plots of true risks versus estimated risks for simulation scenario 10 when the true model has the form of a multinomial logistic regression.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

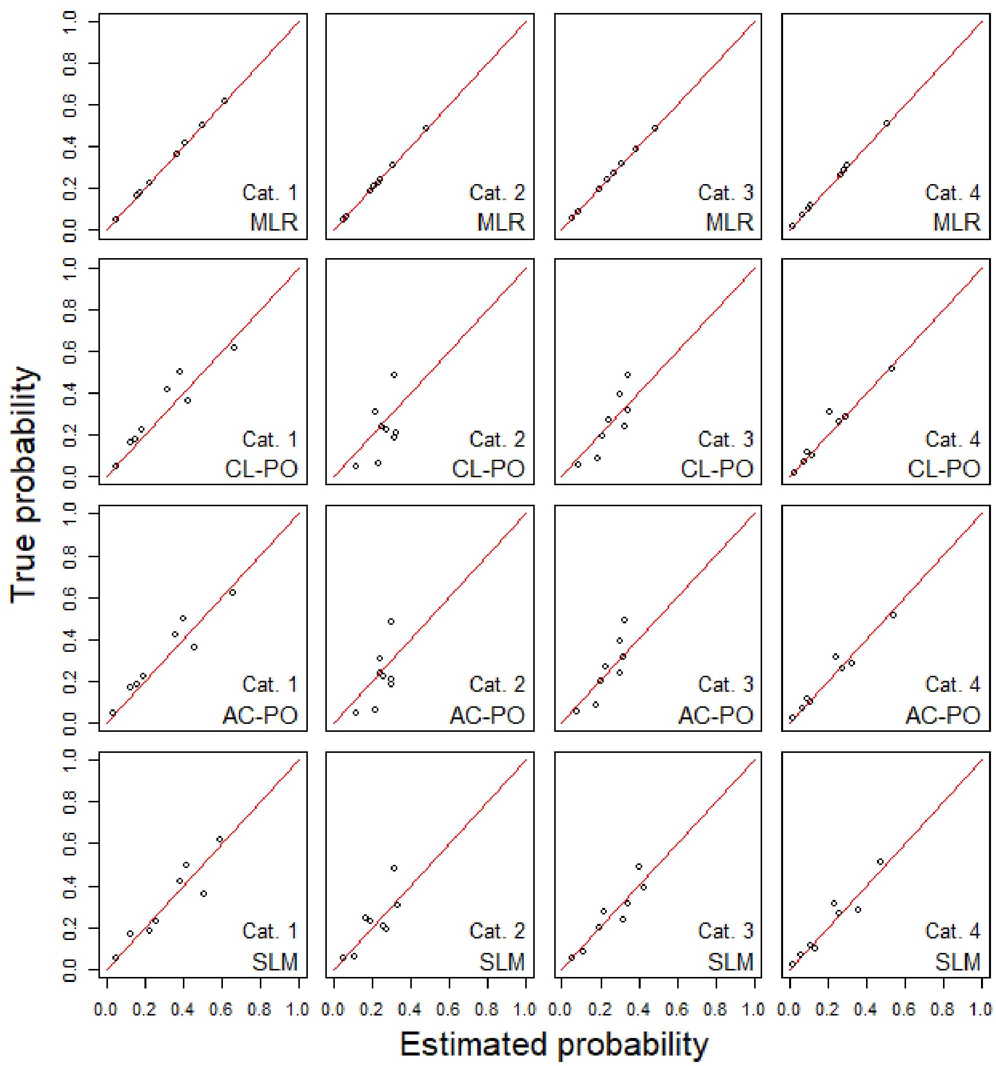


Figure S11. Scatter plots of true risks versus estimated risks for simulation scenario 11 when the true model has the form of a multinomial logistic regression. The plots are based on a random subset of 1,000 cases from all 200,000 cases.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

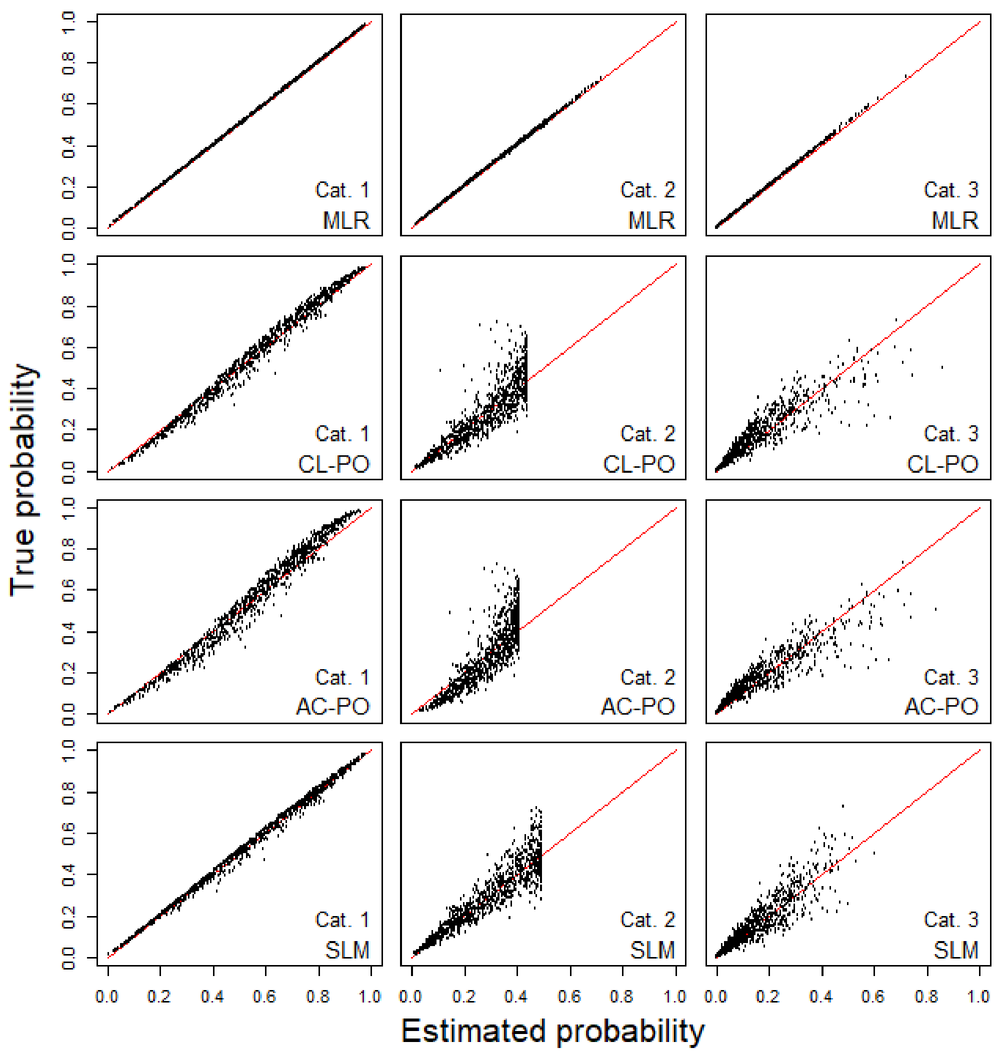


Figure S12. Flexible smoothed calibration curves per outcome category for simulation scenario 5 when the true model has the form of a multinomial logistic regression (green for category 1, orange for category 2, red for category 3). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

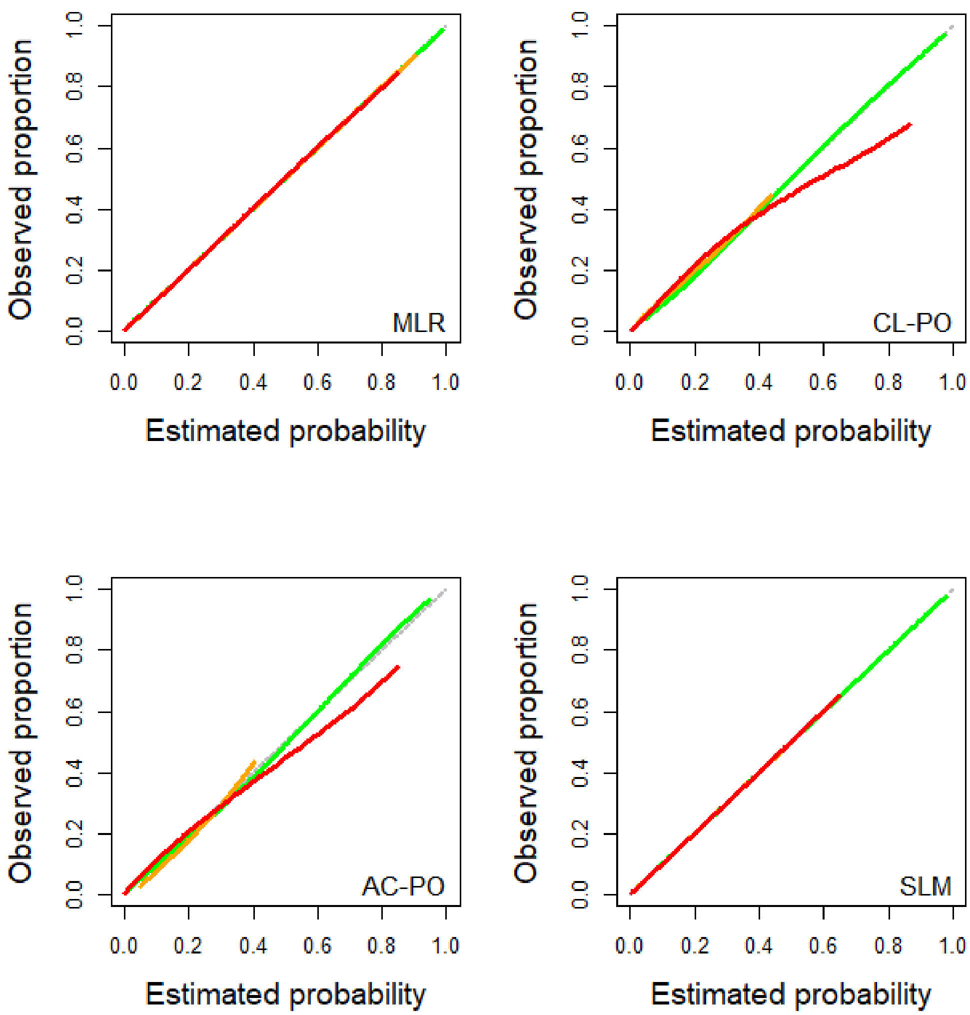


Figure S13. Flexible smoothed calibration curves per outcome category for simulation scenario 6 when the true model has the form of a multinomial logistic regression (green for category 1, orange for category 2, red for category 3, brown for category 4). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

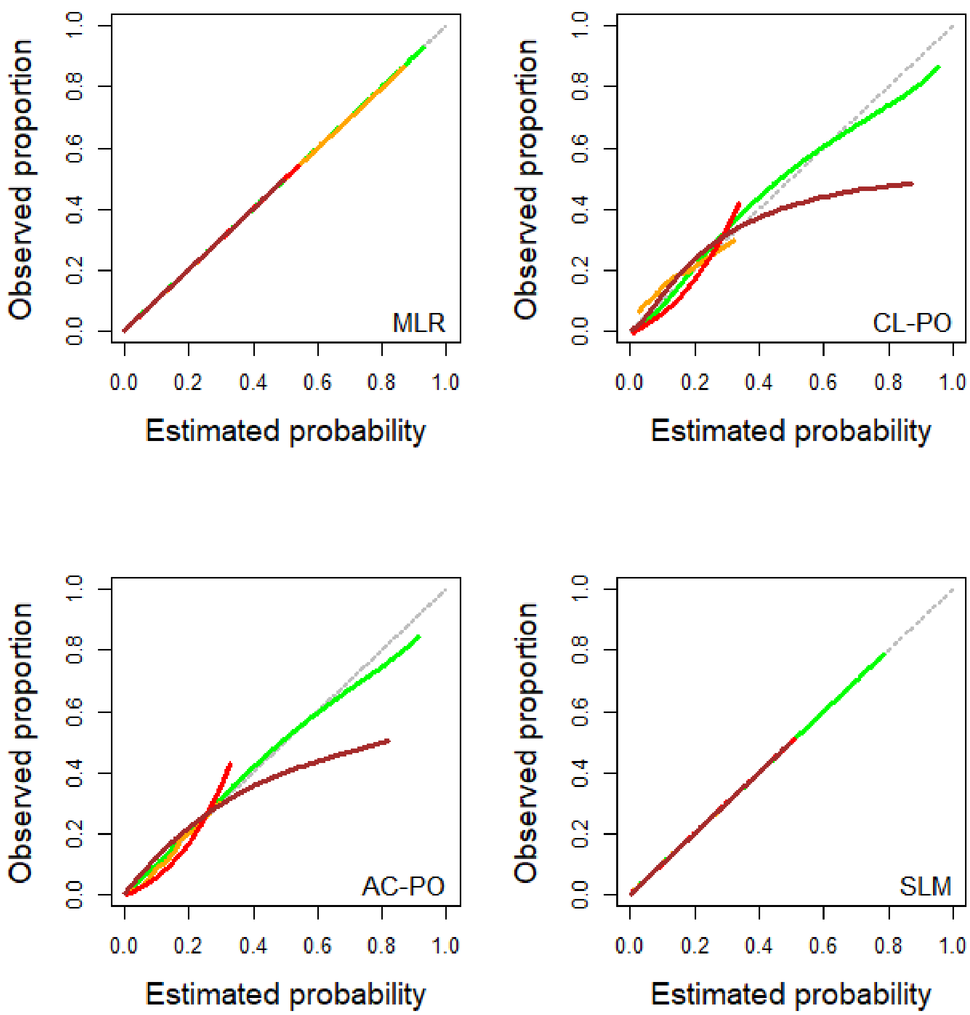


Figure S14. Flexible smoothed calibration curves per outcome category for simulation scenario 7 when the true model has the form of a multinomial logistic regression (green for category 1, orange for category 2, red for category 3, brown for category 4). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

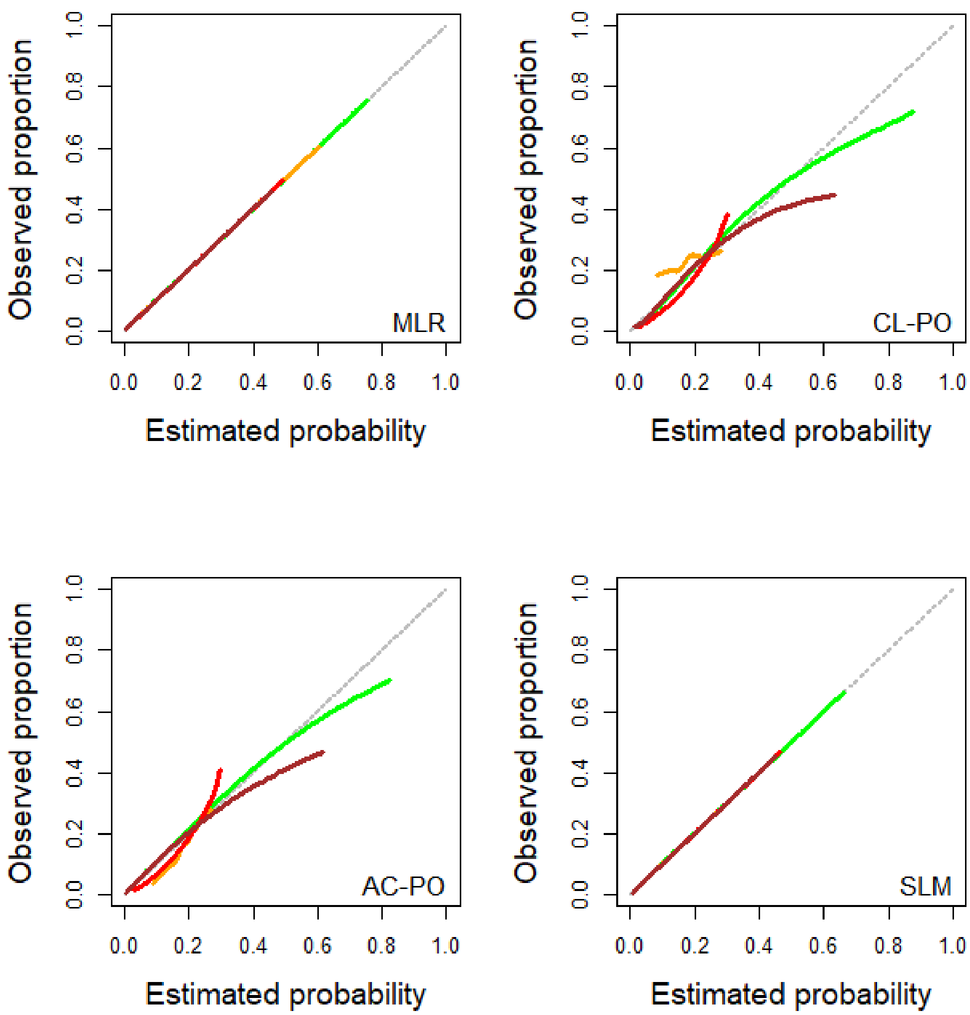


Figure S15. Flexible smoothed calibration curves per outcome category for simulation scenario 8 when the true model has the form of a multinomial logistic regression (green for category 1, orange for category 2, red for category 3, brown for category 4). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

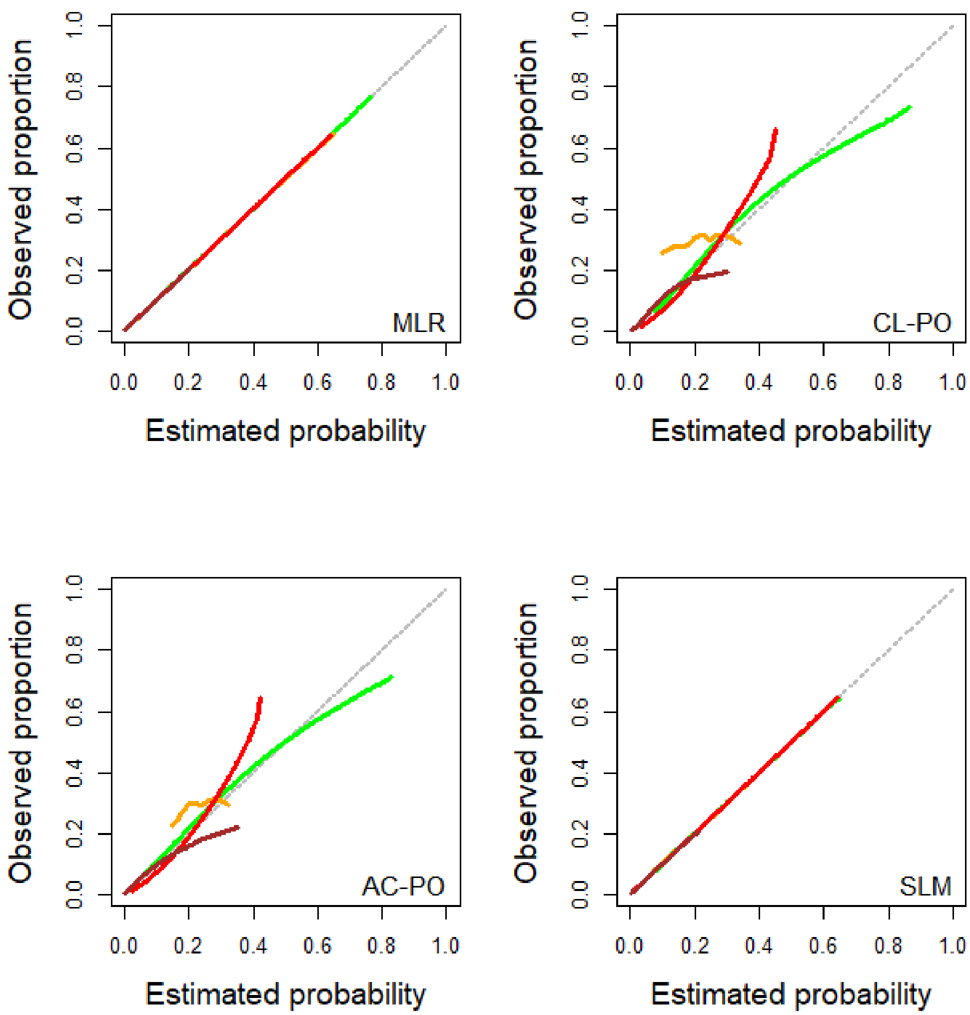


Figure S16. Calibration scatter plots per outcome category for simulation scenario 9 when the true model has the form of a multinomial logistic regression (green for category 1, orange for category 2, red for category 3). These graphs are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). Because all predictors are binary, no flexible curves are shown.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

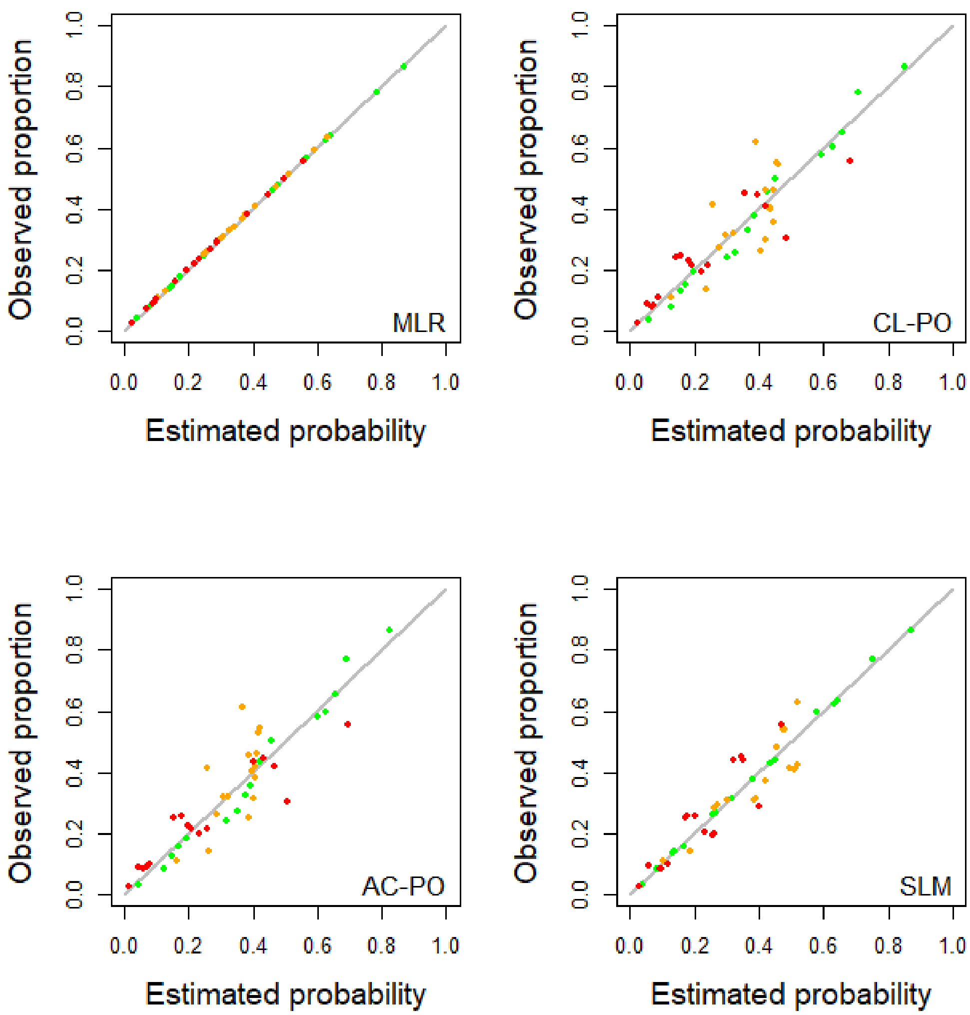


Figure S17. Calibration scatter plots per outcome category for simulation scenario 10 when the true model has the form of a multinomial logistic regression (green for category 1, orange for category 2, red for category 3, brown for category 4). These graphs are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). Because all predictors are binary, no flexible curves are shown.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

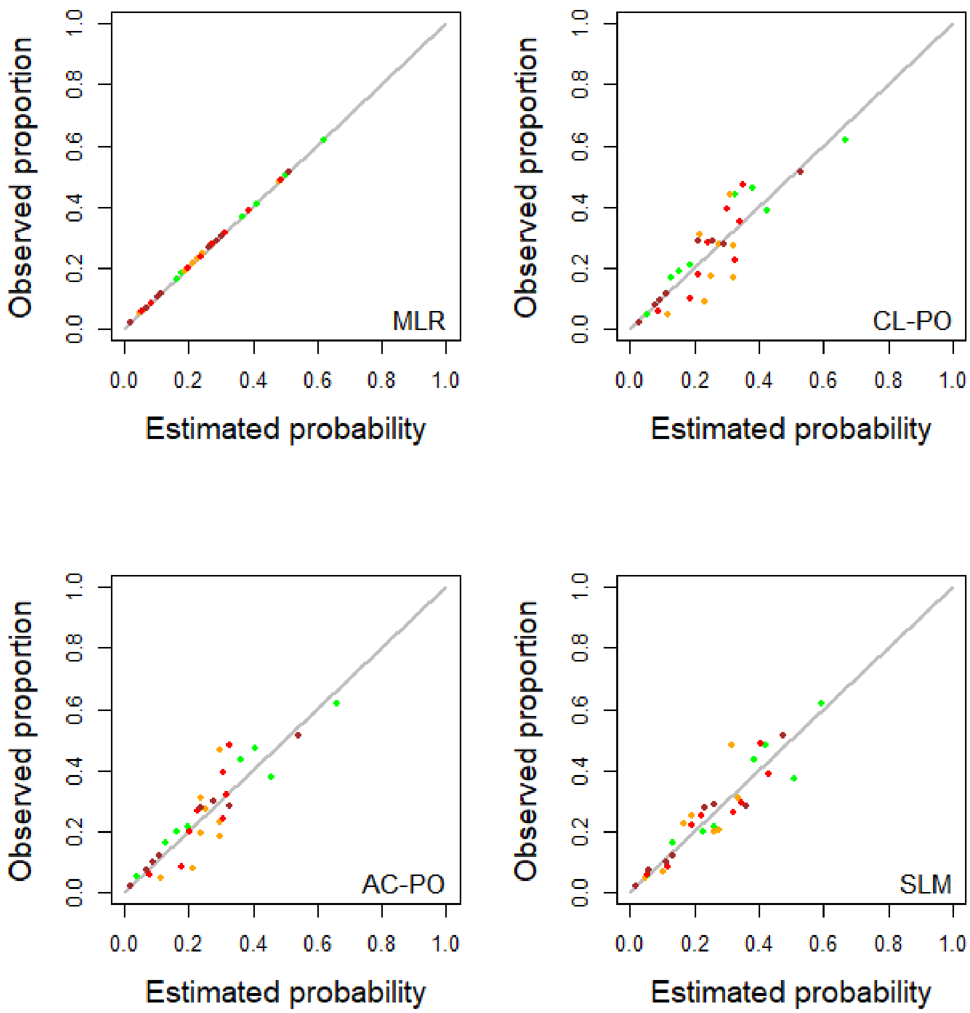


Figure S18. Flexible smoothed calibration curves per outcome category for simulation scenario 11 when the true model has the form of a multinomial logistic regression (green for category 1, orange for category 2, red for category 3). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

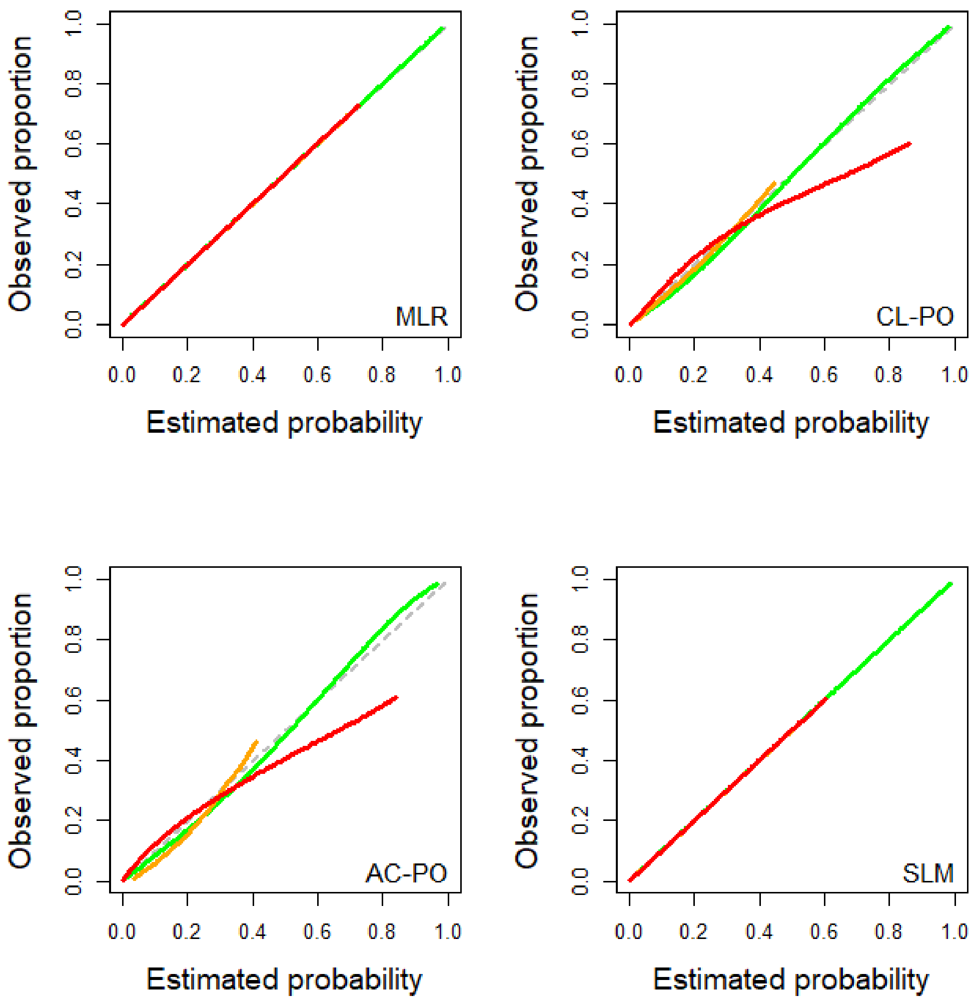


Figure S19. Scatter plots of true risks versus estimated risks for simulation scenario 4 when the true model has the form of a cumulative logit model with proportional odds. The plots are based on a random subset of 1,000 cases from all 200,000 cases.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

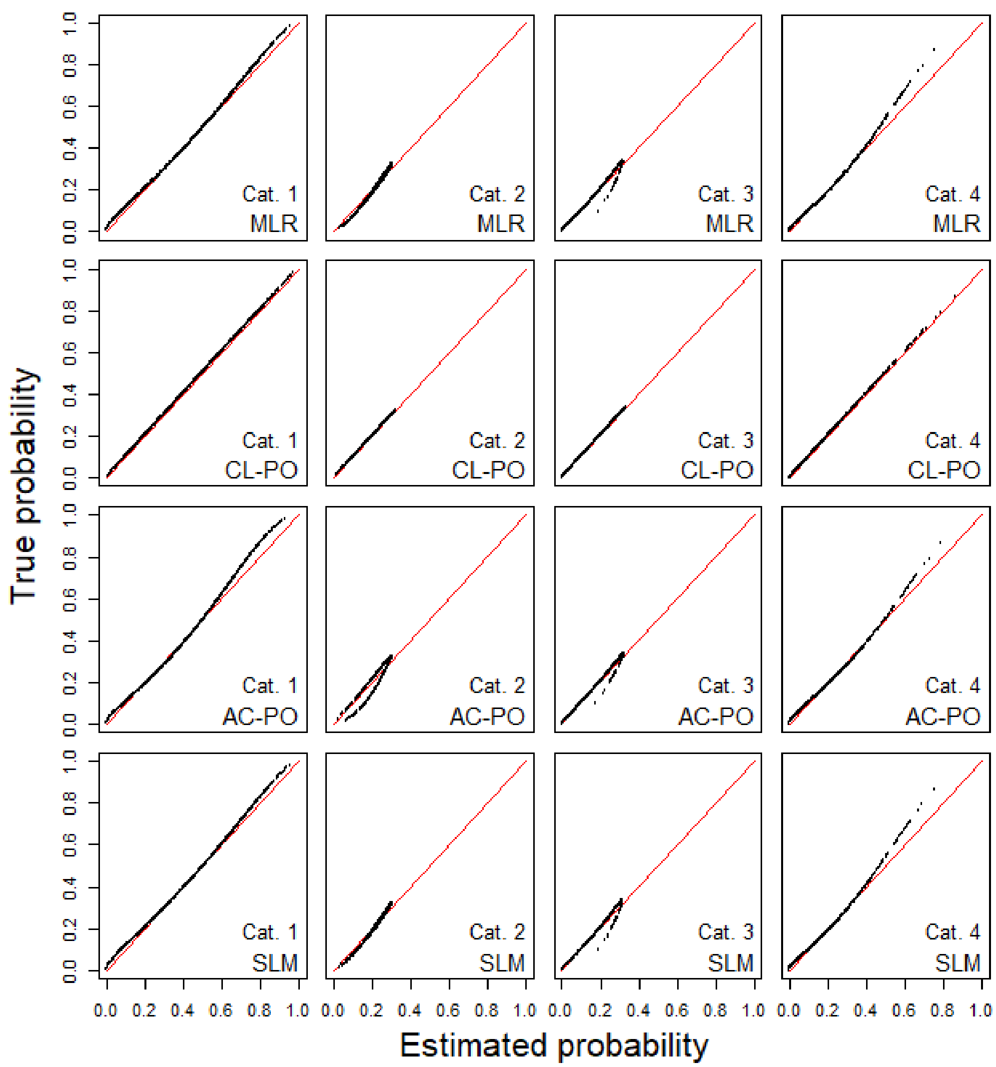


Figure S20. Scatter plots of true risks versus estimated risks for simulation scenario 5 when the true model has the form of a cumulative logit model with proportional odds. The plots are based on a random subset of 1,000 cases from all 200,000 cases.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

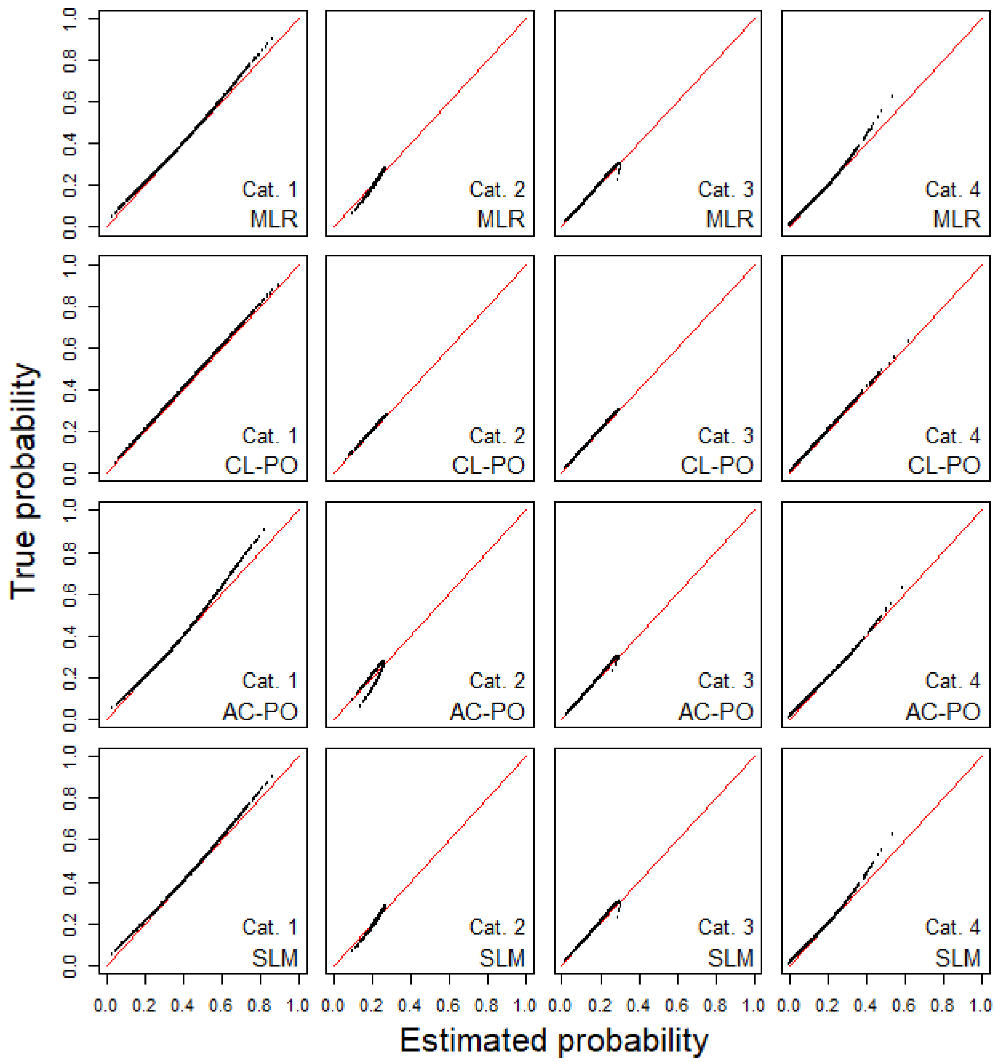


Figure S21. Scatter plots of true risks versus estimated risks for simulation scenario 6 when the true model has the form of a cumulative logit model with proportional odds. The plots are based on a random subset of 1,000 cases from all 200,000 cases.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

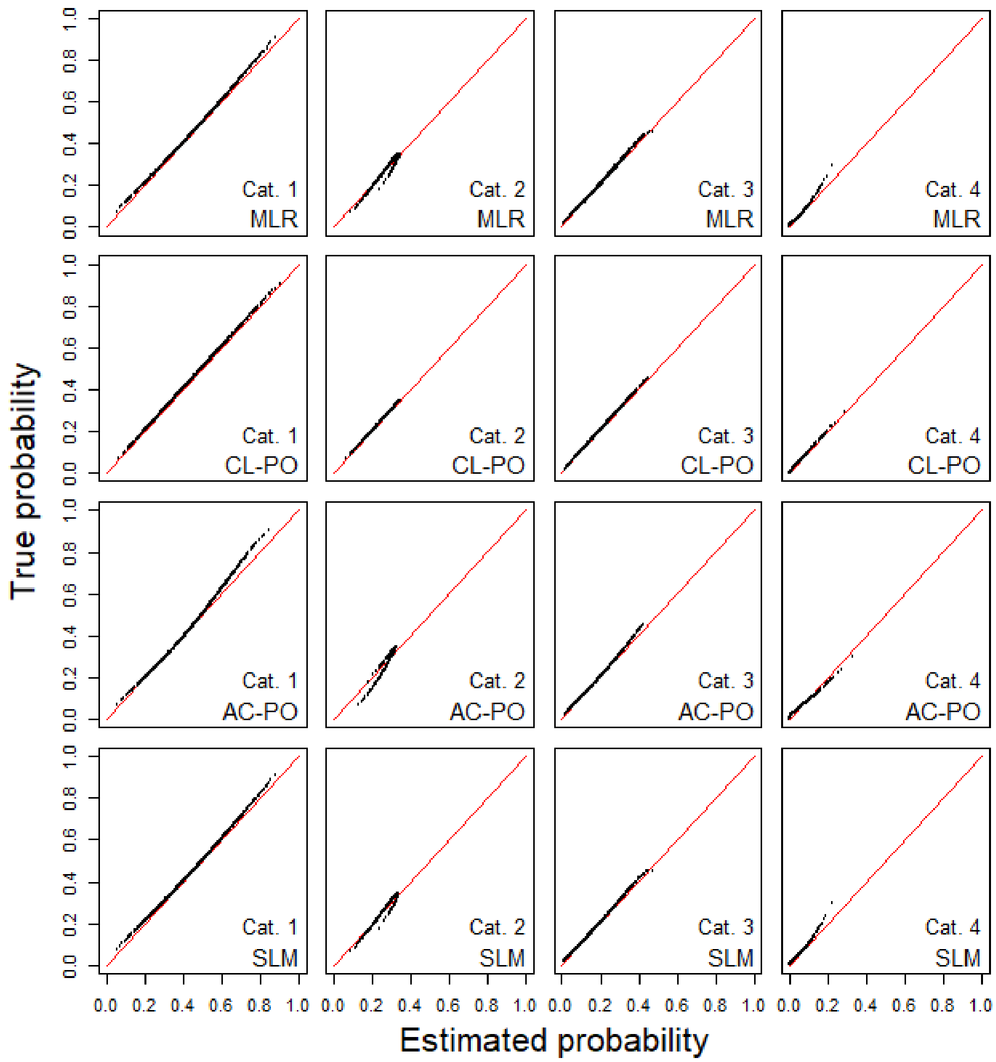


Figure S22. Scatter plots of true risks versus estimated risks for simulation scenario 7 when the true model has the form of a cumulative logit model with proportional odds.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

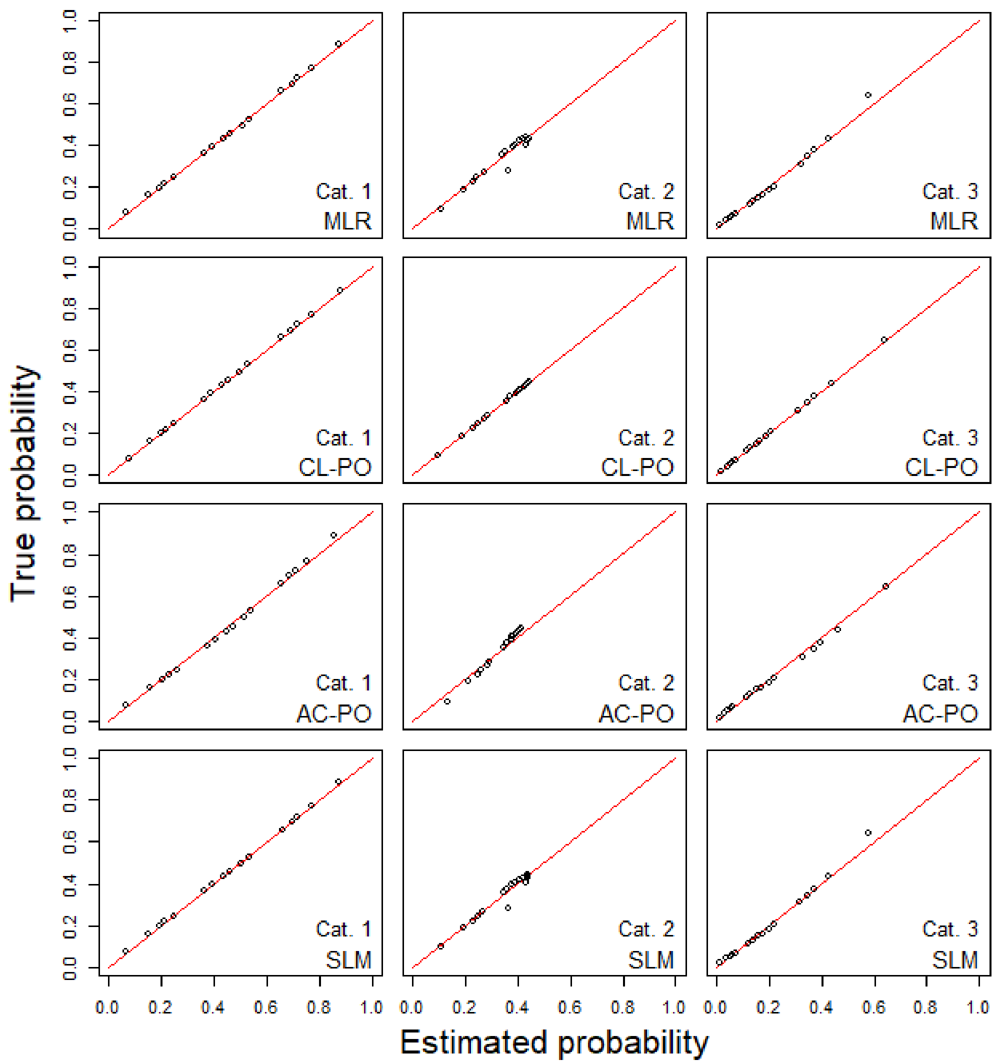


Figure S23. Scatter plots of true risks versus estimated risks for simulation scenario 8 when the true model has the form of a cumulative logit model with proportional odds.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

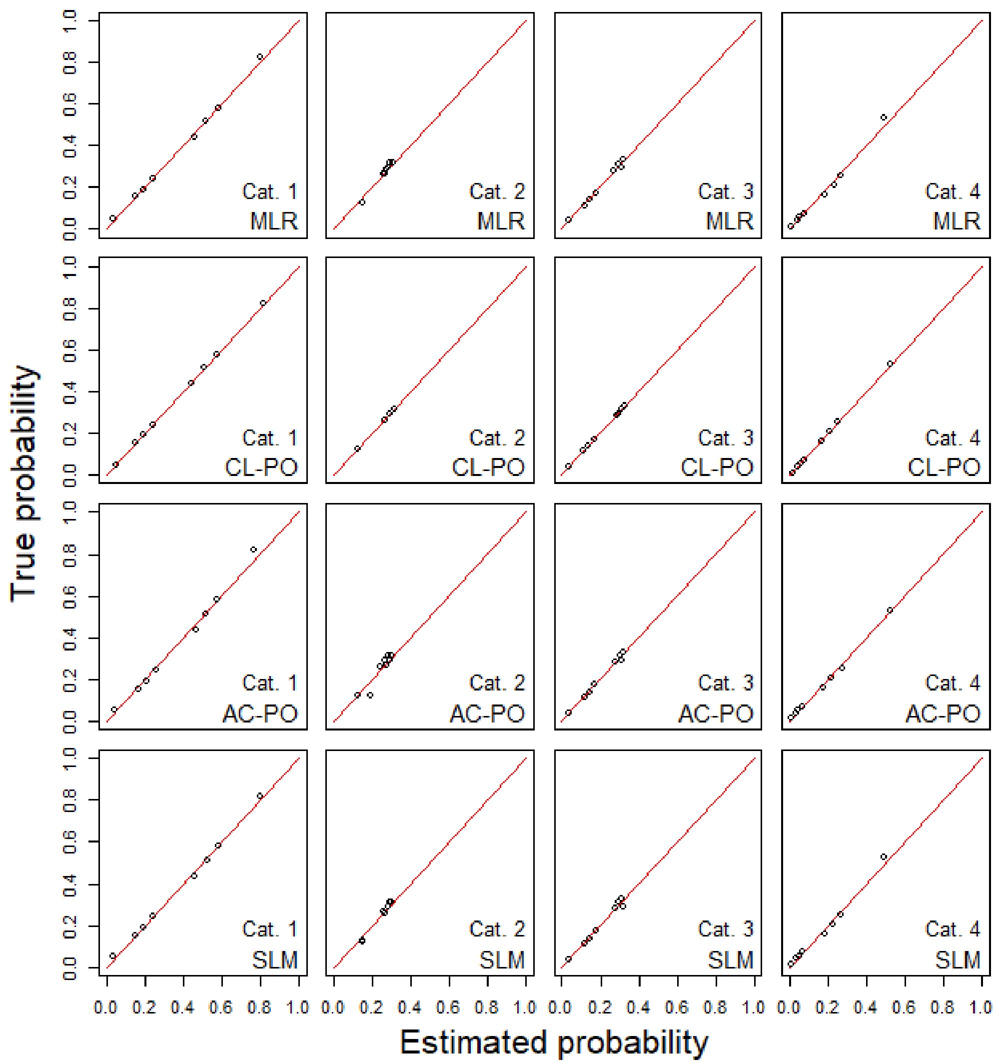


Figure S24. Scatter plots of true risks versus estimated risks for simulation scenario 9 when the true model has the form of a cumulative logit model with proportional odds. The plots are based on a random subset of 1,000 cases from all 200,000 cases.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

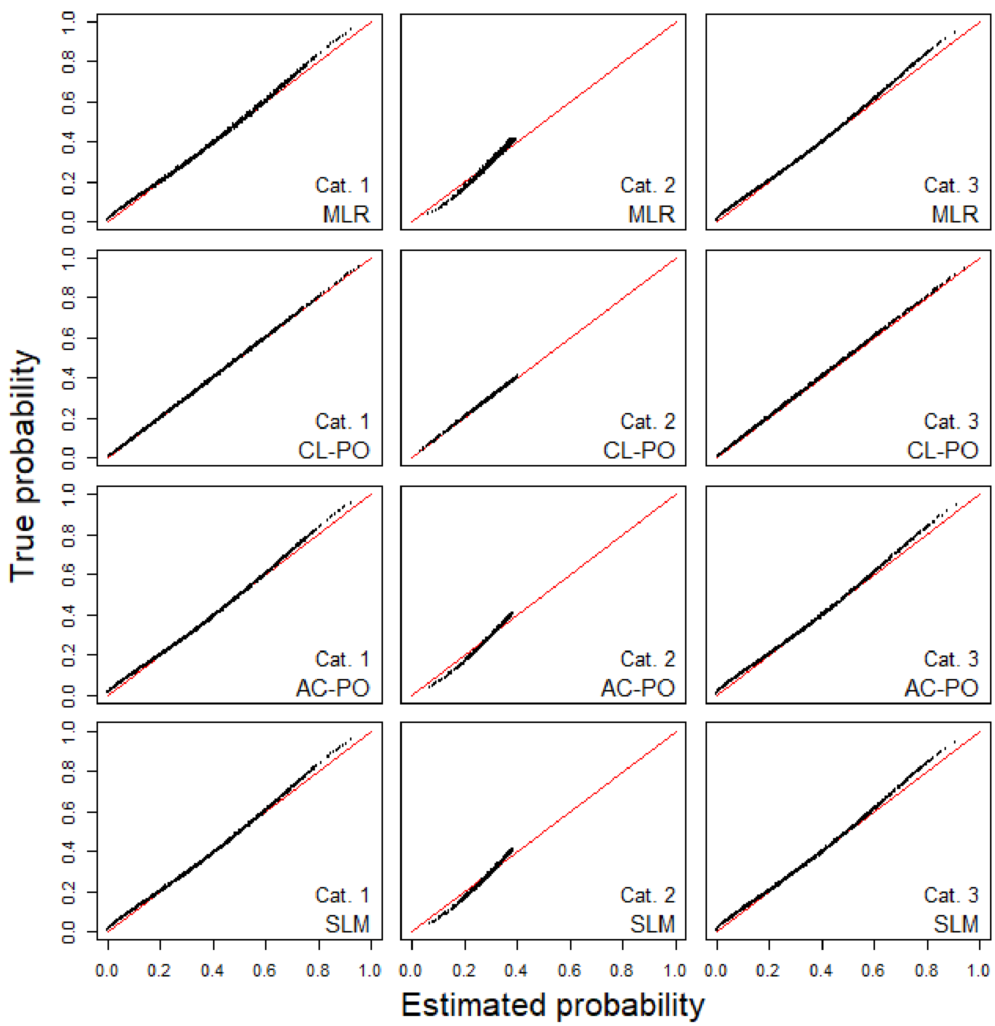


Figure S25. Flexible smoothed calibration curves per outcome category for simulation scenario 1 when the true model has the form of a cumulative logit model with proportional odds (green for category 1, orange for category 2, red for category 3). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

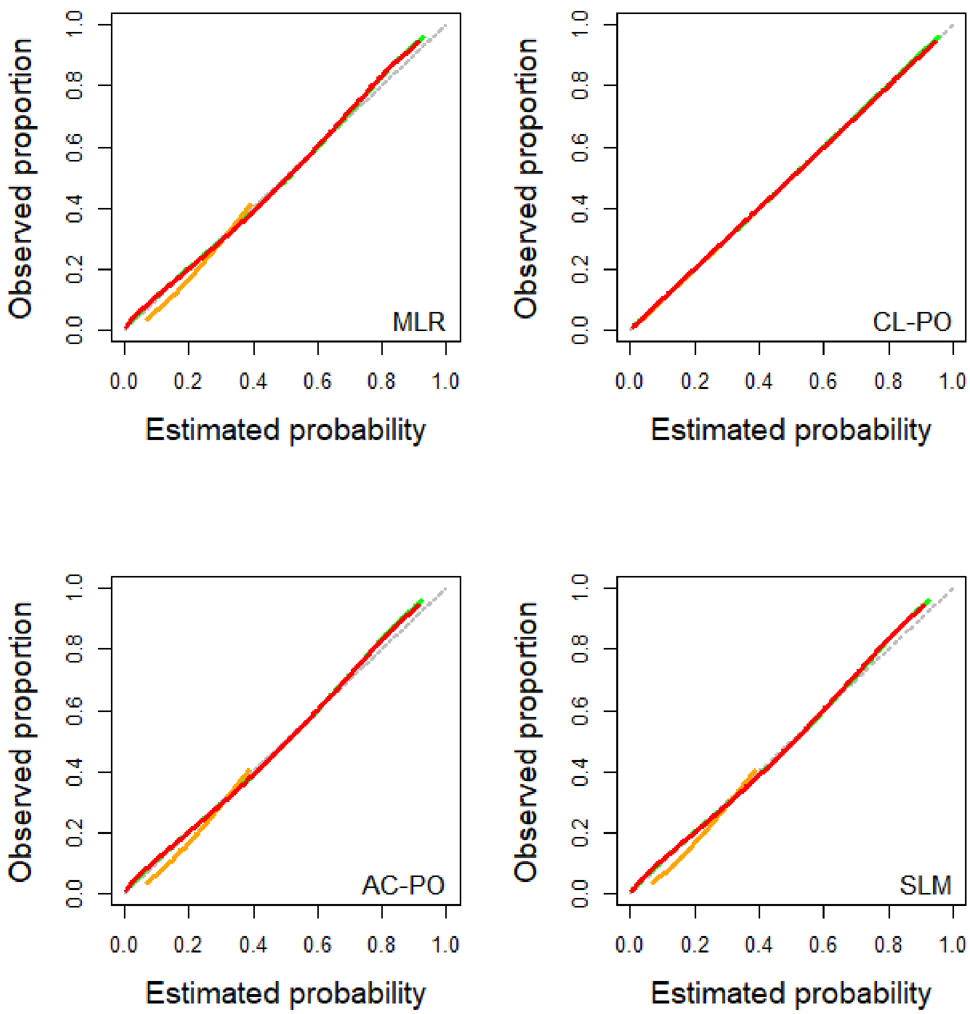


Figure S26. Flexible smoothed calibration curves per outcome category for simulation scenario 2 when the true model has the form of a cumulative logit model with proportional odds (green for category 1, orange for category 2, red for category 3). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

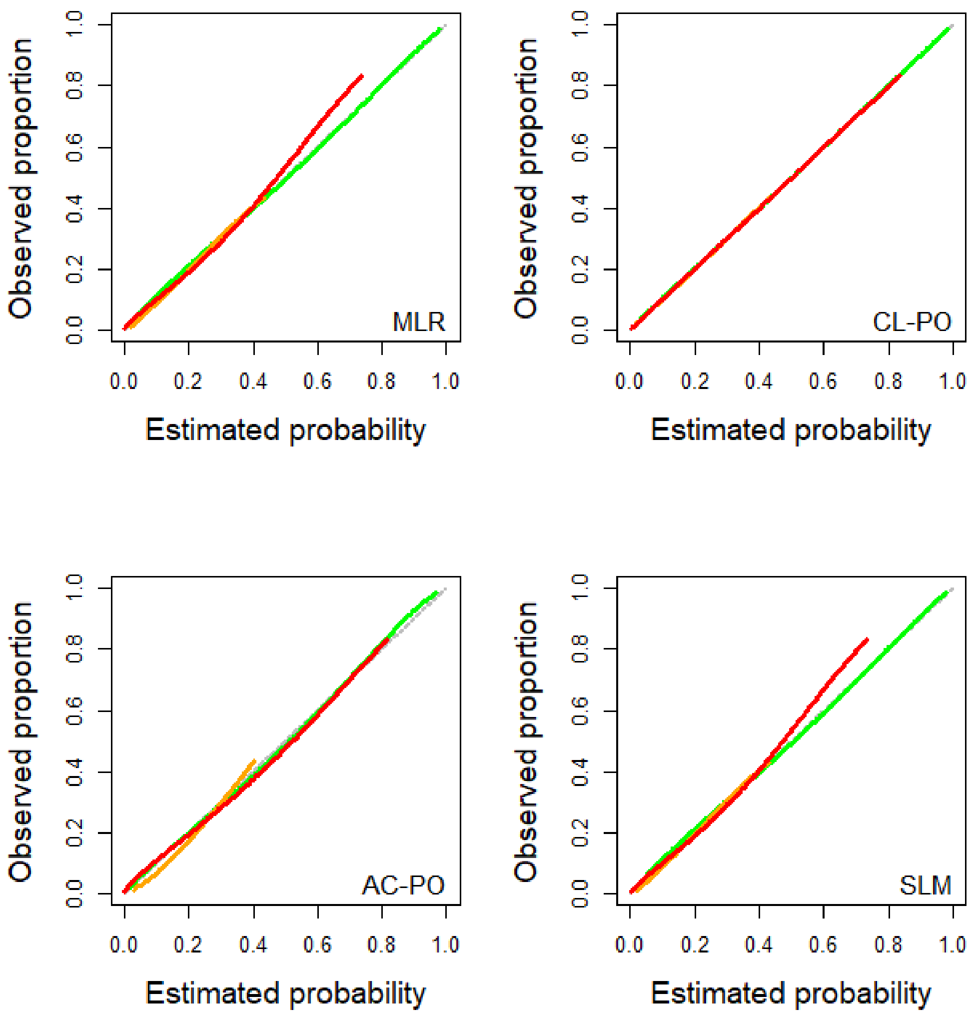


Figure S27. Flexible smoothed calibration curves per outcome category for simulation scenario 3 when the true model has the form of a cumulative logit model with proportional odds (green for category 1, orange for category 2, red for category 3). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

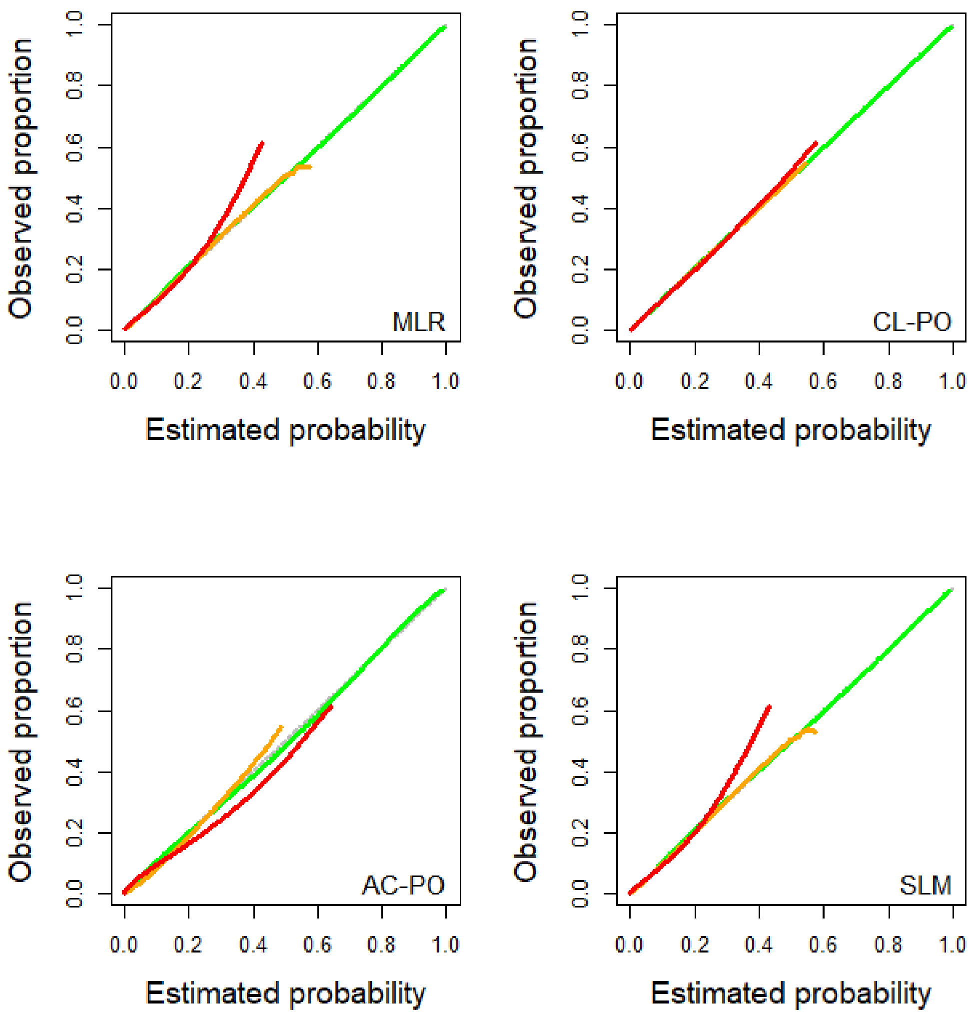


Figure S28. Flexible smoothed calibration curves per outcome category for simulation scenario 4 when the true model has the form of a cumulative logit model with proportional odds (green for category 1, orange for category 2, red for category 3, brown for category 4). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

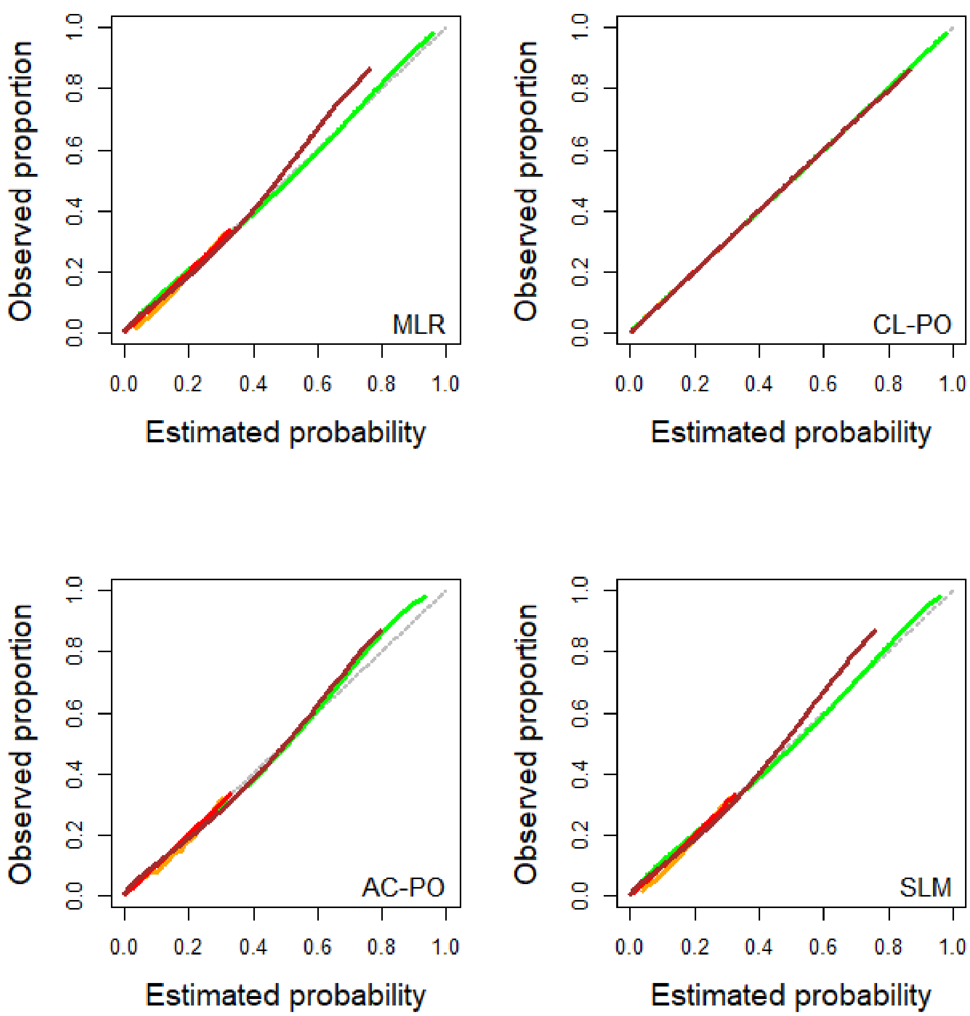


Figure S29. Flexible smoothed calibration curves per outcome category for simulation scenario 5 when the true model has the form of a cumulative logit model with proportional odds (green for category 1, orange for category 2, red for category 3, brown for category 4). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

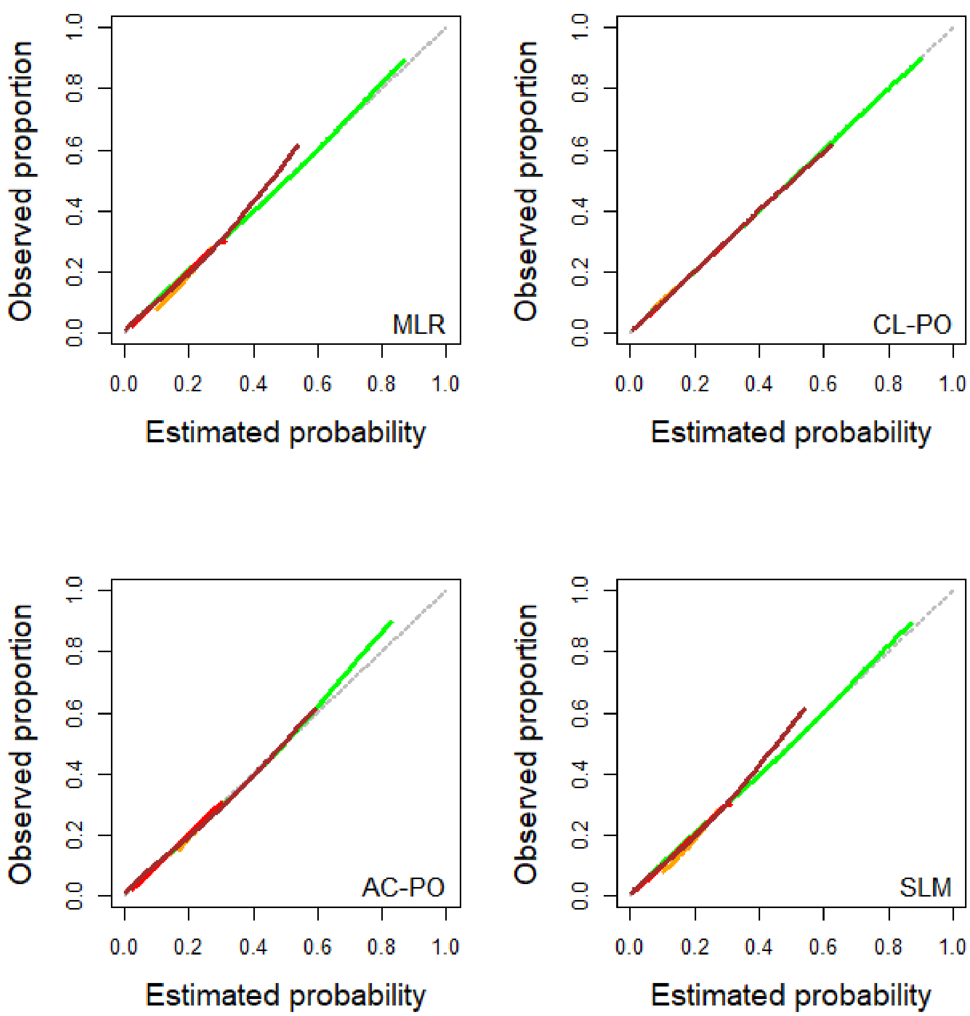


Figure S30. Flexible smoothed calibration curves per outcome category for simulation scenario 6 when the true model has the form of a cumulative logit model with proportional odds (green for category 1, orange for category 2, red for category 3, brown for category 4). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

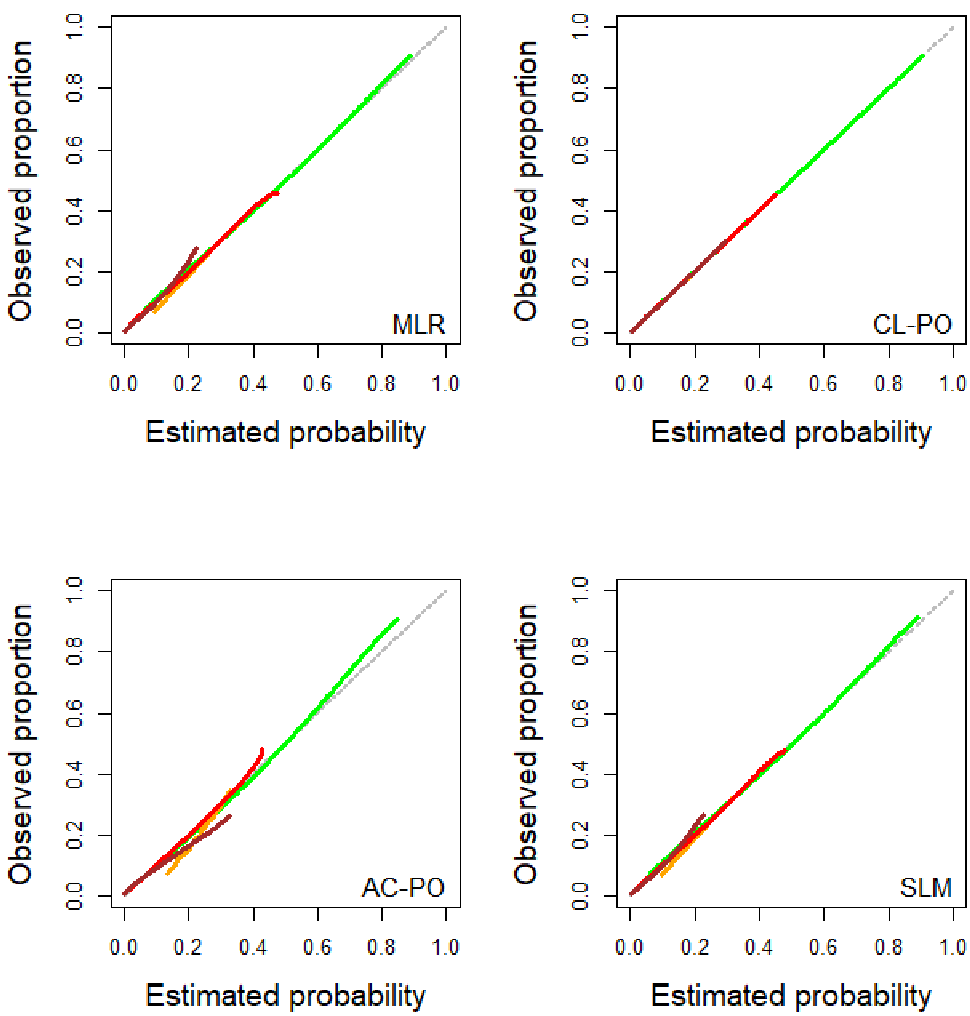


Figure S31. Calibration scatter plots per outcome category for simulation scenario 7 when the true model has the form of a cumulative logit model with proportional odds (green for category 1, orange for category 2, red for category 3). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). Because all predictors are binary, no flexible curves are shown.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

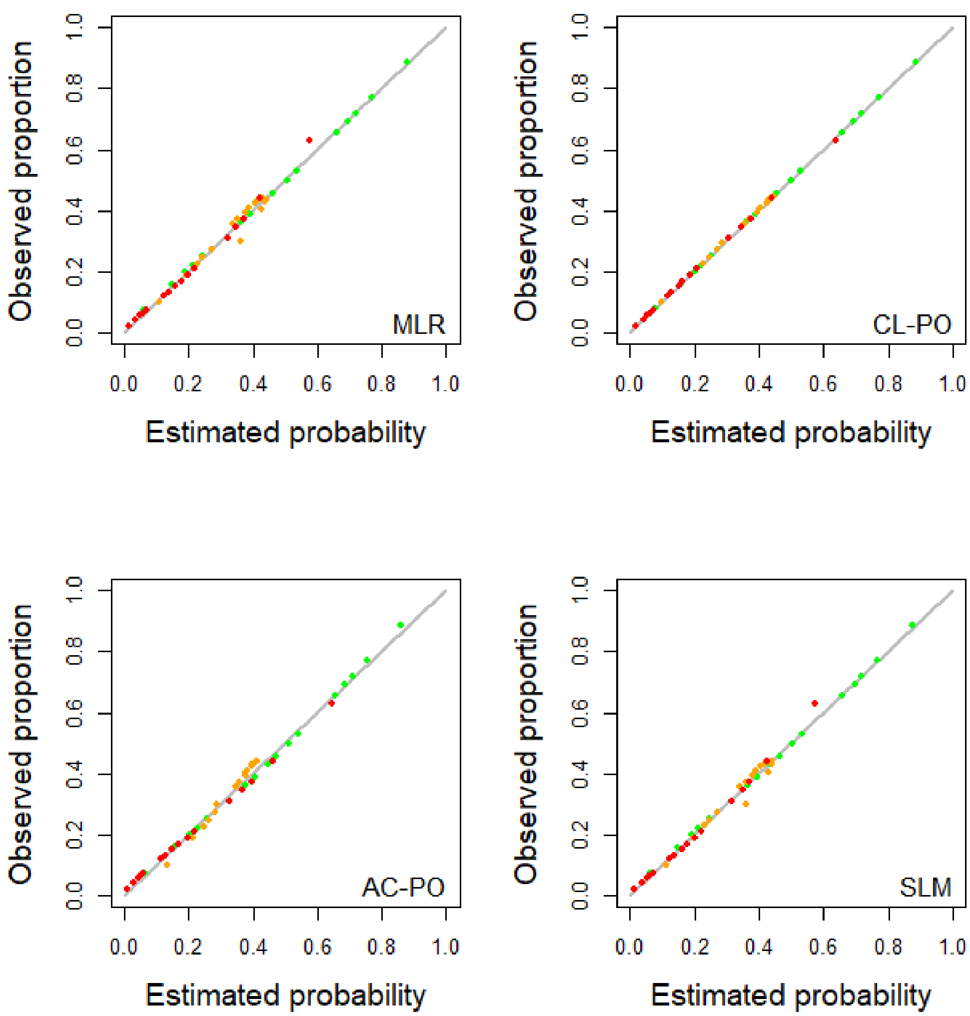


Figure S32. Calibration scatter plots per outcome category for simulation scenario 8 when the true model has the form of a cumulative logit model with proportional odds (green for category 1, orange for category 2, red for category 3, brown for category 4). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). Because all predictors are binary, no flexible curves are shown.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

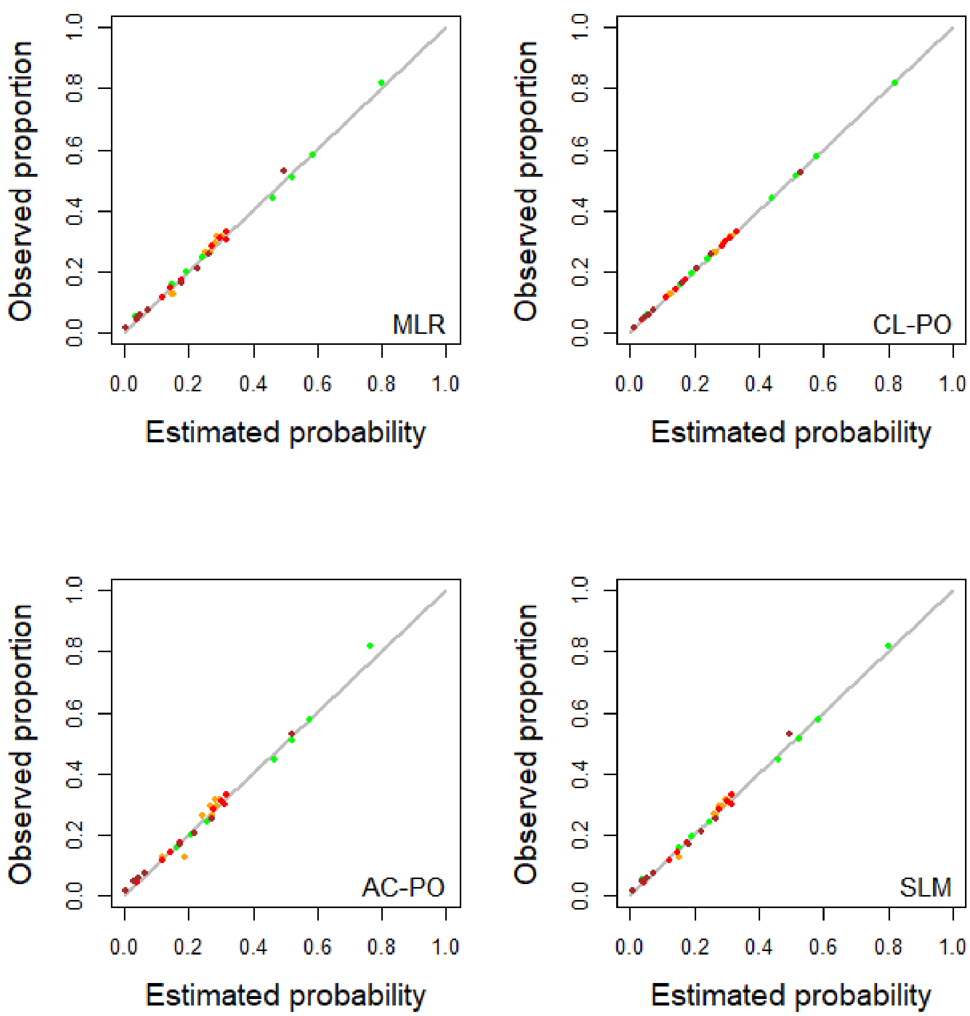


Figure S33. Flexible smoothed calibration curves per outcome category for simulation scenario 9 when the true model has the form of a cumulative logit model with proportional odds (green for category 1, orange for category 2, red for category 3, brown for category 4). These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves (n=200,000). For some models, lines overlap.

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; SLM, stereotype logit model.

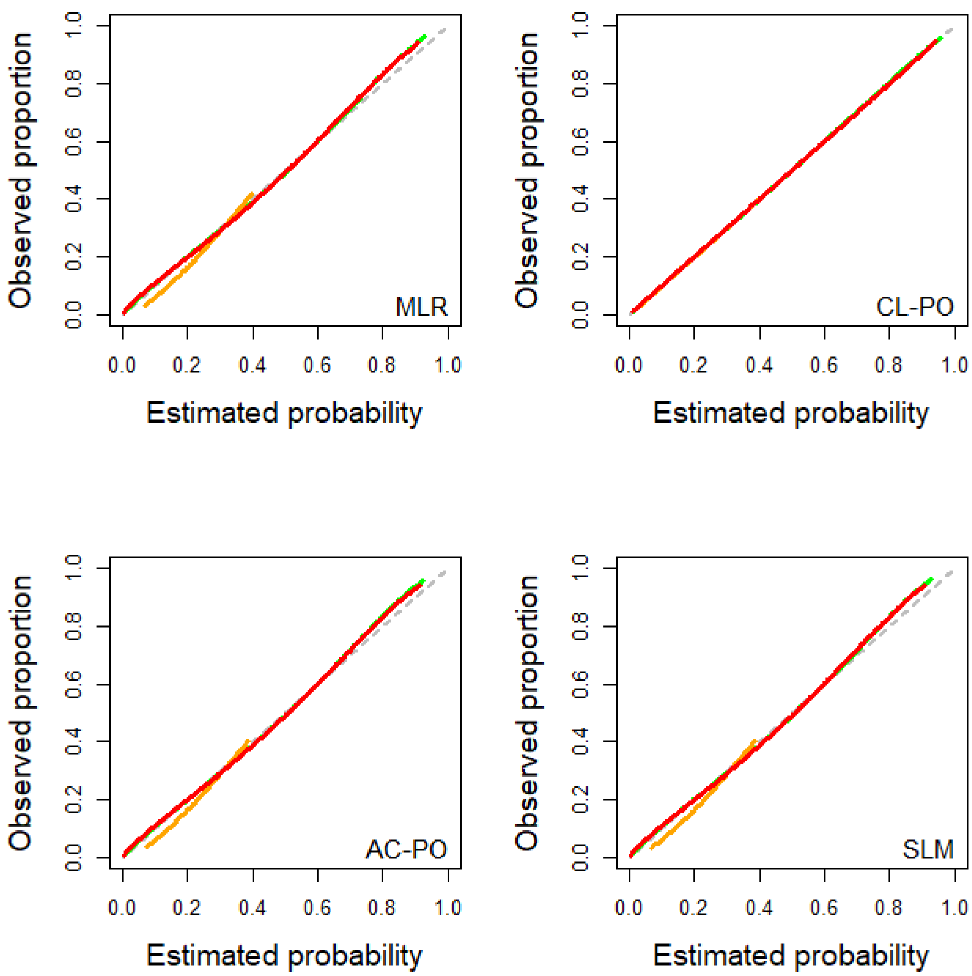


Figure S34. Scatter plot of estimated probabilities for having no coronary artery disease in the case study (n=4,888).

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; CR-PO, continuation ratio logit model with proportional odds; CR-NP, continuation ratio logit model without proportional odds; SLM, stereotype logit model

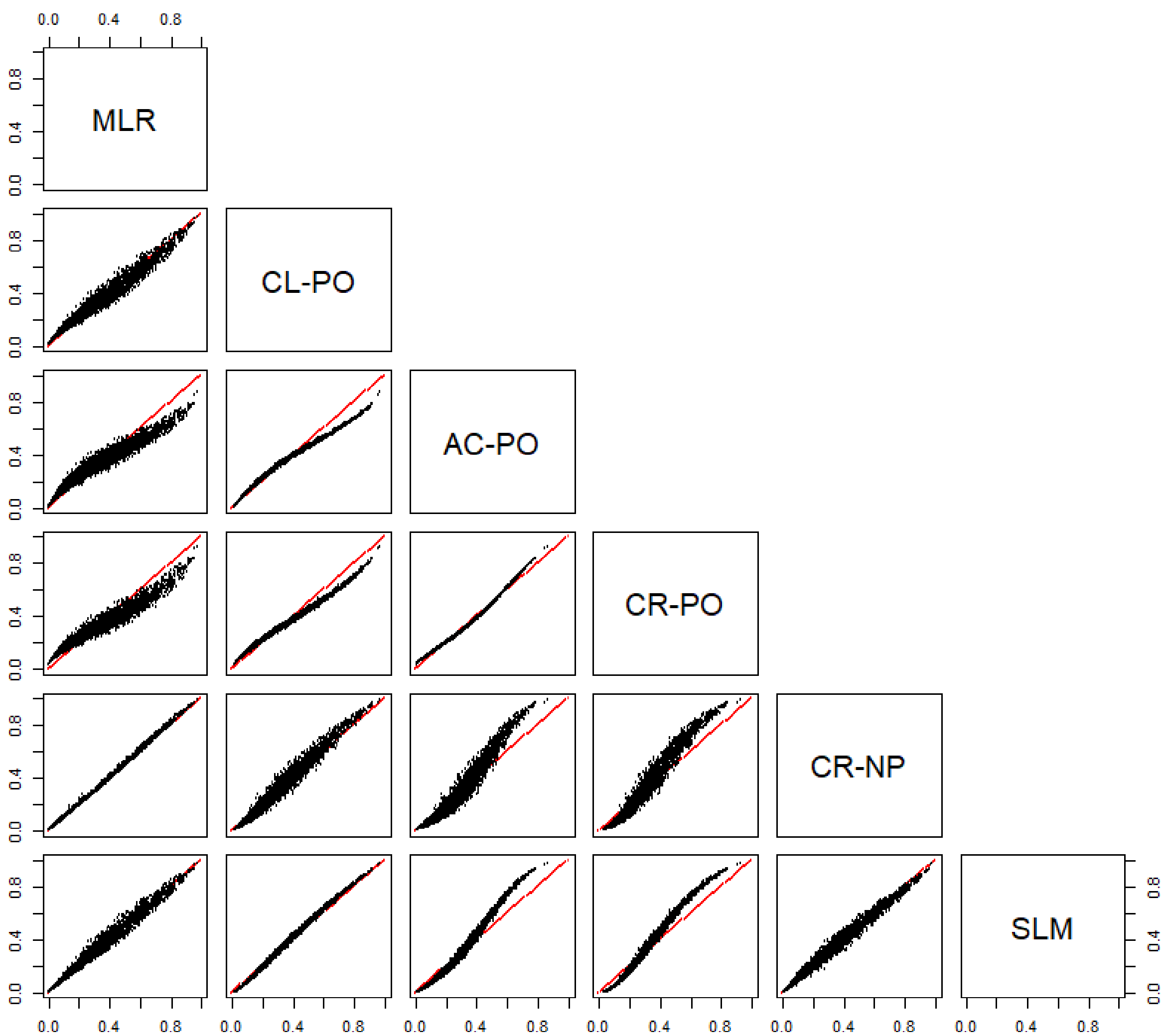


Figure S35. Scatter plot of estimated probabilities for having one-vessel disease in the case study (n=4,888).

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; CR-PO, continuation ratio logit model with proportional odds; CR-NP, continuation ratio logit model without proportional odds; SLM, stereotype logit model

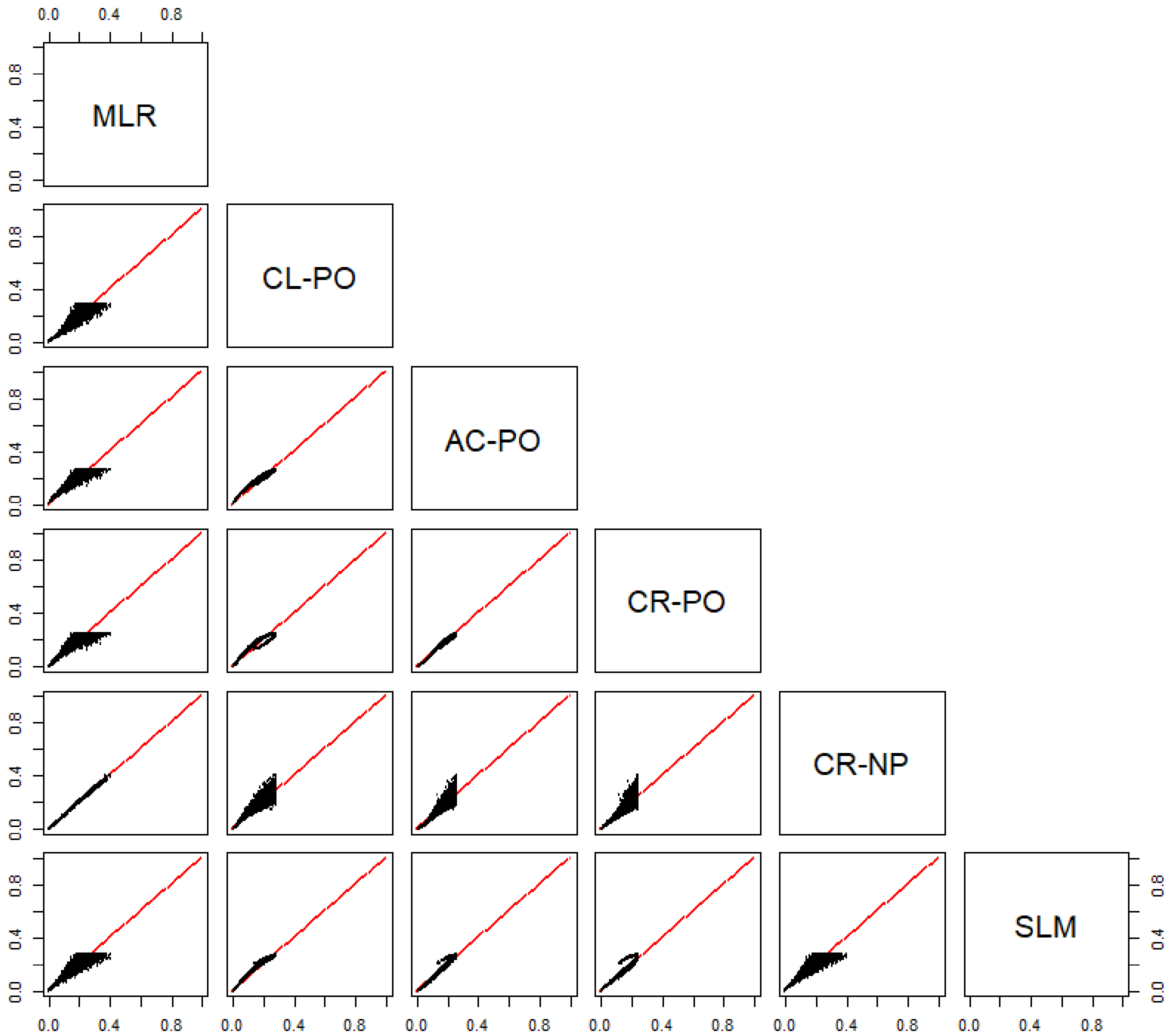


Figure S36. Scatter plot of estimated probabilities for having two-vessel disease in the case study (n=4,888).

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; CR-PO, continuation ratio logit model with proportional odds; CR-NP, continuation ratio logit model without proportional odds; SLM, stereotype logit model

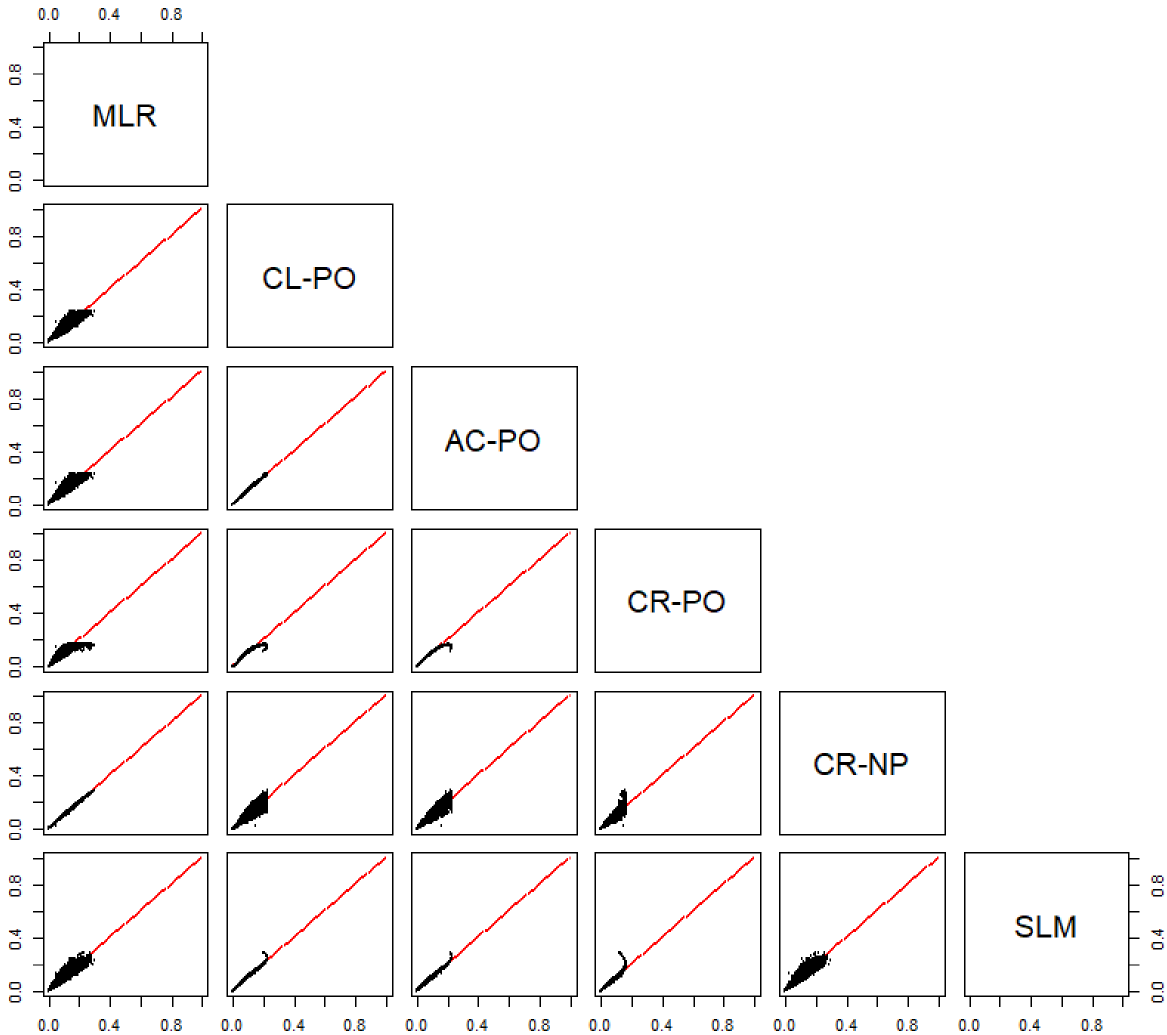


Figure S37. Scatter plot of estimated probabilities for having three-vessel disease in the case study (n=4,888).

MLR, multinomial logistic regression; CL-PO, cumulative logit model with proportional odds; AC-PO, adjacent category logit model with proportional odds; CR-PO, continuation ratio logit model with proportional odds; CR-NP, continuation ratio logit model without proportional odds; SLM, stereotype logit model

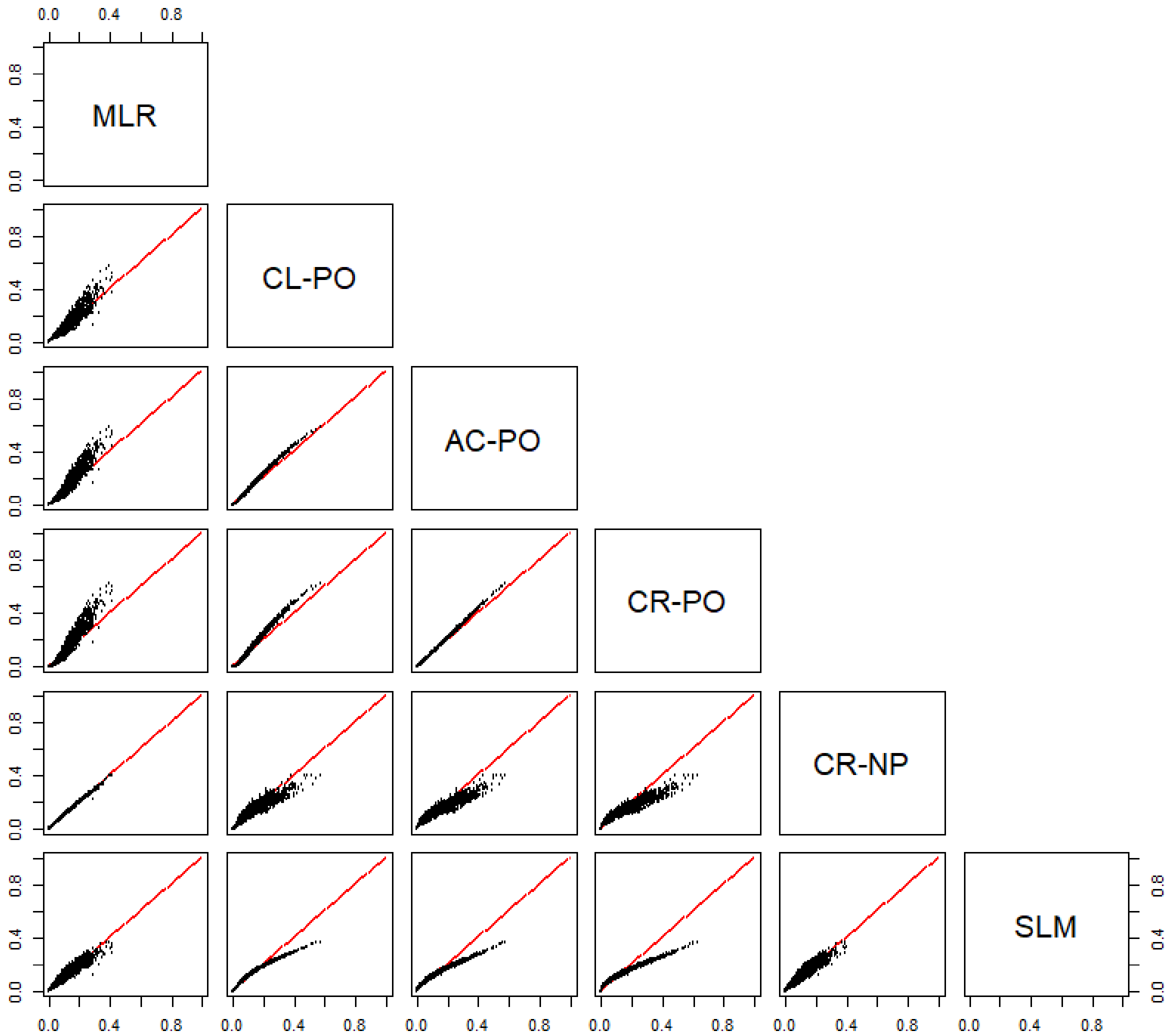


Figure S38. Calibration plots for the MLR model in the case study. These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves. The top left plot shows the flexible calibration scatter plot per outcome category, the top right plot the flexible calibration curves per outcome category, the bottom left plot the flexible calibration scatter plot per outcome dichotomy, and the bottom right the flexible calibration curves per outcome dichotomy.

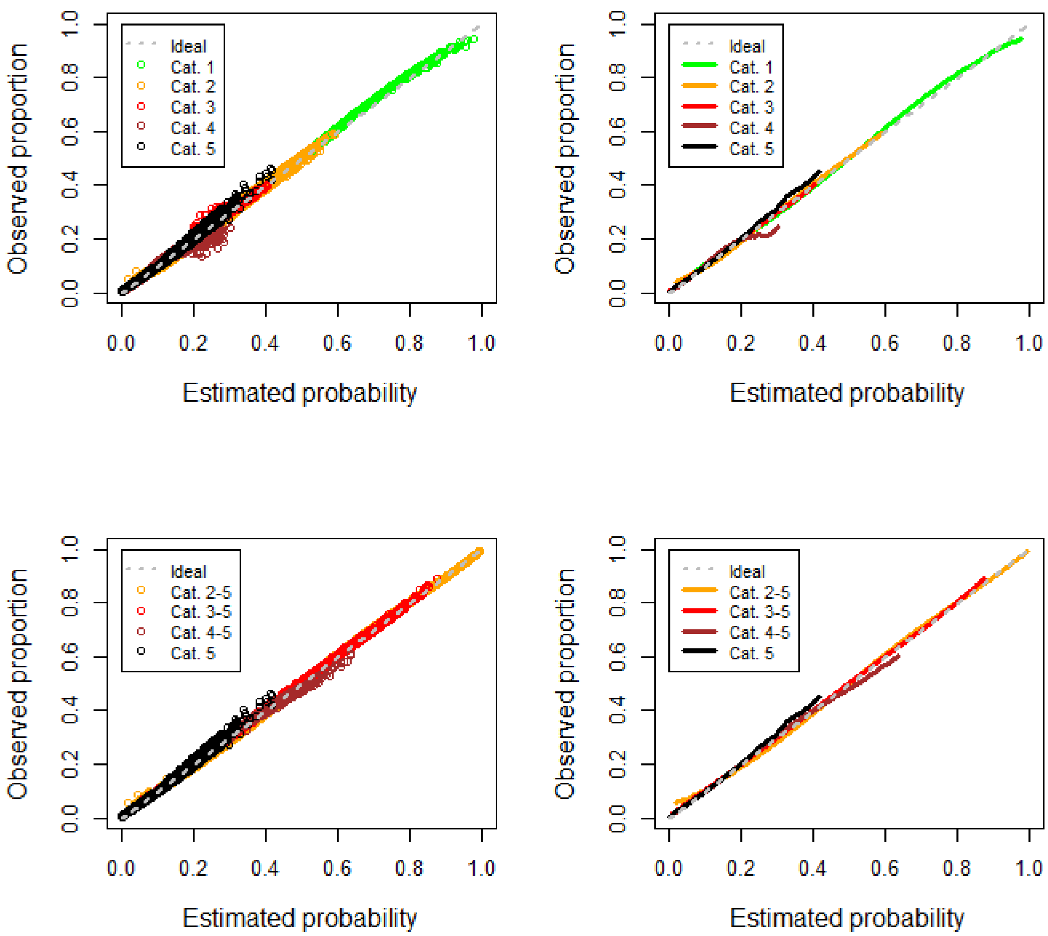


Figure S39. Calibration plots for the CL-PO model in the case study. These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves. The top left plot shows the flexible calibration scatter plot per outcome category, the top right plot the flexible calibration curves per outcome category, the bottom left plot the flexible calibration scatter plot per outcome dichotomy, and the bottom right the flexible calibration curves per outcome dichotomy.

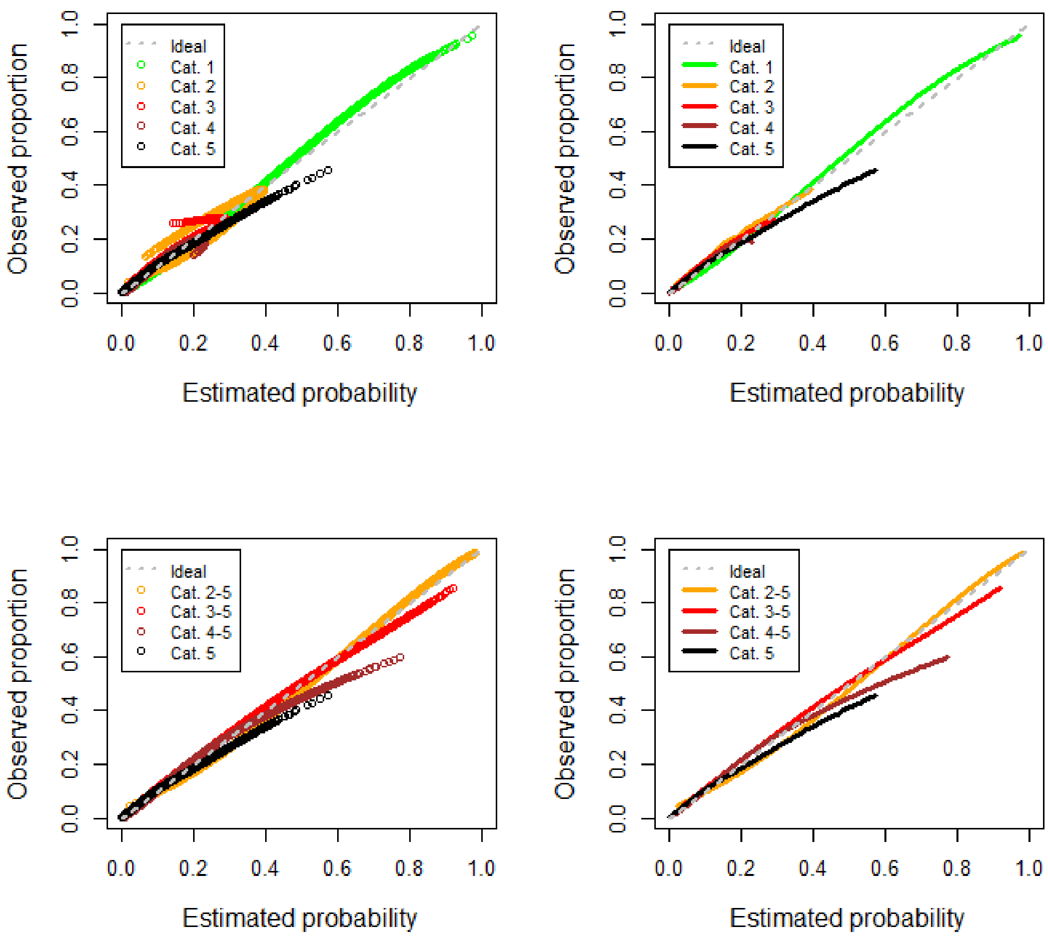


Figure S40. Calibration plots for the AC-PO model in the case study. These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves. The top left plot shows the flexible calibration scatter plot per outcome category, the top right plot the flexible calibration curves per outcome category, the bottom left plot the flexible calibration scatter plot per outcome dichotomy, and the bottom right the flexible calibration curves per outcome dichotomy.

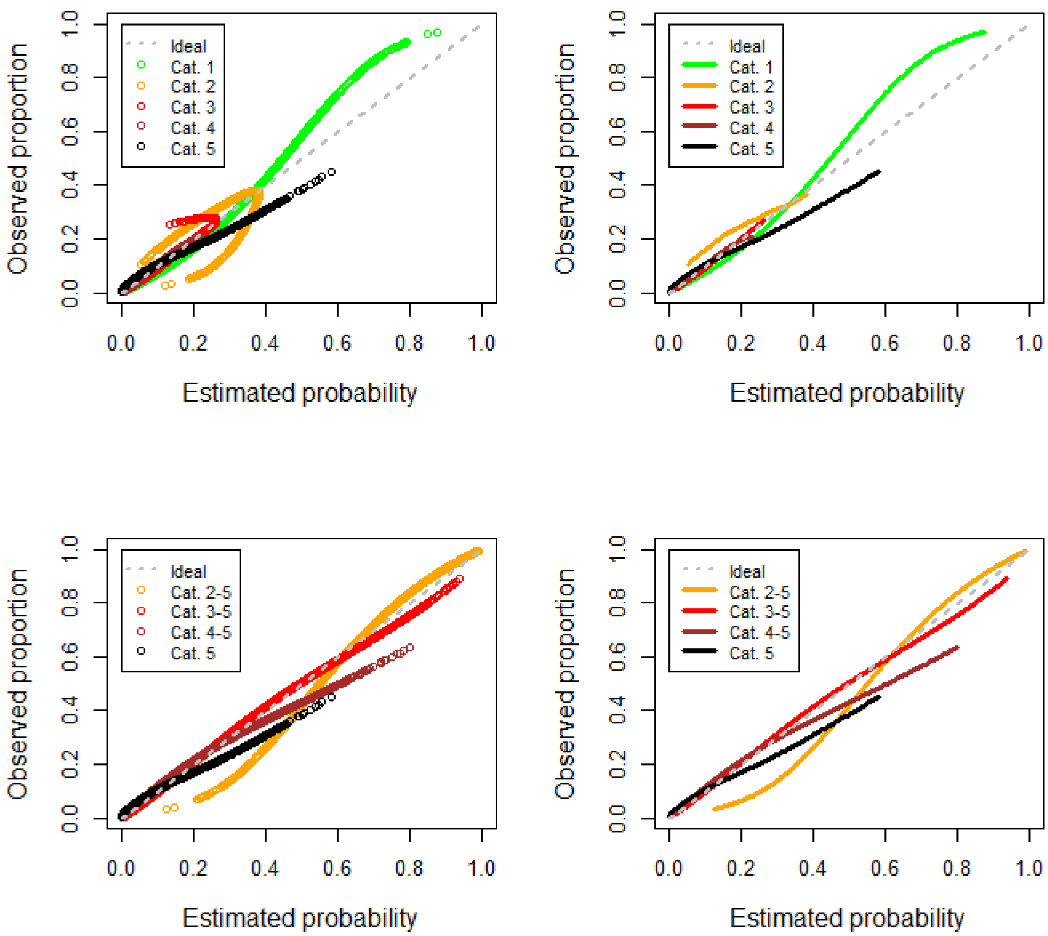


Figure S41. Calibration plots for the CR-PO model in the case study. These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves. The top left plot shows the flexible calibration scatter plot per outcome category, the top right plot the flexible calibration curves per outcome category, the bottom left plot the flexible calibration scatter plot per outcome dichotomy, and the bottom right the flexible calibration curves per outcome dichotomy.

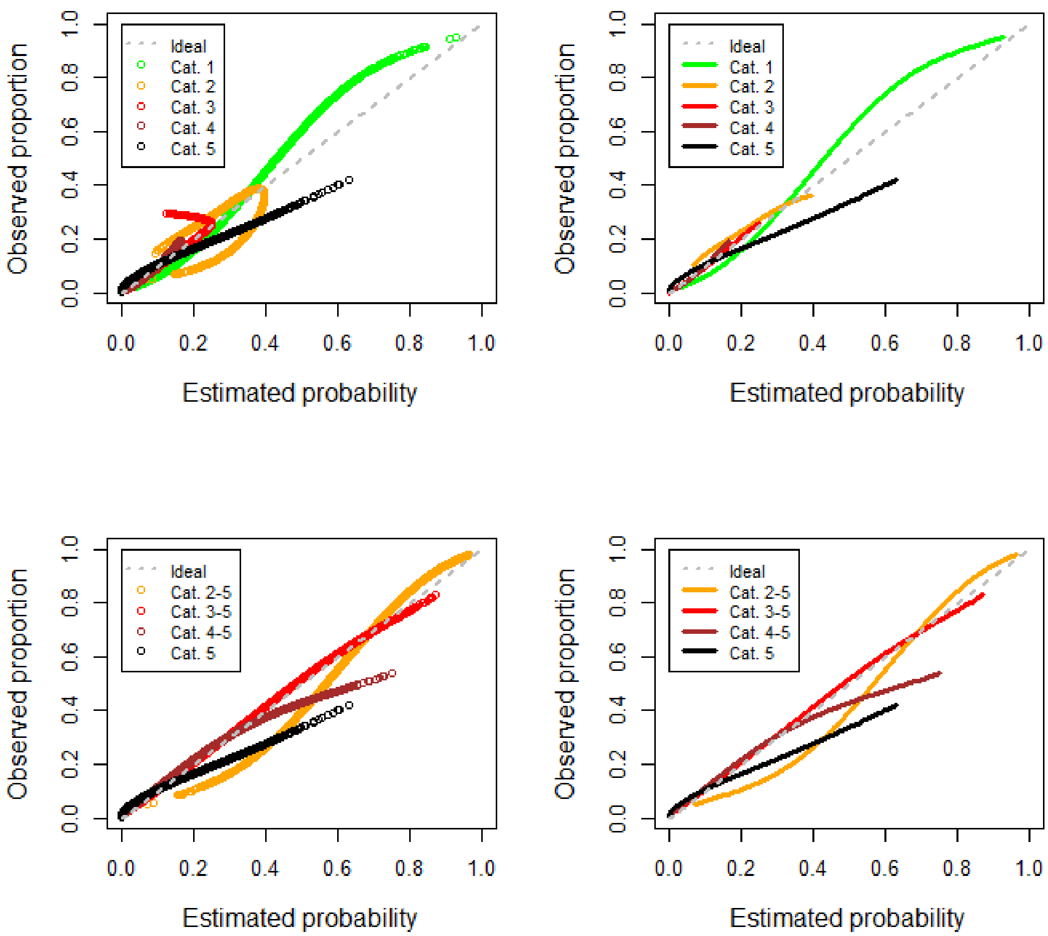


Figure S42. Calibration plots for the CL-NP model in the case study. These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves. The top left plot shows the flexible calibration scatter plot per outcome category, the top right plot the flexible calibration curves per outcome category, the bottom left plot the flexible calibration scatter plot per outcome dichotomy, and the bottom right the flexible calibration curves per outcome dichotomy.

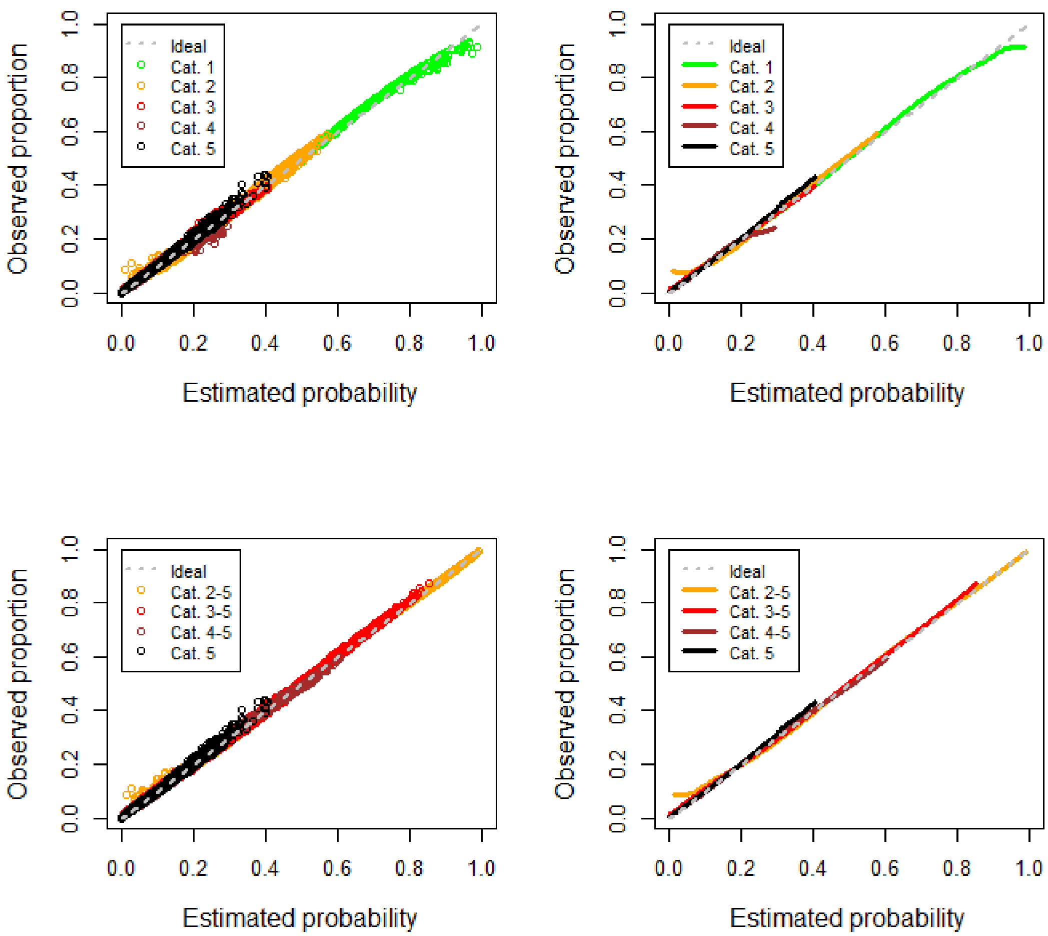


Figure S43. Calibration plots for the SLM model in the case study. These curves are based on the dataset used to develop the model and are therefore apparent (or unvalidated) curves. The top left plot shows the flexible calibration scatter plot per outcome category, the top right plot the flexible calibration curves per outcome category, the bottom left plot the flexible calibration scatter plot per outcome dichotomy, and the bottom right the flexible calibration curves per outcome dichotomy.