THE LANCET
Diabetes & Endocrinology

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Basu S, Sussman JB, Berkowitz SA, Hayward RA, Yudkin JS. Development and validation of Risk Equations for Complications Of type 2 Diabetes (RECODe) using individual participant data from randomised trials. *Lancet Diabetes Endocrinol* 2017; published online Aug 10. http://dx.doi.org/10.1016/S2213-8587(17)30221-8.

Web Extra Material

Statistical code.

Statistical code (in *R*) for replication and extension of the analyses conducted in this study is available at: https://github.com/sanjaybasu/t2dmriskeqns

Elastic net regularization.

Elastic net regularization seeks to select predictors from a set of candidate variables with the goal of generating parsimonious models by minimizing overfitting, while preserving a high degree of predictive power, assessed through repeated internal cross-validation.[39,40] Elastic net regularization utilizes a combination of ridge regression and lasso regression. Ridge regression shrinks coefficients of correlated predictors towards each other, while lasso regression penalizes nonzero regression coefficients, choosing one correlated predictor and discarding the others. Elastic net mixes ridge and lasso regression by adjusting the penalty parameter to balance the two methods.

Given a standard Cox hazard model of the form:

[1] $$h_i(t) = h_0(t)e^{x_i^T\beta},$$

where $h_i(t)$ is the hazard for patient i at time t, $h_0(t)$ is a shared baseline hazard rate, and β is a vector of model coefficients for a vector of predictors x_i, the elastic net regularization approach penalizes the negative log of the partial likelihood function for the Cox model, given by:

[2] $$L(\beta) = \prod_{i=1}^{m} \frac{e^{x_{j(i)}^T\beta}}{\sum_{j\in R_i} e^{x_j^T\beta}}$$

where R is the set of indices j for failure among participants at risk at time t_i across all possible times 1 to m. The objective is to find β that maximizes $L(\beta)$, while also minimizing over-fitting. This is equivalent to maximizing a scaled log partial likelihood, which in a scaled Lagranian reformulation derived previously,[39] producing the objective function:

[3] $$\hat{\beta} = argmax_\beta \left[\frac{2}{n}\left(\sum_{i=1}^{m} x_{j(i)}^T - log\left(\sum_{j\in R_i} e^{x_j^T\beta}\right)\right) - \lambda P_\alpha(\beta)\right]$$

where $\lambda P_\alpha(\beta)$ is the elastic net penalty:

[4] $$\lambda P_\alpha(\beta) = \lambda\left(\alpha\sum_{i=1}^{p}|\beta_i| + \frac{1}{2}(1-\alpha)\sum_{i=1}^{p}\beta_i^2\right)$$

As shown in equation 4, the parameter α balances between lasso regression (α=1, which tends to pick one correlated parameter and discard the others) and ridge regression (α=0, which tends to shrink correlated predictor coefficients toward each other). The parameter λ controls the overall degree of the penalty and selects the degree to which the model will be more or less parsimonious (i.e., higher λ values select fewer candidate parameters to produce a less over-fit model).

The method was implemented using the glmnet package in the statistical program *R*. The glmnet algorithm uses cyclical coordinate descent, which successively optimizes the objective function over each parameter with others fixed, and cycles repeatedly until convergence.[41]

The parameter λ value that minimized the partial likelihood deviance of the model (the error between the model and observed outcomes) over the course of 10-times repeated internal cross-validation in the ACCORD sample was chosen for each Cox model. In 10-fold internal cross-validation, the original sample is randomly partitioned into 10 equal sized subsamples. Of the 10 subsamples, a single subsample is retained as the validation data for testing the model, and the remaining 9subsamples are used as training data. The cross-validation process is then repeated 10 times, with each of the 10 subsamples used exactly once as the validation data. The 10 results are then averaged to provide a coefficient estimate. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once. The choice of ten cross-validations, or any number of such validations, is inherently arbitrary and ten is conventional, but we also checked that the choice of ten validations was produced stable estimates of covariates selected and coefficient values.

Additional details on derivation and validation datasets

Derivation data.

ACCORD included a randomized, controlled trial of intensive versus standard glycemic control (open-label target of hemoglobin A1c <6.0% versus 7.0-7.9%, respectively), with a multi-factorial design in which participants in the glycemic control trial were additionally randomized to intensive versus standard lipid treatment (double-blinded assignment to fibrate plus statin or placebo plus statin, respectively), or intensive versus standard blood pressure treatment (open-label target of systolic blood pressure <120 mmHg or <140 mmHg, respectively). The study was conducted in the United States and Canada between January 2001 and June 2009. The study was terminated early due to higher mortality in the intensive compared to standard glycemic treatment group. Participants were 40 to 79 years old with type 2 diabetes and a hemoglobin A1c level of at least 7.5%, and either prior evidence of cardiovascular disease (CVD) or risk factors for CVD (dyslipidemia, hypertension, smoking or obesity, main text **Table 1**). Exclusion criteria for ACCORD included having a body mass index greater than 45 kg/m^2, serum creatinine greater than 1.5 mg/dL, or serious illnesses that might limit trial participation or life expectancy.

Validation data.

The Diabetes Prevention Program Outcomes Study (DPPOS) was used for validation of the microvascular risk equations. DPPOS was the follow-up to the randomized, controlled Diabetes Prevention Program (DPP) trial of metformin, troglitazone, lifestyle or placebo (with double-blinding of medication arms). DPP was conducted in the United States between January 1996 and July 2001, followed by the

DPPOS follow-up through October 2013. All DPP completers were offered DPPOS enrolment. DPP participants in the metformin arm continued open-label metformin in DPPOS until diabetes diagnosis, at which point study medication was deferred to the patient's primary care provider but participation in DPPOS continued. DPP participants in the lifestyle and placebo arms were offered group lifestyle classes in DPPOS. DPP/DPPOS inclusion criteria included age at least 25 years old, body mass index of $\geq$24 kg/m^2 for non-Asians or $\geq$22 kg/m^2 for Asians, a plasma glucose concentration of 95 to 125 mg/dL fasting, and glucose of 140 to 199 mg/dL two hours after a 75-g oral glucose load (main text **Table 1**). DPP/DPPOS excluded persons taking medications known to alter glucose tolerance or persons having serious illnesses that might limit trial participation or life expectancy. Trial data from individual participants who developed type 2 diabetes at any point during the DPP/DPPOS, across all study arms except the troglitazone arm (which was canceled early in the DPP due to liver toxicity), were included for microvascular equation validation (main text **Table 1**).

The Action for Health in Diabetes trial (Look AHEAD), used for validation of the macrovascular risk equations, was a randomized, controlled trial of intensive lifestyle modification versus diabetes support and education. Look AHEAD was conducted in the United States between January 2001 and September 2012. Look AHEAD inclusion criteria included age 45 to 75 years old; having type 2 diabetes with a haemoglobin A1c of $\leq$11%, a blood pressure of <160/100 mmHg, and triglycerides <600 mg/dL; and having a body mass index of at least 25 kg/m^2 (main text **Table 1**). Look AHEAD excluded persons unable to complete a maximal exercise test, those not having a primary care provider, and those having serious illnesses that might limit trial participation or life expectancy. Trial data from individual participants, across all study arms, were included for macrovascular equation validation (main text **Table 1**).

Web Extra Material Table 1. Model performance if including persons with missing data, using multiple imputation with chained equations for missing covariates.

	Internal validation: RECODe		**External validation: RECODe**		**Alternative risk equations: UKPDS OM2**		**Alternative risk equations: ACC/AHA PCE's**	
Equation	Discrimination: C-statistic	Calibration: Slope/intercept/χ^2, P-value*	Discrimination: C-statistic	Calibration: Slope/intercept/χ^2, P-value*	Discrimination: C-statistic	Calibration: Slope/intercept/χ^2, P-value*	Discrimination: C-statistic	Calibration: Slope/intercept/χ^2, P-value*
Microvascular outcomes	ACCORD study (*N* = 9,635, 2001-2009)		DPPOS study (*N* = 1,018, 1996-2001)		ACCORD study (*N* = 9,635, 2001-2009)			
Nephropathy								
(i) Microalbuminuria	0.62 (0.61, 0.64)	0.94/0.015/5.7, 0.77						
(ii) Macro-albuminuria	0.84 (0.82, 0.86)	1.14/-0.009/79.4*, <0.0001						
(iii) Renal failure/ESRD	0.60 (0.56, 0.64)	1.28/0.0003/30.8*, <0.0001			0.54 (0.50, 0.59)	0.19/0.035/242.6*, <0.0001		
(iv) 2x SCr or >20 mL/min decr in eGFR	0.76 (0.75, 0.77)	0.91/0.053/42.9*, <0.0001						
(v) Any of ii-iv	0.74 (0.73, 0.75)	0.88/0.073/64.3*, <0.0001						
(vi) Any of i-iii	0.61 (0.60, 0.63)	0.96/0.011/4.6, 0.87	0.65 (0.61, 0.70)	1.31/-0.15/9.3, 0.16				
Retinopathy								
(i) photocoagulation or vitrectomy	0.65 (0.63, 0.67)	1.03/-0.003/15.7, 0.07	0.57 (0.51, 0.63)	0.72/0.12/13.9, 0.05				
(ii) cataract extraction	0.68 (0.66, 0.69)	0.97/0.004/18.7*, 0.03						
(iii) 3-line change in acuity	0.56 (0.55, 0.57)	0.87/0.052/11.1, 0.27						
(iv) severe vision loss	0.62 (0.60, 0.64)	1.01/-0.001/6.9, 0.65			0.59 (0.57, 0.62)	1.12/0.041/59.0*, <0.0001		
(v) any of i or iv	0.63 (0.62, 0.65)	0.97/0.004/11.5, 0.24						
Neuropathy								
(i) MNSI >2	0.60 (0.59, 0.62)	1.01/-0.005/14.4, 0.11						
(ii) vibratory sensation loss	0.64 (0.63, 0.66)	0.99/0.003/17.2, 0.05						
(iii) ankle jerk loss	0.57 (0.55, 0.58)	0.96/0.019/5.0, 0.84						
(iv) pressure sensation loss	0.62 (0.61, 0.64)	1.00/-0.0005/9.7, 0.37	0.69 (0.63, 0.74)	1.01/-0.002/1.0, 0.91				

Macrovascular outcomes	ACCORD study (*N* = 9,635, 2001-2009)		Look AHEAD study (*N* = 4,760, 2001-2012)		ACCORD study (*N* = 9,635, 2001-2009) and Look AHEAD study (*N* = 4,760, 2001-2012)		ACCORD study (*N* = 9,635, 2001-2009) and Look AHEAD study (*N* = 4,760, 2001-2012)	
(i) ASCVD	0.69 (0.67, 0.71)	1.06/-0.005/13.7, 0.14	0.73 (0.71, 0.75)	1.13/-0.071/203.1*, <0.0001	0.62 (0.60, 0.63) in ACCORD, 0.67 (0.64, 0.69) in Look AHEAD	0.36/0.043/602.6*, <0.0001 in ACCORD, 0.53/-0.013/746.6*, <0.0001 in Look AHEAD	0.61 (0.59, 0.63) in ACCORD, 0.66 (0.64, 0.69) in Look AHEAD	0.30/0.077/468.8*, <0.0001 in ACCORD, 0.39/0.032/444.0*, <0.0001 in Look AHEAD
(ii) MI (fatal/nonfatal)	0.69 (0.67, 0.70)	1.00/0.0003/6.4, 0.70	0.71 (0.68, 0.74)	1.08/-0.016/17.0, 0.05	0.62 (0.59, 0.64) in ACCORD, 0.67 (0.65, 0.70) in Look AHEAD	0.80/0.106/47.6*, <0.0001 in ACCORD, 0.94/-0.038/270.9*, <0.0001 in Look AHEAD		
(iii) Stroke (fatal/nonfatal)	0.70 (0.66, 0.74)	1.16/-0.003/7.4, 0.38	0.67 (0.63, 0.71)	0.99/0.006/8.2, 0.22	0.61 (0.56, 0.66) in ACCORD, 0.63 (0.58, 0.68) in Look AHEAD	0.063/0.023/2275.6*, <0.0001 in ACCORD, 0.279/0.007/659.5*, <0.0001 in Look AHEAD		
(iv) CHF	0.75 (0.73, 0.77)	1.01/-0.0004/3.1, 0.93	0.76 (0.73, 0.80)	1.13/-0.011/11.7, 0.07	0.61 (0.58, 0.65) in ACCORD, 0.61 (0.57, 0.65) in Look AHEAD	0.46/0.006/345.8*, <0.0001 in ACCORD, 0.24/0.010/1246.5*, <0.0001 in Look AHEAD		
(v) CVD death	0.74 (0.71, 0.77)	0.96/0.001/7.8, 0.46	0.79 (0.74, 0.83)	1.00/-0.010/44.5*, <0.0001				
All-cause mortality	0.70 (0.68, 0.72)	1.03/-0.002/14.7, 0.10	0.71 (0.68, 0.74)	1.10/-0.012/16.3, 0.06				

See Methods for definitions of microvascular outcomes. 95% confidence intervals in parentheses around C-statistics are from 10-times cross-validations. Calibration slopes/intercepts are calculated between deciles of predicted and observed Kaplan-Meier event rates, with lower numbers of centiles than deciles used if <5 events are observed per group, to prevent unstable inferences per current guidelines.

Legend: MNSI = Michigan Neuropathy Screening Instrument; CVD = cardiovascular disease; ASCVD = atherosclerotic cardiovascular disease (nonfatal or fatal myocardial infarction or stroke); MI = myocardial infarction; CHF = congestive heart failure

*P-values <0.05 reflect larger difference between predicted and observed Kaplan-Meier event rates by the Greenwood-D'Agostino-Nam test (see Figure 1 for calibration plots).

Web Extra Material Table 2. Reclassification numbers and net reclassification indices (NRI) by outcome, for the derivation cohort, ACCORD (*N* = 9,635, 2001-2009).

Risk predictions from new RECODe equations compared to predictions of older risk equations from the UK Prospective Diabetes Study Outcomes Model 2 (UKPDS OM2) and the American College of Cardiology/American Heart Association Pooled Cohort Equations (ACC/AHA PCEs). The tables show how many people with and without observed events were correctly and incorrectly classified as high or low risk by the older risk equations and by the newer RECODe equations. Positive NRI values indicate overall improvement from the older to the newer RECODe equations.

(A)

Nephropathy: renal failure/ESRD

People without any events

	New RECODe equations		
	<=10%	>10%	Total
UKPDS OM2			
<=10%	7,193	9	7,202
>10%	1456	70	1,526
Total	8,649	79	8,728

People with events

	<=10%	>10%	Total
UKPDS OM2			
<=10%	194	0	194
>10%	54	19	73
Total	248	19	267
		NRI	0.161875188

(B)

Retinopathy: severe vision loss

People without any events

	New RECODe equations		
UKPDS OM2	<=10%	>10%	Total
<=10%	6,459	577	7,036
>10%	641	583	1,224
Total	7,100	1,160	8,260

People with events

UKPDS OM2	<=10%	>10%	Total
<=10%	459	100	559
>10%	73	103	176
Total	532	203	735
		NRI	0.027712311

(C)

ASCVD

New RECODe equations

People without any events

UKPDS OM2	<=10%	>10%	Total
<=10%	461	7	468
>10%	6289	1,368	7,657
Total	6,750	1,375	8,125

People with events

UKPDS OM2	<=10%	>10%	Total
<=10%	26	2	28
>10%	496	346	842
Total	522	348	870

NRI	12.6071981

(D)

MI

People without any events

UKPDS OM2	New RECODe equations <=5%	>5%	Total
<=5%	63	143	206
>5%	330	7,726	8,056
Total	393	7,869	8,262

People with events

UKPDS OM2	<=5%	>5%	Total
<=5%	0	6	6
>5%	6	721	727
Total	6	727	733

NRI	0.882075472

(E)

Stroke

New RECODe equations

People without any events

		<=5%	>5%	Total
UKPDS OM2	<=5%	1,288	11	1,299
	>5%	7082	457	7,539
	Total	8,370	468	8,838

People with events

		<=5%	>5%	Total
UKPDS OM2	<=5%	16	1	17
	>5%	113	27	140
	Total	129	28	157

NRI	5.358515071

(F)

CHF

New RECODe equations

People without any events

		<=10%	>10%	Total
UKPDS OM2	<=10%	1,225	219	1,444
	>10%	5777	1,372	7,149
	Total	7,002	1,591	8,593

People with events

		<=10%	>10%	Total
UKPDS OM2	<=10%	28	27	55
	>10%	164	183	347
	Total	192	210	402

	NRI	3.689528789

(G)

ACC/AHA ASCVD

New RECODe equations

People without any events

	<=10%	>10%	Total
<=10%	1,015	100	1,115
>10%	4191	2,819	7,010
Total	5,206	2,919	8,125

ACC/AHA PCEs

People with events

	<=10%	>10%	Total
<=10%	30	20	50
>10%	292	528	820
Total	322	548	870

ACC/AHA PCEs

NRI	3.476849796

Web Extra Material Table 3. Reclassification numbers and net reclassification indices by outcome, for the validation cohort for which both UKPDS and newer equations were available for the same outcomes (Look AHEAD study, *N* = 4,760, 2001-2012). Risk predictions from new equations derived from the ACCORD study sample (*N* = 9,635, 2001-2009) compared to predictions of older risk equations from the UK Prospective Diabetes Study Outcomes Model 2 (UKPDS OM2) and the American College of Cardiology/American Heart Association Pooled Cohort Equations (ACC/AHA PCEs). The tables show how many people with and without observed events were correctly and incorrectly classified as high or low risk by the older risk equations and by the newer risk equations.

(A)

ASCVD

New RECODe equations

People without any events

UKPDS OM2	<=10%	>10%	Total
<=10%	282	7	289
>10%	2682	1,307	3,989
Total	2,964	1,314	4,278

People with events

UKPDS OM2	<=10%	>10%	Total
<=10%	6	2	8
>10%	162	291	453
Total	168	293	461

NRI 8.970714196

(B)

MI

New RECODe equations

People without any events

UKPDS OM2	<=5%	>5%	Total
<=5%	1,640	131	1,771
>5%	1881	756	2,637
Total	3,521	887	4,408

People with events

UKPDS OM2	<=5%	>5%	Total
<=5%	37	10	47
>5%	122	162	284
Total	159	172	331

NRI	0.924253226

(C)

Stroke

New RECODe equations

People without any events

		<=5%	>5%	Total
UKPDS OM2	<=5%	2,845	10	2,855
	>5%	1627	100	1,727
	Total	4,472	110	4,582

People with events

		<=5%	>5%	Total
UKPDS OM2	<=5%	64	0	64
	>5%	86	7	93
	Total	150	7	157

NRI	0.506704087

(D)

CHF

New RECODe equations

People without any events

UKPDS OM2	<=10%	>10%	Total
<=10%	1,351	77	1,428
>10%	2756	345	3,101
Total	4,107	422	4,529

People with events

UKPDS OM2	<=10%	>10%	Total
<=10%	31	10	41
>10%	87	82	169
Total	118	92	210

NRI	1.800142184

(E)

ACC/AHA ASCVD — New RECODe equations

People without any events

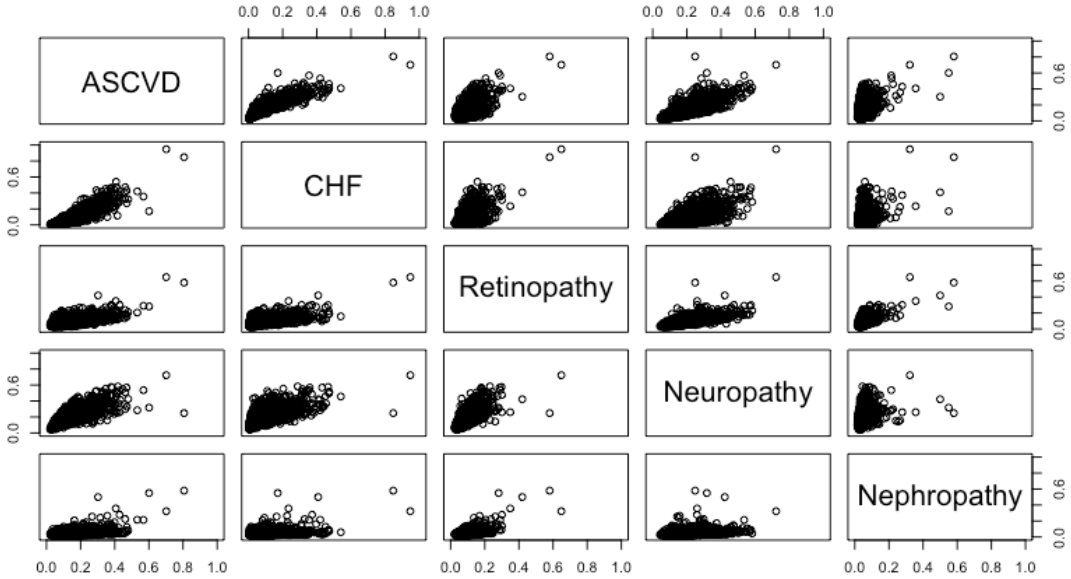
	<=10%	>10%	Total
ACC/AHA PCEs <=10%	1,641	143	1,784
ACC/AHA PCEs >10%	1323	1,171	2,494
Total	2,964	1,314	4,278

People with events

	<=10%	>10%	Total
ACC/AHA PCEs <=10%	64	20	84
ACC/AHA PCEs >10%	104	273	377
Total	168	293	461

NRI	0.602433551

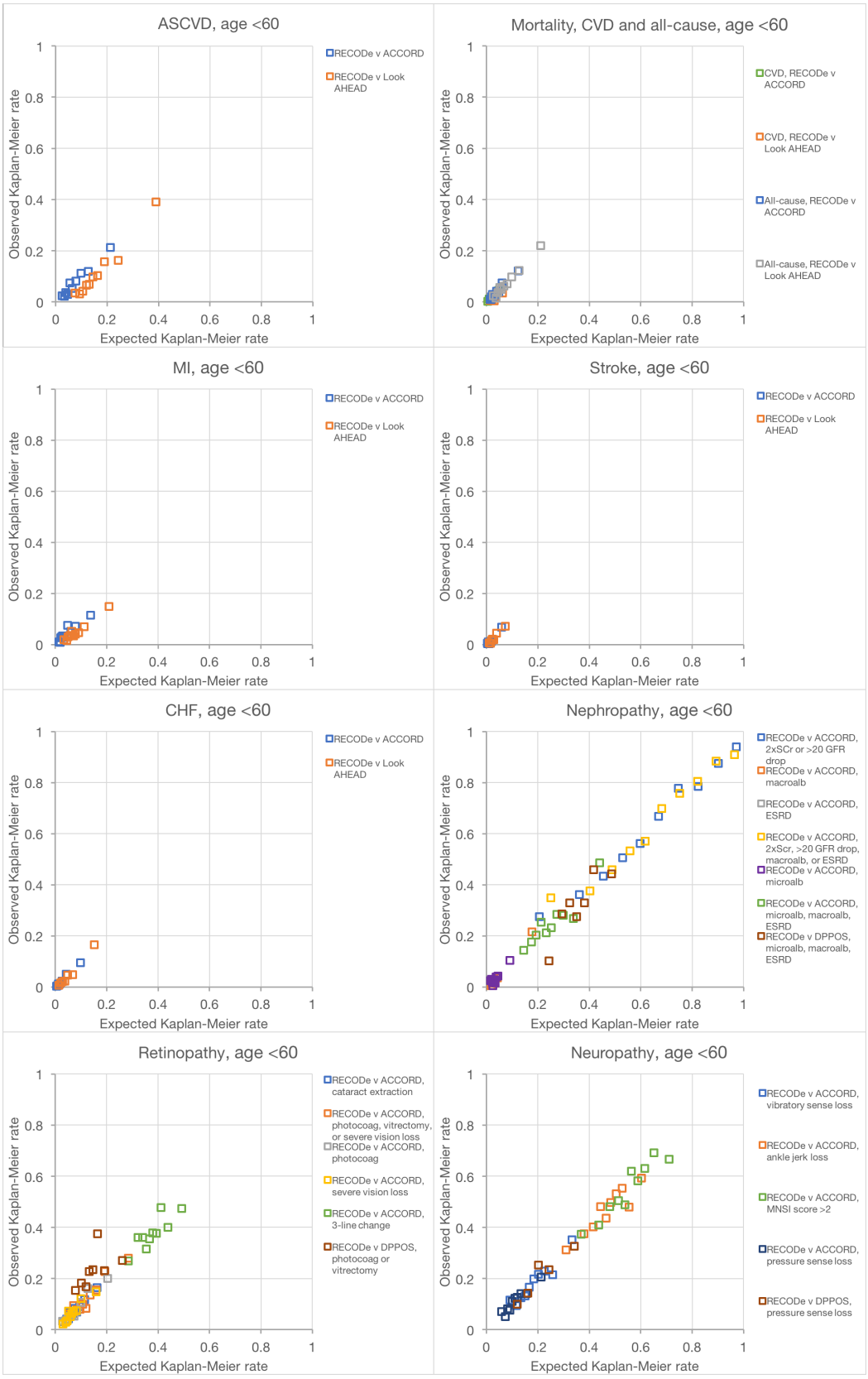
Web Extra Material Figure 1. Correlelogram of predicted risks by RECODe equations.
Correlations between predicted macrovascular and microvascular risk for participants in the ACCORD study (*N* = 9,635, 2001-2009). Nephropathy was defined as renal failure/ESRD; neuropathy as pressure sensation loss; and retinopathy as severe vision loss (<20/200 visual acuity by Snellen chart). Axes are all arranged from probability 0 to probability 1 of the outcome over a predicted 10-year time horizon.



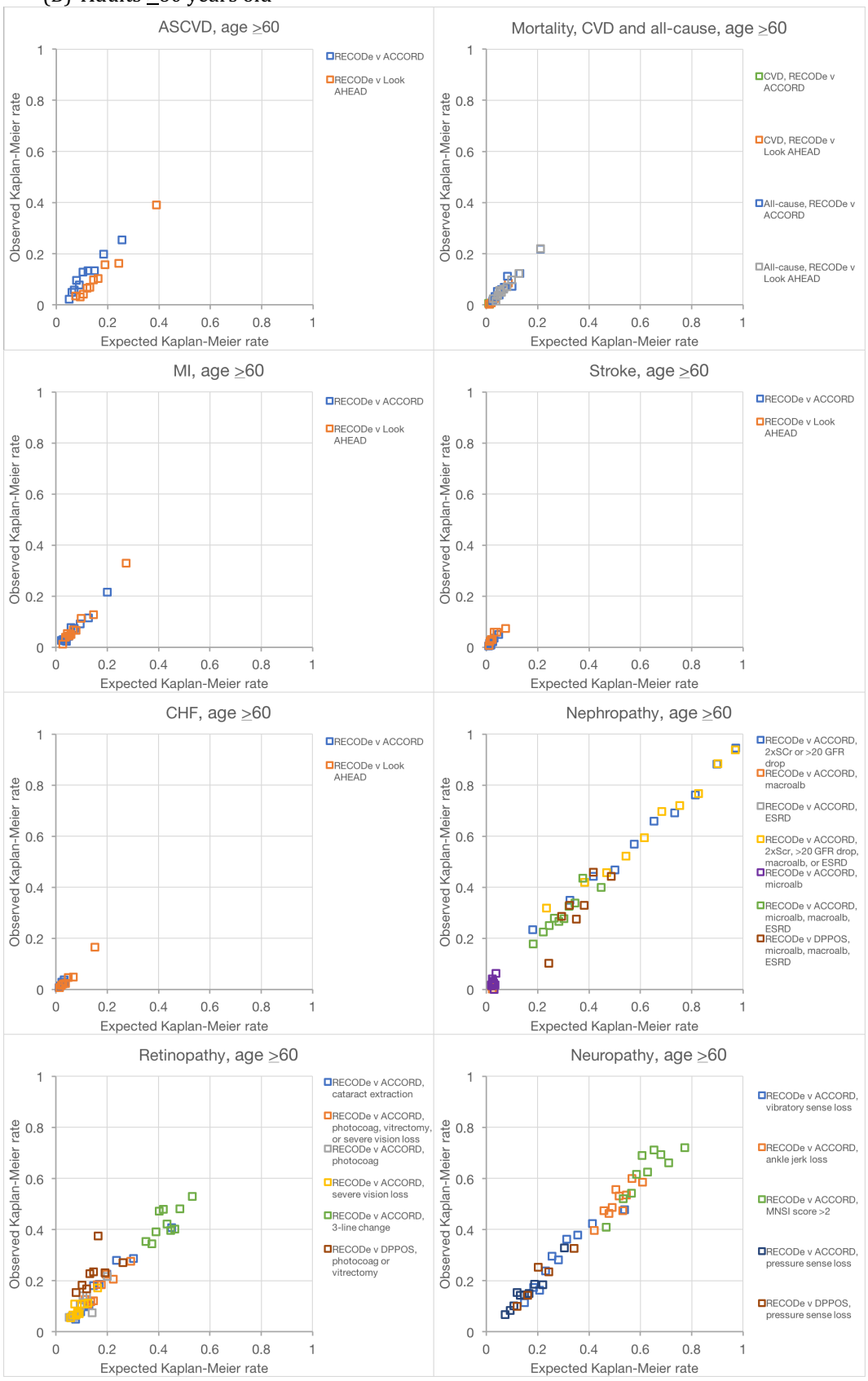
Web Extra Material Figure 2. Calibration plots among subgroups.

Plots display Kaplan-Meier event rates over 10 years predicted by the RECODe equations versus observed rates in the ACCORD study (*N* = 9,635, 2001-2009), DPPOS study (*N* = 1,018, 1996-2001), and Look AHEAD study (*N* = 4,760, 2001-2012). Also displayed, where available for each outcome, are predictions from older equations from the United Kingdom Prospective Diabetes Study Outcomes Model 2 (UKPDS OM2) and the American College of Cardiology/American Heart Association Pooled Cohort Equations (ACC/AHA PCEs) for atherosclerotic cardiovascular disease (nonfatal/fatal myocardial infarction or stroke). Points are displayed for deciles of predicted and observed Kaplan-Meier event rates, with lower numbers of centiles than deciles used if <5 events are observed per group, to prevent unstable inferences per current guidelines. Legend: MNSI = Michigan Neuropathy Screening Instrument; CVD = cardiovascular disease; ASCVD = atherosclerotic cardiovascular disease (nonfatal or fatal myocardial infarction or stroke); MI = myocardial infarction; CHF = congestive heart failure.

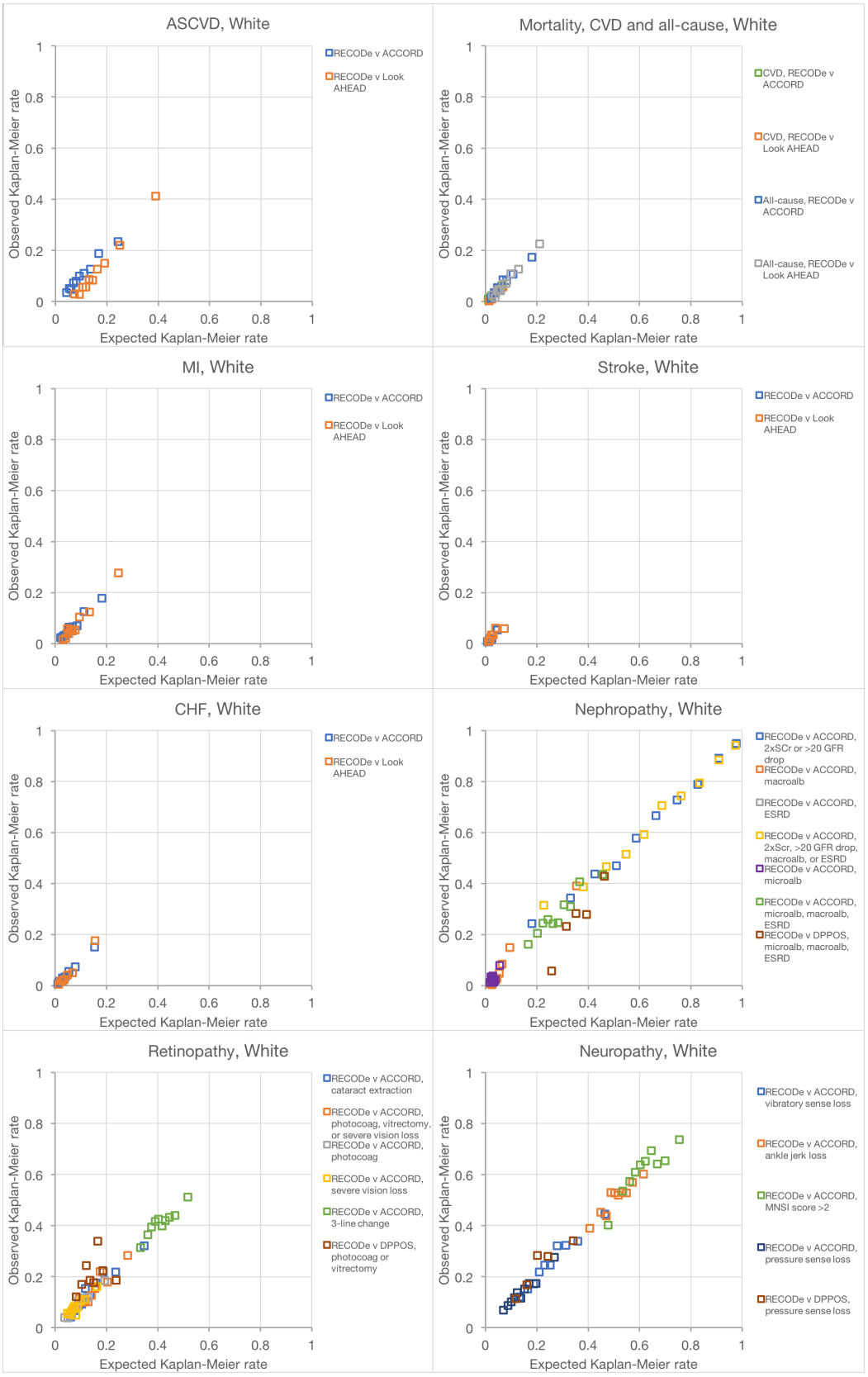
(A) Adults <60 years old



(B) Adults ≥60 years old



(C) White race


ASCVD, White
Observed Kaplan-Meier rate
Expected Kaplan-Meier rate
RECODe v ACCORD
RECODe v Look AHEAD
Mortality, CVD and all-cause, White
CVD, RECODe v ACCORD
CVD, RECODe v Look AHEAD
All-cause, RECODe v ACCORD
All-cause, RECODe v Look AHEAD
MI, White
Stroke, White
CHF, White
Nephropathy, White
RECODe v ACCORD, 2xSCr or >20 GFR drop
RECODe v ACCORD, macroalb
RECODe v ACCORD, ESRD
RECODe v ACCORD, 2xScr, >20 GFR drop, macroalb, or ESRD
RECODe v ACCORD, microalb
RECODe v ACCORD, microalb, macroalb, ESRD
RECODe v DPPOS, microalb, macroalb, ESRD
Retinopathy, White
RECODe v ACCORD, cataract extraction
RECODe v ACCORD, photocoag, vitrectomy, or severe vision loss
RECODe v ACCORD, photocoag
RECODe v ACCORD, severe vision loss
RECODe v ACCORD, 3-line change
RECODe v DPPOS, photocoag or vitrectomy
Neuropathy, White
RECODe v ACCORD, vibratory sense loss
RECODe v ACCORD, ankle jerk loss
RECODe v ACCORD, MNSI score >2
RECODe v ACCORD, pressure sense loss
RECODe v DPPOS, pressure sense loss

(D) Non-White race