

Supplementary Online Content

Comparisons of Polyexposure, Polygenic, and Clinical Risk Scores Implicate Prominent Roles of Non-Genetic Factors in Risk Prediction and Screening of Type 2 Diabetes

Yixuan He^{1,2}, Chirag M Lakhani², Danielle Rasooly^{2,3}, Arjun K Manrai^{2,3}, Ioanna Tzoulaki⁴, Chirag J Patel²

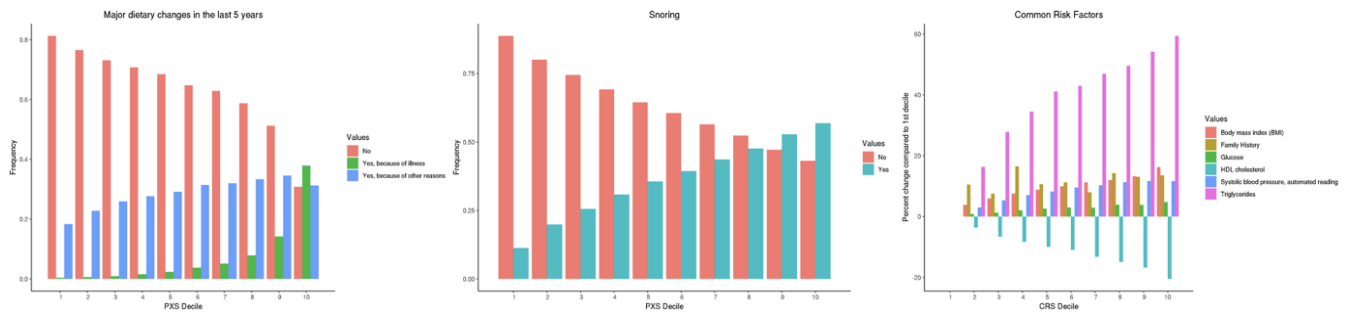
¹Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA, USA

²Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA.

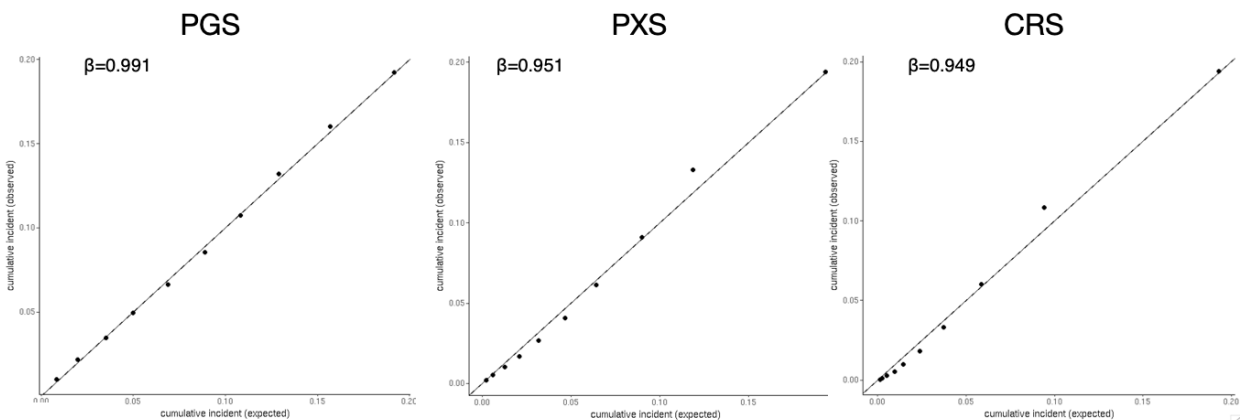
³Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA.

⁴Department of Epidemiology and Biostatistics, Imperial College London School of Public Health, London, United Kingdom

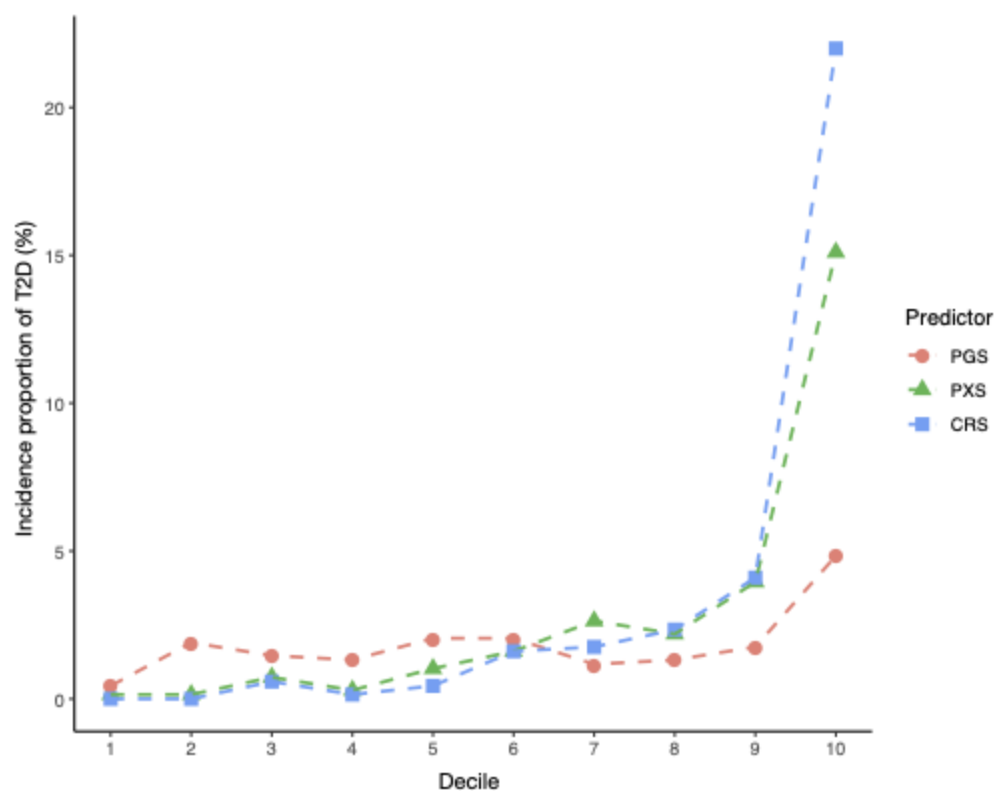
Supplementary Figure S1. Changes in variable responses over risk gradients. Left and middle: the frequency of each response was calculated within each PXS decile and plotted for the top two exposure variables. Right: The average value of each clinical risk variable was calculated for each decile, and the percent change compared to the first decile is plotted.



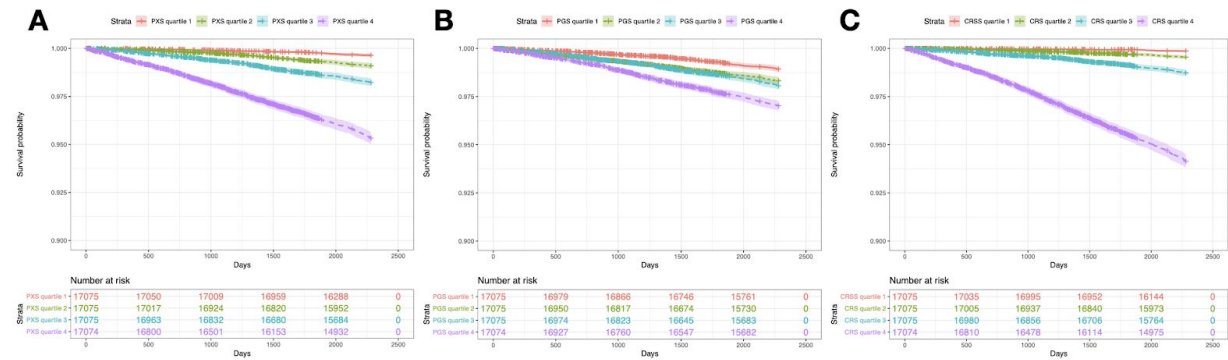
Supplementary Figure S2. Calibration plots for PGS, PXS, and CRS. The observed probability (Kaplan-Meier estimates) is plotted against the mean predicted probability within tenths of the predicted probabilities at t=2000 days.



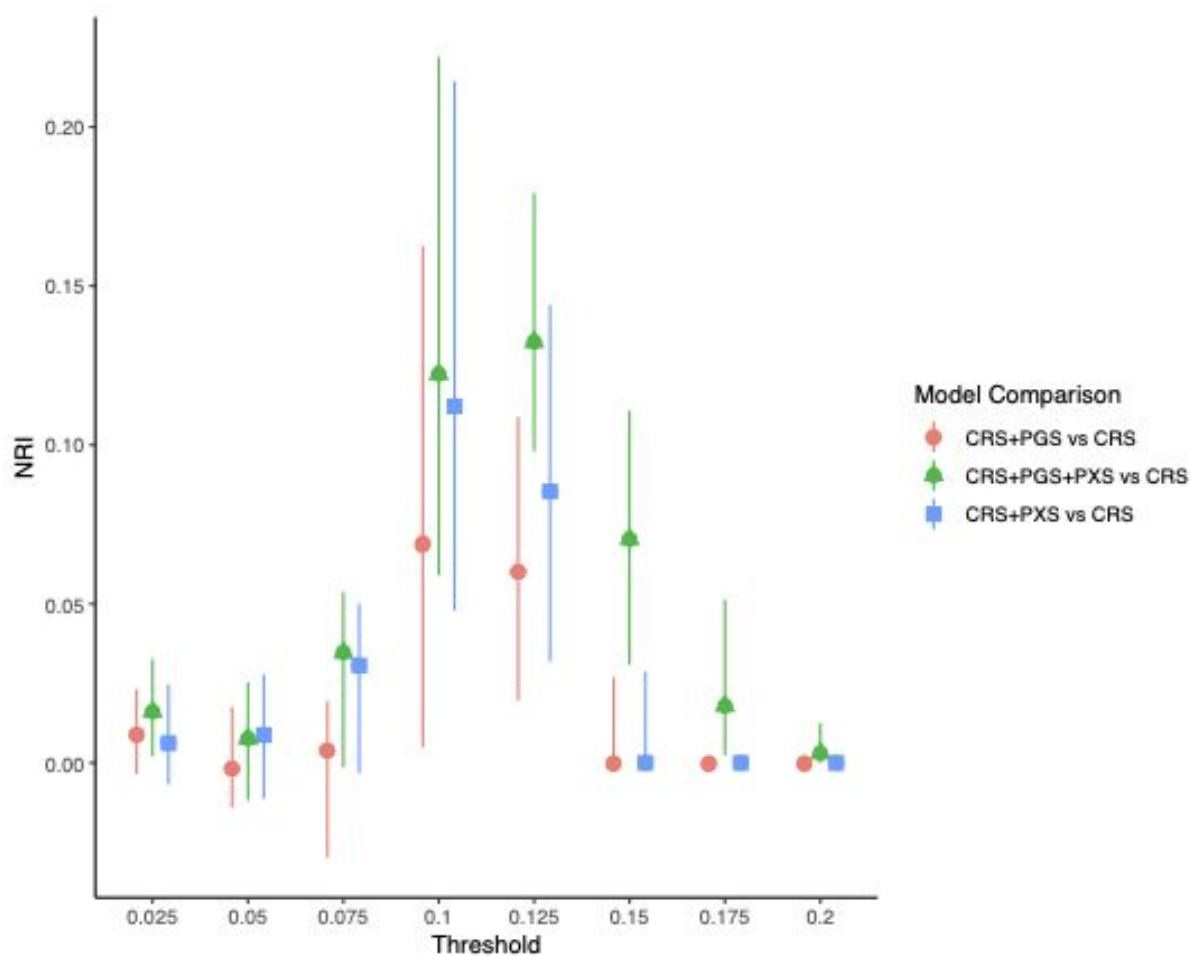
Supplementary Figure S3: Risk gradient for T2D. The incidence of T2D in 10 groups binned according to the decile of polygenic risk score (PGS), polyexposure risk score (PXS), or clinical risk score (CRS).



Supplementary Figure S4: Kaplan Meier (KM) plots of time-to-events for risk scores. The KM plots illustrate significant differences in time-to-events between different quartiles using A) polygenic risk score (PGS), B) polyexposure risk score (PXS), or C) clinical risk score (CRS).



Supplementary Figure S5: NRI of risk models over a range of risk thresholds. The categorical NRI for different risk models.



Supplementary Figure S6: Categorical NRI of CRS+PGS model over range of risk thresholds

	CRS Model	CRS+PGS Model		
		< 0.025	>= 0.025	Total
Events	< 0.025	259	33	292
	>= 0.025	35	954	989
	Total	294	987	1281
Nonevents	< 0.025	49197	1053	50250
	>= 0.025	1759	15009	16768
	Total	50956	16062	67018

	CRS Model	CRS+PGS Model		
		< 0.05	>= 0.05	Total
Events	< 0.05	482	41	523
	>= 0.05	57	701	758
	Total	539	742	1281
Nonevents	< 0.05	57208	1007	58215
	>= 0.05	1725	7078	8803
	Total	58933	8085	67018

	CRS Model	CRS+PGS Model		
		< 0.075	>= 0.075	Total
Events	< 0.075	688	103	791
	>= 0.075	96	394	490
	Total	784	497	1281
Nonevents	< 0.075	62145	1266	63411
	>= 0.075	1163	2444	3607
	Total	63308	3710	67018

	CRS Model	CRS+PGS Model		
		< 0.1	>= 0.1	Total
Events	< 0.1	966	165	1131
	>= 0.1	60	90	150
	Total	1026	255	1281
Nonevents	< 0.1	65383	1103	66486
	>= 0.1	246	286	532
	Total	65629	1389	67018

	CRS Model	CRS+PGS Model		
		< 0.125	>= 0.125	Total
Events	< 0.125	1198	83	1281
	>= 0.125	0	0	0
	Total	1198	83	1281
Nonevents	< 0.125	66703	315	67018
	>= 0.125	0	0	0
	Total	66703	315	67018

	CRS Model	CRS+PGS Model		
		< 0.15	>= 0.15	Total
Events	< 0.15	1281	0	1281
	>= 0.15	0	0	0
	Total	1281	0	1281
Nonevents	< 0.15	67018	0	67018
	>= 0.15	0	0	0
	Total	67018	0	67018

	CRS Model	CRS+PGS Model		
		< 0.175	>= 0.175	Total
Events	<17.5%	1281	0	1281
	>=17.5%	0	0	0
	Total	1281	0	1281
Nonevents	<17.5%	67018	0	67018
	>=17.5%	0	0	0
	Total	67018	0	67018

	CRS Model	CRS+PGS Model		
		< 0.2	>= 0.2	Total
Events	< 0.2	1281	0	1281
	>= 0.2	0	0	0
	Total	1281	0	1281
Nonevents	< 0.2	67018	0	67018
	>= 0.2	0	0	0
	Total	67018	0	67018

Incorrectly reclassified
 Correctly reclassified

Supplementary Figure S7: Categorical NRI of CRS+PXS model over range of risk thresholds

		CRS+PXS Model			
		CRS Model	< 0.025	>= 0.025	Total
Events	< 0.025		248	44	292
	>= 0.025		54	935	989
	Total		302	979	1281
Nonevents	< 0.025		48730	1520	50250
	>= 0.025		2460	14308	16768
	Total		51190	15828	67018

		CRS+PXS Model			
		CRS Model	< 0.05	>= 0.05	Total
Events	< 0.05		455	68	523
	>= 0.05		71	687	758
	Total		526	755	1281
Nonevents	< 0.05		56890	1325	58215
	>= 0.05		2072	6731	8803
	Total		58962	8056	67018

		CRS+PXS Model			
		CRS Model	< 0.075	>= 0.075	Total
Events	< 0.075		666	125	791
	>= 0.075		80	410	490
	Total		746	535	1281
Nonevents	< 0.075		62038	1373	63411
	>= 0.075		1075	2532	3607
	Total		63113	3905	67018

		CRS+PXS Model			
		CRS Model	< 0.1	>= 0.1	Total
Events	< 0.1		923	208	1131
	>= 0.1		44	106	150
	Total		967	314	1281
Nonevents	< 0.1		65227	1259	66486
	>= 0.1		196	336	532
	Total		65423	1595	67018

		CRS+PXS Model			
		CRS Model	< 0.125	>= 0.125	Total
Events	< 0.125		1165	116	1281
	>= 0.125		0	0	0
	Total		1165	116	1281
Nonevents	< 0.125		66672	346	67018
	>= 0.125		0	0	0
	Total		66672	346	67018

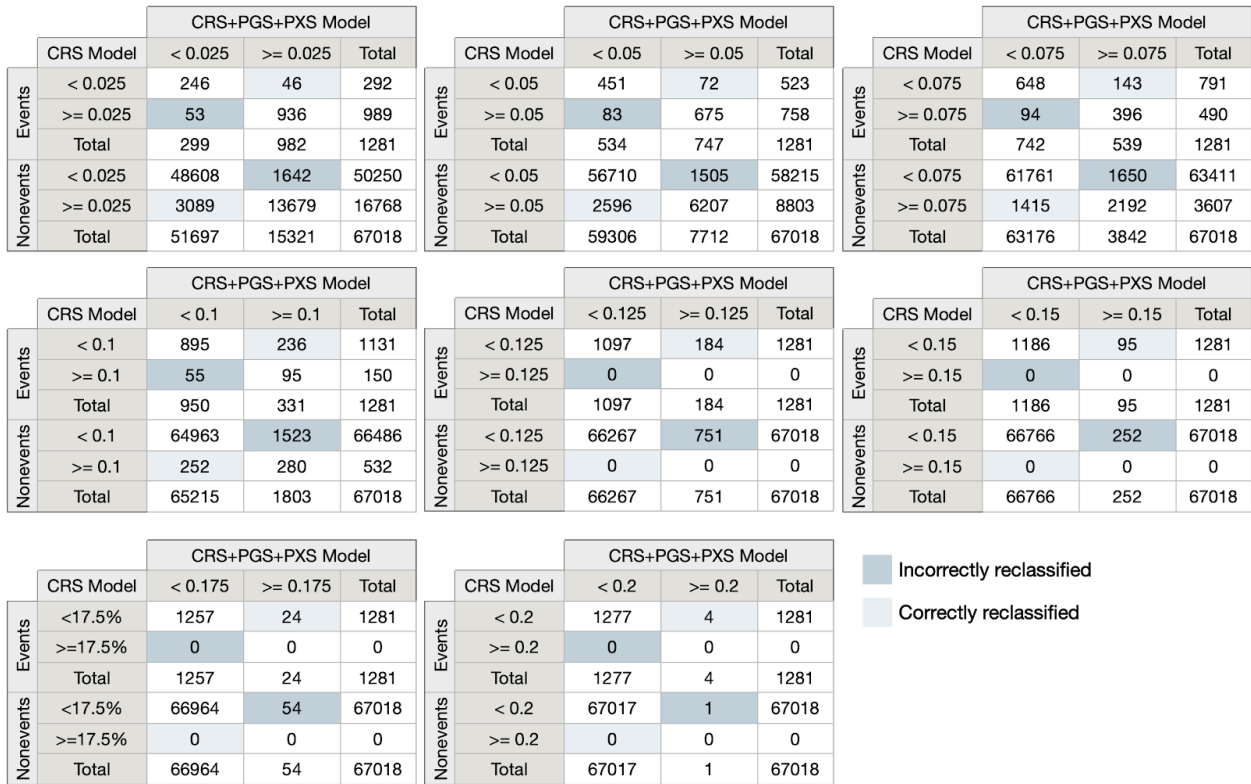
		CRS+PXS Model			
		CRS Model	< 0.15	>= 0.15	Total
Events	< 0.15		1281	0	1281
	>= 0.15		0	0	0
	Total		1281	0	1281
Nonevents	< 0.15		67018	0	67018
	>= 0.15		0	0	0
	Total		67018	0	67018

		CRS+PXS Model			
		CRS Model	< 0.175	>= 0.175	Total
Events	<17.5%		1281	0	1281
	>=17.5%		0	0	0
	Total		1281	0	1281
Nonevents	<17.5%		67018	0	67018
	>=17.5%		0	0	0
	Total		67018	0	67018

		CRS+PXS Model			
		CRS Model	< 0.2	>= 0.2	Total
Events	< 0.2		1281	0	1281
	>= 0.2		0	0	0
	Total		1281	0	1281
Nonevents	< 0.2		67018	0	67018
	>= 0.2		0	0	0
	Total		67018	0	67018

Incorrectly reclassified
 Correctly reclassified

Supplementary Figure S8: Categorical NRI of CRS+PGS+PXS model over range of risk thresholds



Supplementary Figure S9: Reclassification of predicted type 2 diabetes risk with glucose removed from CRS. The reclassified predicted risk with addition of polygenic risk score (PGS) (A), polygenic risk score (PXS) (B), or PGS+PXS (C) to clinical risk score (CRS) model without blood glucose in the continuous case or the categorical case, with a threshold of 12.5% risk.

A

	CRS+PGS Model		
CRS Model	# Participants	Continuous NRI	Categorical NRI
Cases	1281	0.157 (0.118 to 0.296)	0.016 (0.000 to 0.049)
Noncases	67018	0.084 (0.066 to 0.102)	-0.002 (-0.004, 0.00)
Full population	68299	0.241 (0.186 to 0.296)	0.015 (0.000 to 0.045)

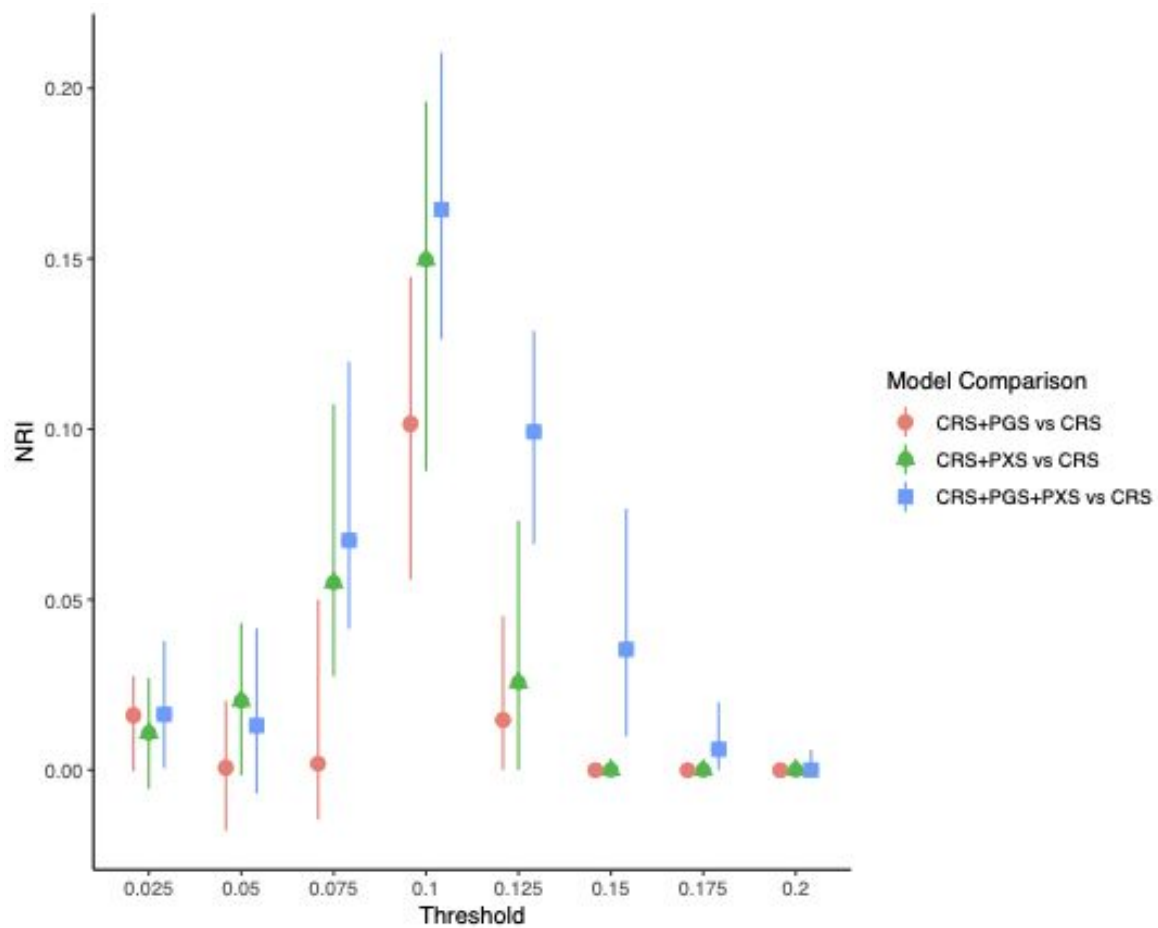
B

	CRS+PXS Model		
CRS Model	# Participants	Continuous NRI	Categorical NRI
Cases	1281	0.315 (0.276 to 0.351)	0.027 (0.000 to 0.078)
Noncases	67018	0.179 (0.156 to 0.202)	-0.001 (-0.005 to 0.000)
Full population	68299	0.494 (0.440 to 0.548)	0.026 (0.00 to 0.073)

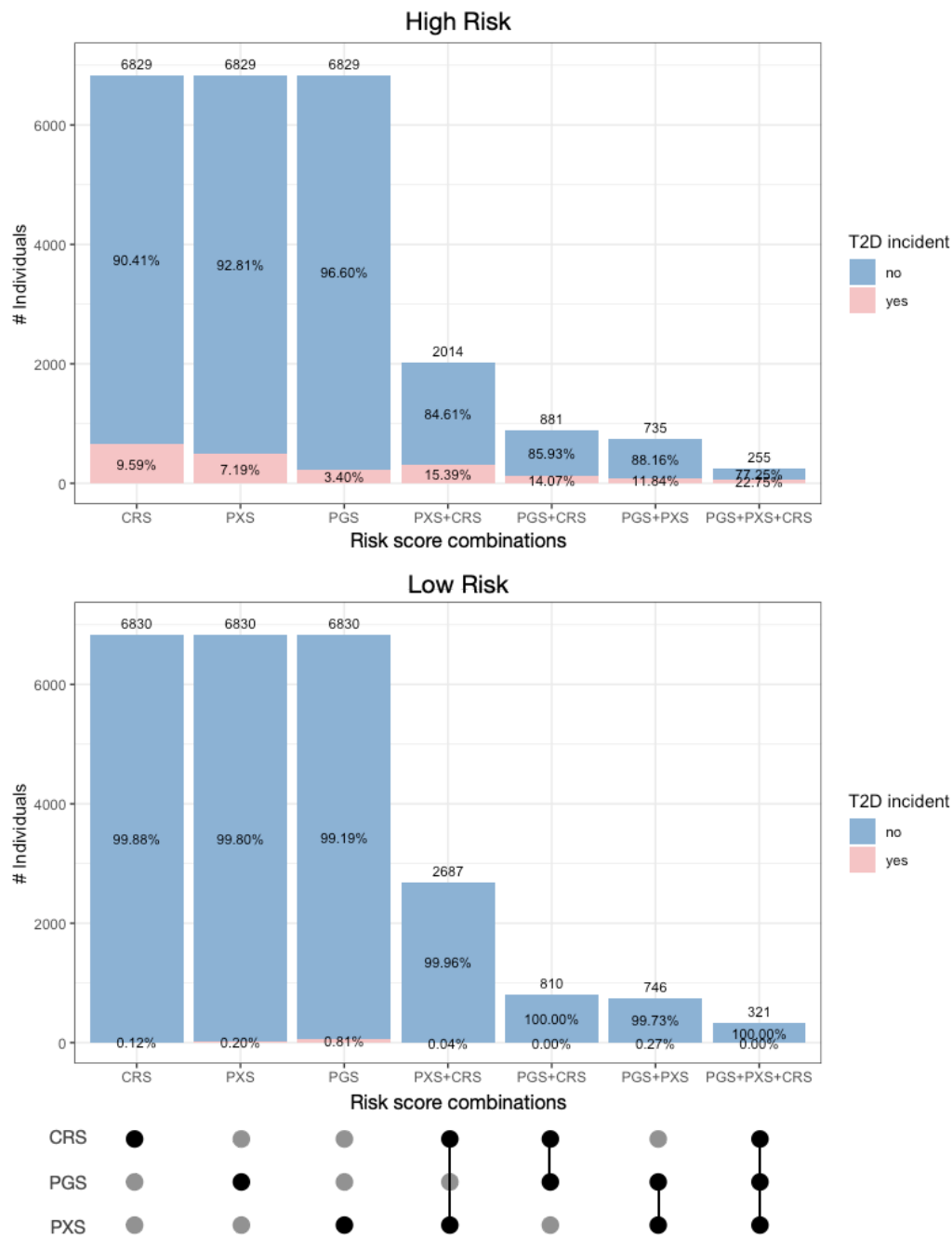
C

	CRS+PGS+PXS Model		
CRS Model	# Participants	Continuous NRI	Categorical NRI
Cases	1281	0.258 (0.214 to 0.305)	0.107 (0.071 to 0.140)
Noncases	67018	0.235 (0.207 to 0.256)	-0.008 (-0.012 to -0.005)
Full population	68299	0.493 (0.430 to 0.552)	0.099 (0.066 to 0.129)

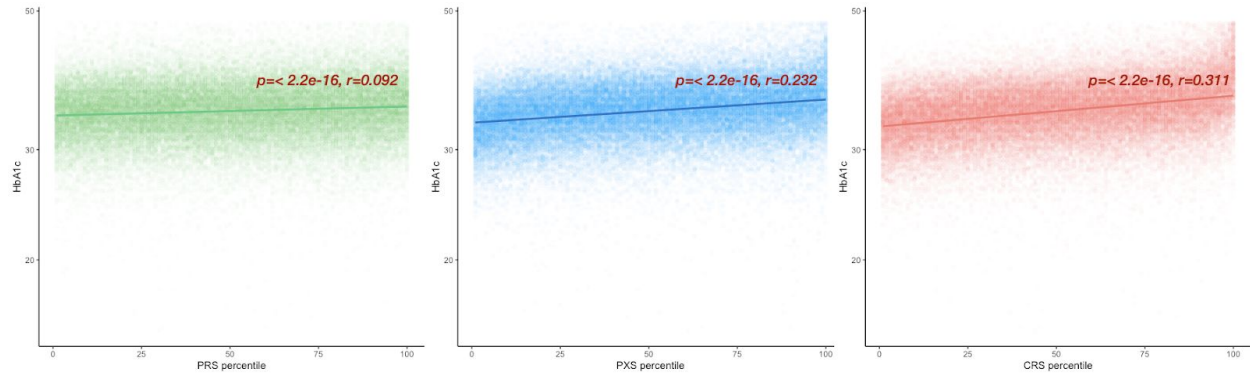
Supplementary Figure S10: NRI of risk models over a range of risk thresholds with blood glucose removed from CRS. The categorical NRI for different risk models.



Supplementary Figure S11: T2D incidence in Individuals with high (top) and low (bottom) risk CRS, PGS, and/or PXS. Individuals with scores in the top 10 percentiles were classified as high risk (top). Those with scores in the bottom 10 percentiles were classified as low risk (bottom). The total number of individuals within each risk score group is indicated at the top of each bar. Within each high risk group, the percent of individuals who received a formal diagnosis is shaded in pink. A black circle denotes the group that consists of that high risk group; a greyed circle denotes the group that does not consist of that high risk group.



Supplementary Figure S12: Correlation between HbA1c and risk scores in the test set. P values and pearson correlation coefficients are shown above the linear regression lines.



Supplementary Table S1: eXposure wide association study. Summary statistics of univariate XWAS. (see attached .xlsx).

Supplementary Table S2: All exposures extracted from UK Biobank. (see attached .xlsx)

Supplementary Table S3: T2D polyexposure risk score (PXS). The PXS was composed of 12 variables that were indicators of alcohol, diet, early life factors, household information, sleep, smoking. The hazard ratio and p values are shown. Hazard ratio of categorical responses are in comparison to the reference responses. Variables are colored by indicator category.

Field	PXS		Reference	HR	P	+ CRS		+ PGS		+CRS+ PGS		
	Response					HR	P	HR	P	HR	P	
Sex	Male	Female		1.63	0.00	1.32	0.00	1.64	0.00	1.35	0.00	
Year of birth				0.94	0.00	0.94	0.00	0.94	0.00	0.93	0.00	
Own or rent accommodation lived in	None of the above	Own outright (by you or someone in your household)		1.21	0.68	0.91	0.87	1.25	0.62	0.93	0.90	household
Own or rent accommodation lived in	Own with a mortgage	Own outright (by you or someone in your household)		1.25	0.00	1.24	0.00	1.25	0.00	1.25	0.00	sleep
Own or rent accommodation lived in	Rent - from local authority, local council, housing association	Own outright (by you or someone in your household)		1.38	0.00	1.19	0.14	1.36	0.00	1.18	0.16	smoking
Own or rent accommodation lived in	Rent - from private landlord or letting agency	Own outright (by you or someone in your household)		2.10	0.00	1.66	0.00	2.05	0.00	1.65	0.00	diet
Own or rent accommodation lived in	Pay part rent and part mortgage (shared ownership)	Own outright (by you or someone in your household)		2.24	0.02	2.65	0.01	2.09	0.04	2.56	0.01	alcohol
Own or rent accommodation lived in	Live in accommodation rent free	Own outright (by you or someone in your household)		0.98	0.96	0.76	0.51	1.04	0.90	0.78	0.55	early life
Average total household income before tax	18,000 to 30,999	Less than 18,000		0.53	0.00	0.67	0.01	0.54	0.00	0.68	0.01	
Average total household income before tax	31,000 to 51,999	Less than 18,000		0.83	0.09	0.86	0.22	0.84	0.11	0.88	0.30	
Average total household income before tax	52,000 to 100,000	Less than 18,000		0.87	0.09	0.91	0.28	0.86	0.07	0.89	0.20	
Average total household income before tax	Greater than 100,000	Less than 18,000		0.94	0.31	0.88	0.06	0.95	0.40	0.89	0.09	
Sleeplessness / insomnia	Sometimes	Never/rarely		1.34	0.00	1.33	0.00	1.32	0.00	1.32	0.00	
Sleeplessness / insomnia	Usually	Never/rarely		1.02	0.69	0.96	0.46	1.02	0.67	0.98	0.60	
Snoring	No	Yes		0.75	0.00	0.91	0.03	0.76	0.00	0.92	0.03	
Past tobacco smoking	Smoked occasionally	Smoked on most or all days		0.71	0.00	0.72	0.00	0.72	0.00	0.73	0.00	
Past tobacco smoking	Just tried once or twice	Smoked on most or all days		1.36	0.00	1.25	0.00	1.35	0.00	1.25	0.00	
Past tobacco smoking	I have never smoked	Smoked on most or all days		0.95	0.44	0.96	0.58	0.94	0.38	0.93	0.40	
Milk type used	Full cream	Semi-skimmed		0.91	0.42	0.95	0.71	0.93	0.53	0.98	0.87	
Milk type used	Skimmed	Semi-skimmed		1.14	0.04	1.07	0.31	1.12	0.05	1.08	0.24	
Milk type used	Soya	Semi-skimmed		0.48	0.00	0.55	0.01	0.49	0.00	0.56	0.01	
Milk type used	Other type of milk	Semi-skimmed		0.89	0.61	1.04	0.88	0.91	0.70	1.09	0.76	
Milk type used	Never/rarely have milk	Semi-skimmed		0.79	0.16	0.70	0.08	0.80	0.17	0.71	0.09	
Spread type	Never/rarely use spread	Other type of spread/margarine		0.65	0.00	0.72	0.00	0.65	0.00	0.70	0.00	
Spread type	Butter/spreadable butter	Other type of spread/margarine		1.04	0.52	1.06	0.39	1.04	0.51	1.05	0.42	
Spread type	Flora Pro-Active/Benecol	Other type of spread/margarine		1.52	0.15	1.72	0.09	1.50	0.16	1.72	0.09	
Tea intake				0.88	0.00	0.93	0.01	0.88	0.00	0.92	0.00	
Never eat eggs, dairy, wheat, sugar	Eggs or foods containing eggs	I eat all of the above		1.32	0.06	1.26	0.17	1.30	0.07	1.24	0.19	
Never eat eggs, dairy, wheat, sugar	Dairy products	I eat all of the above		1.46	0.07	1.35	0.25	1.47	0.07	1.33	0.28	
Never eat eggs, dairy, wheat, sugar	Wheat products	I eat all of the above		0.99	0.97	1.24	0.33	0.99	0.96	1.24	0.32	
Never eat eggs, dairy, wheat, sugar	Sugar or foods/drinks containing sugar	I eat all of the above		1.77	0.00	1.51	0.00	1.74	0.00	1.52	0.00	
Major dietary changes in the last 5 years	Yes, because of illness	No		3.69	0.00	2.79	0.00	3.57	0.00	2.72	0.00	
Major dietary changes in the last 5 years	Yes, because of other reasons	No		1.33	0.00	1.23	0.00	1.32	0.00	1.22	0.01	
Alcohol intake frequency	Three or four times a week	Daily or almost daily		2.05	0.00	1.45	0.00	2.01	0.00	1.46	0.00	
Alcohol intake frequency	Once or twice a week	Daily or almost daily		0.93	0.29	0.99	0.93	0.94	0.40	1.00	0.98	
Alcohol intake frequency	One to three times a month	Daily or almost daily		0.82	0.00	0.86	0.04	0.82	0.00	0.87	0.05	
Alcohol intake frequency	Special occasions only	Daily or almost daily		0.98	0.74	0.99	0.85	0.96	0.57	0.98	0.75	
Alcohol intake frequency	Never	Daily or almost daily		1.00	0.99	1.03	0.72	0.99	0.85	1.02	0.77	
Comparative body size at age 10	Plumper	Thinner		0.87	0.00	0.80	0.00	0.87	0.00	0.80	0.00	
Comparative body size at age 10	About average	Thinner		0.85	0.00	1.15	0.03	0.87	0.01	1.16	0.02	
BMI				-	-	1.10	0.00	-	-	1.00	0.04	
Systolic blood pressure				-	-	0.42	0.00	-	-	1.10	0.00	
Glucose				-	-	1.59	0.00	-	-	1.58	0.00	
HDL cholesterol				-	-	0.42	0.00	-	-	0.42	0.00	
Triglycerides				-	-	1.09	0.00	-	-	1.07	0.00	
Family History of Diabetes				-	-	1.90	0.00	-	-	1.81	0.00	
PXS				-	-	-	-	24.89	0.00	14.32	0.00	

Supplementary Table S4: Clinical Risk Score (CRS). Weights of 8 clinical risk factors. *Sex served as an interactor, not a mediator, in the Framingham Heart Study.

CRS model					Meigs et al.	
Field	Class	HR	95% CI	P-val	Field	OR
Male (versus Female)	Binary	1.30	1.20 - 1.40	2.58E-07	Male (versus Female)	0.66*
Year of birth	Continuous	0.94	0.93 - 0.94	6.38E-80	Age (years)	1.05
Systolic blood pressure (mmHg)	Continuous	1.00	1.00 - 1.01	6.80E-03	Systolic blood pressure (mmHg)	1.01
BMI (kg/m2)	Continuous	1.12	1.12 - 1.13	1.96E-215	BMI (kg/m2)	1.14
Glucose (mmol/L)	Continuous	1.60	1.57 - 1.63	2.57E-275	Glucose (mg/dL)	1.14
HDL cholesterol (mmol/L)	Continuous	0.27	0.09 - 0.45	2.48E-47	HDL cholesterol (mg/dL))	0.99
Triglycerides (mmol/L)	Continuous	1.12	1.08 - 1.15	8.63E-10	Triglycerides (mg/dL)	1.00
Family History of Diabetes (versus no history)	Binary	1.99	1.90 - 2.05	8.98E-49	Family History of Diabetes (versus no history)	1.72

Supplementary Table S5: The C-statistics of the new PXS model when each of the exposure variables were iteratively dropped from the score in subset of the validation set of N=68,299 with complete information for all variables.

Dropped Exposure	C-Statistic (95% CI)
Major dietary changes in the last 5 years	0.737 (0.724, 0.75)
Snoring	0.756 (0.743, 0.769)
Alcohol intake frequency	0.757 (0.744, 0.77)
Own or rent accommodation lived in	0.759 (0.746, 0.772)
Sleeplessness / insomnia	0.760 (0.747, 0.773)
Never eat eggs, dairy, wheat, sugar	0.760 (0.747, 0.773)
Comparative body size at age 10	0.760 (0.747, 0.773)
Average total household income before tax	0.761 (0.748, 0.774)
Past tobacco smoking	0.761 (0.749, 0.773)
Spread type	0.761 (0.749, 0.773)
Tea intake	0.761 (0.749, 0.773)
Milk type used	0.767 (0.755, 0.779)

Supplementary Table S6: C-Statistics for evaluating performance of risk models in sensitivity analysis. The performance was evaluated in the full test population as well as stratified by sex and year of birth where blood glucose removed from CRS. * Denotes models adjusted for all covariates (sex, age, assessment center, and genetic principal components) in the full test population, all covariates except sex in the sex stratified analysis, and all covariates except age in year of birth stratified analysis.

	C-Statistic (95% CI), blood glucose removed from CRS						
	All	Male	Female	Born before 1945	Born between 1945-1950	Born between 1950-1958	Born after 1958
N	68299	32657	35642	15032	16529	18547	18191
# of Events	1281	844	437	468	377	291	145
Sex+Age	0.670 (0.656, 0.684)	0.629 (0.612, 0.646)	0.637 (0.612, 0.662)	0.594 (0.569, 0.619)	0.606 (0.579, 0.633)	0.624 (0.593, 0.655)	0.592 (0.548, 0.636)
PGS*	0.709 (0.696, 0.722)	0.680 (0.663, 0.697)	0.705 (0.682, 0.728)	0.658 (0.634, 0.682)	0.674 (0.648, 0.7)	0.713 (0.687, 0.739)	0.719 (0.678, 0.760)
PXS*	0.762 (0.749, 0.775)	0.732 (0.716, 0.748)	0.774 (0.753, 0.795)	0.714 (0.69, 0.738)	0.718 (0.692, 0.744)	0.785 (0.759, 0.811)	0.782 (0.743, 0.821)
CRS*	0.820 (0.810, 0.820)	0.789 (0.775, 0.803)	0.842 (0.824, 0.86)	0.774 (0.754, 0.794)	0.791 (0.77, 0.812)	0.838 (0.818, 0.858)	0.845 (0.811, 0.879)
PGS+PXS*	0.776 (0.764, 0.788)	0.749 (0.734, 0.764)	0.786 (0.765, 0.807)	0.729 (0.706, 0.752)	0.730 (0.705, 0.755)	0.803 (0.779, 0.827)	0.802 (0.764, 0.84)
CRS+PGS*	0.826 (0.816, 0.836)	0.797 (0.783, 0.811)	0.848 (0.831, 0.865)	0.781 (0.762, 0.8)	0.796 (0.775, 0.817)	0.849 (0.83, 0.868)	0.856 (0.824, 0.888)
CRS+PXS*	0.835 (0.825, 0.845)	0.807 (0.793, 0.821)	0.855 (0.838, 0.872)	0.790 (0.771, 0.809)	0.801 (0.708, 0.822)	0.861 (0.842, 0.88)	0.856 (0.822, 0.890)
CRS+PXS+PGS*	0.841 (0.831, 0.851)	0.814 (0.801, 0.827)	0.86 (0.843, 0.877)	0.797 (0.778, 0.816)	0.805 (0.784, 0.826)	0.87 (0.852, 0.888)	0.866 (0.834, 0.898)

Supplementary Table S7: Hazard ratio (HR) for individuals with highest risk of T2D. The HR and 95 CI for individuals in the top 1% (versus the bottom 99%), 5% (versus the bottom 95th), 10% (versus the bottom 90th), and 20% (versus the bottom 80th) of using the polygenic risk score (PGS), polyexposure risk score (PXS), or clinical risk score (CRS) models.

		Hazard Ratio (95% CI)		
		PGS	PXS	CRS
Top % score	1%	2.64 (1.87, 3.73)	9.74 (7.96, 11.91)	15.11 (12.74, 17.92)
	5%	2.27 (1.90, 2.71)	6.72 (5.92, 7.63)	10.54 (9.39, 11.83)
	10%	2.00 (1.73, 2.31)	5.90 (5.28, 6.61)	9.97 (8.94, 11.13)
	20%	1.96 (1.75, 2.21)	4.72 (4.23, 5.27)	9.51 (8.44, 10.71)

Supplementary Table S8: Sensitivity of risk scores in participants with undiagnosed T2D. Undiagnosed individuals (N=3,658) with scores in the top 10th of the reference group were classified as high risk. Those with scores in the bottom 10th of the reference group were classified as low risk.

Score type	High Risk N	Low Risk N	High Risk Incidence	Low Risk Incidence
CRS	2869	34	1572 (54.79%)	5 (14.71%)
PXS	1436	75	878 (61.14%)	18 (24.00%)
PGS	83	132	456 (54.68%)	73 (55.30%)
PXS and CRS	1217	7	755 (62.04%)	0 (0.00%)
PGS and CRS	657	1	375 (57.08%)	0 (0.00%)
PGS and PXS	328	3	216 (65.85%)	1 (33.33%)
PGS and PXS and CRS	274	0	184 (67.15%)	0 (NA)