SUPPLEMENTAL APPENDIX

Sensitivity analyses with Framingham Risk score

We assessed the DPP trial with a diabetes prediction model from the Framingham cohort.[1] The Framingham diabetes Risk Score (FRS) was developed in a normal risk population and assessed the risk of developing diabetes over 7 years. To account for the shorter, three-year follow-up in the DPP, we multiplied the risk of developing diabetes by 3/7, a previously-developed technique.[2] We then also calculated predictions based on a recalibrated FRS, updating the prediction model to the DPP cohort. To do this, we refit the Cox proportional hazards regression model with linear FRS as the only predictor, thus adjusting both the baseline hazard function (intercept) and the overall calibration slope. We then evaluated these model predictions (FRS and recalibrated score) in the same manner that we evaluated the internal model.

The FRS had an c-statistic of 0.69, only a small amount worse than the internal model's discrimination of 0.73. However, the FRS substantially under-predicted risk at all levels until the top decile (eFigure 1). After adjusting for length of follow-up, the Framingham model would have predicted 33.3% of placebo patients would develop diabetes; only 26.2% did.

Recalibration did not change the discrimination of the FRS model (c-statistic = 0.69), but did dramatically improve the calibration (eFigure 1).

The value of the recalibration suggests that the main problem is differences in the model constant (estimated risk when risk factors are absent), not the relative effects of individual risk factors. The suboptimal calibration did not substantially alter the model's ability to recognize heterogeneity of treatment effect, which is not surprising given their similar discrimination. The overprediction of the FRS does alter the estimated net benefit assessment of each quartile, however.

Once stratified by risk quartile, the absolute reduction estimated for the Framingham Model is quite similar to that seen with the DPP model. This is because of the similar discrimination between the two models (eFigure 2).

Decision support tools: Net benefit assessment and nomogram

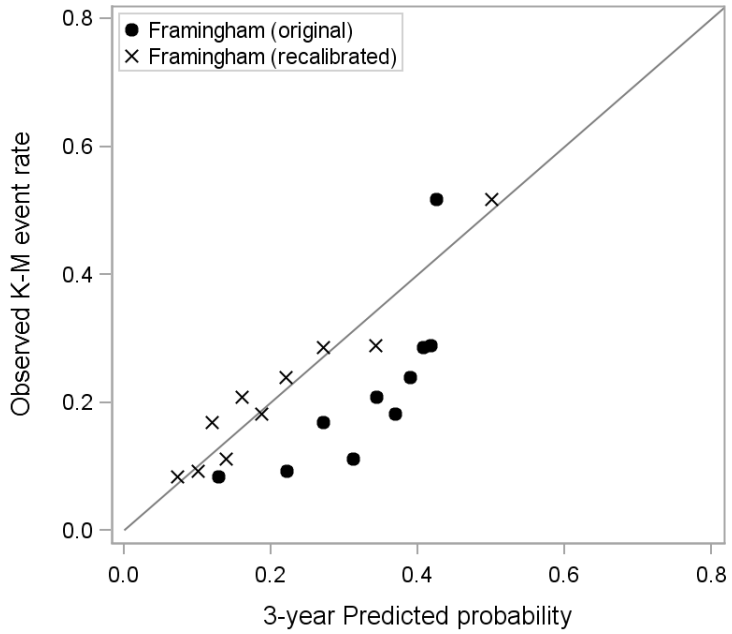
Net benefit assessment:

The primary benefit of understanding the heterogeneity of treatment effect found within a clinical trial is that it allows for greater personalization of decision-making. In particular, it can help decision makers, including patients, clinicians, or policy-makers, decide if the benefit of treatment outweighs the harms for them. One particularly useful method to help make decisions is called the net benefit assessment method.[3,4] This method, which is based on decision curve analysis,[5] allows the comparison of the net population-wide benefit of a risk-based approach relative to treating everyone in that population ("treat all", which follows the common expectation of a clinical trial) versus treating no one ("treat none") over a range of treatment thresholds. These treatment thresholds represent the implicit burden of taking the treatment, so for example a treatment threshold of 0.1 means that the harms of 3 years of treatment for 10 people are considered equivalent to one case of diabetes. This threshold is inherently subjective, so the net benefit assessment allows decision-making that recognizes different threshold values.

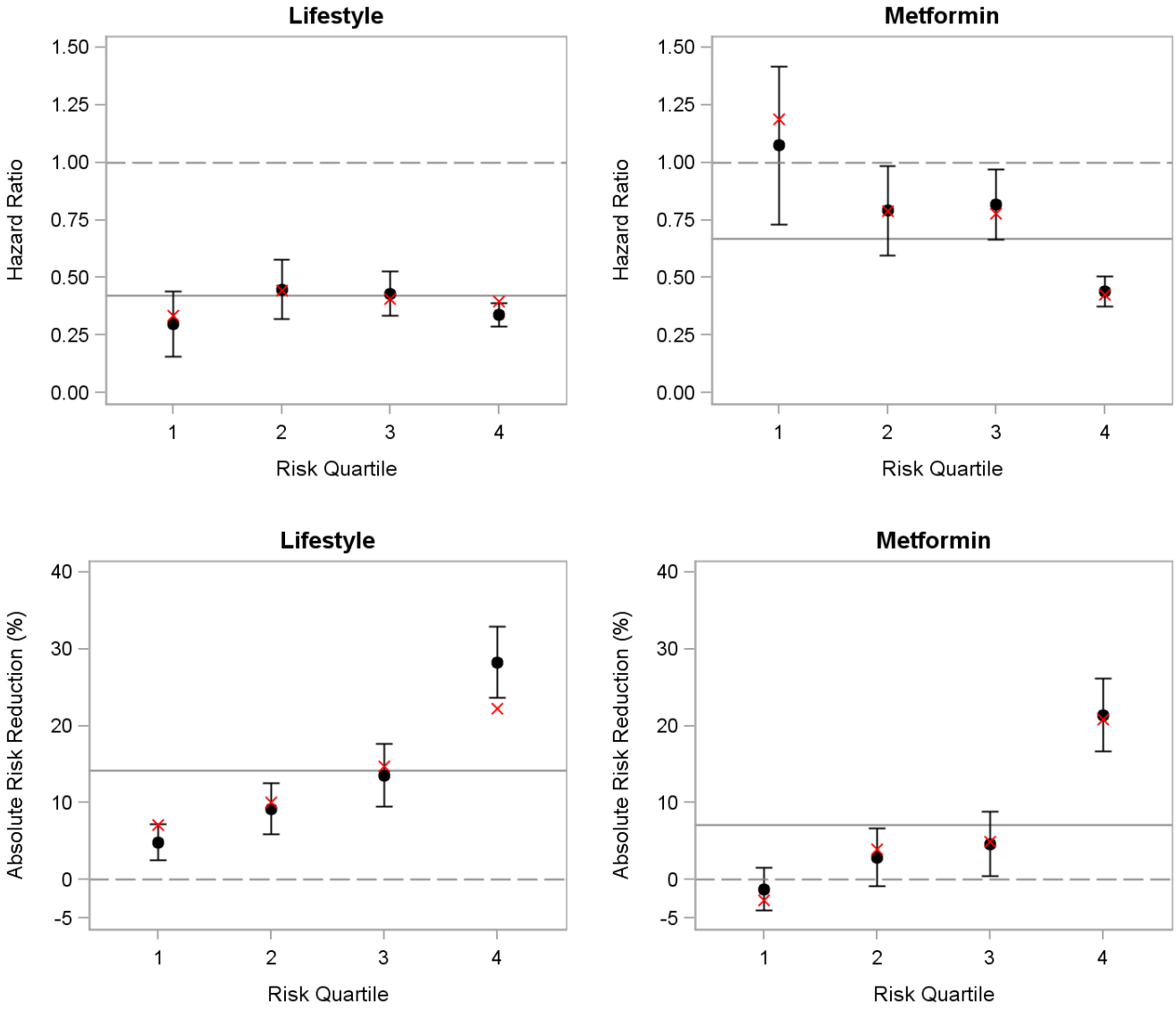
We created net benefit assessment curves and tables using the internal model across the range of possible net benefit thresholds. The results show that basing treatment decisions off the prediction model offers benefit compared to either a treat all or treat none strategy over a range of clinically plausible treatment thresholds for both interventions, particularly for metformin (eFigure 3 and eTable 1). If all patients receive metformin and we assume that metformin has no treatment burden, then there is a net benefit of 6 cases of diabetes prevented for every 100 people over 3 years. However, if it were determined that we should only treat those patients who were likely to have at least a 7% absolute risk reduction greater, then one implicitly is weighting the treatment burden as equally as important as this degree of benefit. At this treatment threshold a treat all and a treat none strategy are equivalent; at any level beyond it, treating all patients is considered on net harmful. But at almost all treatment all thresholds up to 17%, treatment cold be more efficient with a risk-based decision rule than with treating all or treating none. For the lifestyle intervention, from 0 to 10% absolute risk reduction decision threshold nearly everyone would be recommended treatment, but beyond that, using the risk-based guidance could substantially improve care.

Using the Framingham diabetes prediction model still showed no circumstances where treating all or treating none substantially outperformed selecting treatment based on absolute risk reduction, but the amount of benefit was substantially smaller.

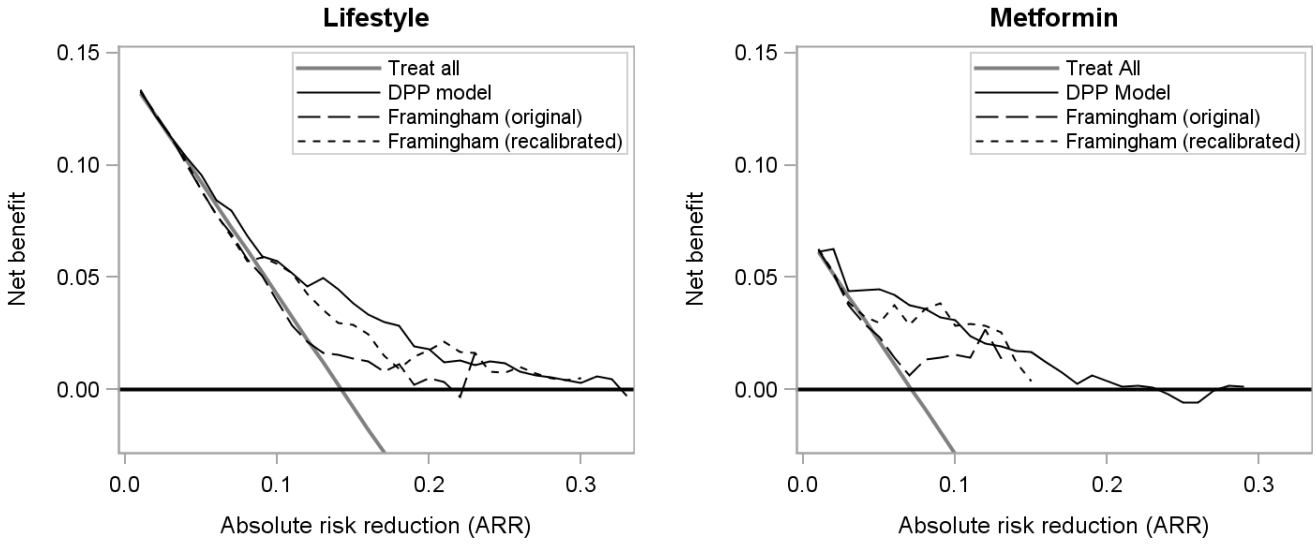
eFigure 1: Framingham risk score: The circles represent the original Framingham risk prediction model, the 'x' represent the re-calibrated Framingham model.



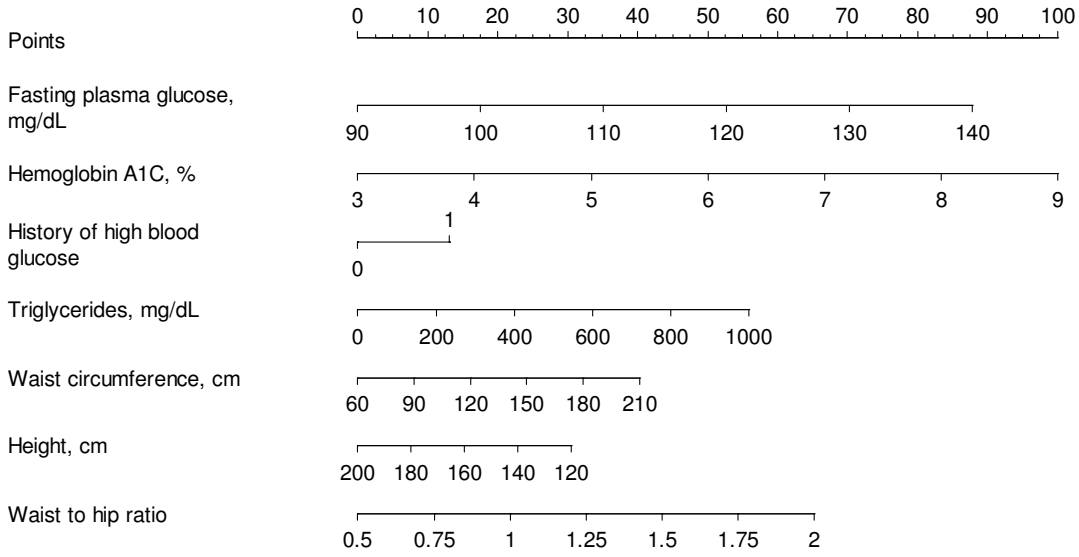
eFigure 2: Efficacy plots: The circle with error bars represent the DPP prediction model estimates, the "x" represent the recalibrated Framingham risk score point estimates.



eFigure 3: Net Benefit Assessment Curves (by intervention arm)



eFigure 4: Nomogram for internal model



eTable 1. Net benefit assessment for DPP 3-year risk of diabetes

				Prediction-based treatment net benefit (treatment rate, %):					
		Treat all patients		DPP Model		Framingham (original)		Framingham (recalibrated)	
NWT	ARR	Metformin	Lifestyle	Metformin	Lifestyle	Metformin	Lifestyle	Metformin	Lifestyle
100	0.01	0.0613	0.1322	0.061 (80)	0.133 (100)	0.062 (100)	0.133 (100)	0.063 (100)	0.133 (100)
50	0.02	0.0513	0.1222	0.063 (70)	0.122 (100)	0.051 (99)	0.123 (100)	0.052 (100)	0.123 (100)
34	0.03	0.0413	0.1122	0.044 (61)	0.113 (98)	0.038 (98)	0.113 (99)	0.039 (98)	0.113 (100)
25	0.04	0.0313	0.1022	0.044 (54)	0.104 (94)	0.029 (96)	0.1 (99)	0.033 (92)	0.102 (99)
20	0.05	0.0213	0.0922	0.045 (48)	0.096 (89)	0.023 (93)	0.089 (99)	0.03 (81)	0.089 (98)
17	0.06	0.0113	0.0822	0.042 (43)	0.084 (83)	0.014 (90)	0.078 (98)	0.038 (68)	0.078 (95)
15	0.07	0.0013	0.0722	0.038 (39)	0.08 (76)	0.006 (85)	0.07 (96)	0.029 (56)	0.068 (90)
13	0.08	-0.0087	0.0622	0.036 (36)	0.068 (70)	0.013 (78)	0.058 (95)	0.036 (45)	0.057 (85)
12	0.09	-0.0187	0.0522	0.032 (33)	0.059 (65)	0.014 (70)	0.051 (93)	0.038 (36)	0.059 (77)
10	0.1	-0.0287	0.0422	0.031 (30)	0.057 (59)	0.016 (63)	0.039 (91)	0.029 (30)	0.056 (70)
10	0.11	-0.0387	0.0322	0.024 (27)	0.052 (54)	0.015 (50)	0.028 (89)	0.029 (24)	0.052 (63)
9	0.12	-0.0487	0.0222	0.021 (25)	0.046 (49)	0.027 (35)	0.021 (87)	0.029 (17)	0.043 (56)
8	0.13	-0.0587	0.0122	0.019 (23)	0.05 (45)	0.014 (8)	0.016 (84)	0.026 (12)	0.036 (49)
8	0.14	-0.0687	0.0022	0.017 (21)	0.045 (41)	. (0)	0.016 (80)	0.012 (8)	0.03 (44)
7	0.15	-0.0787	-0.0078	0.017 (20)	0.038 (38)		0.014 (75)	0.004 (3)	0.029 (39)
7	0.16	-0.0887	-0.0178	0.012 (18)	0.033 (36)		0.013 (71)	. (0)	0.025 (34)
6	0.17	-0.0987	-0.0278	0.008 (16)	0.03 (33)		0.008 (66)		0.015 (31)
6	0.18	-0.1087	-0.0378	0.002 (15)	0.028 (30)		0.011 (61)		0.009 (28)
6	0.19	-0.1187	-0.0478	0.007 (13)	0.019 (28)		0.002 (54)		0.015 (24)
5	0.2	-0.1287	-0.0578	0.004 (12)	0.018 (26)		0.005 (47)		0.018 (21)
5	0.21	-0.1387	-0.0678	0.001 (11)	0.012 (24)		0.004 (39)		0.021 (19)
5	0.22	-0.1487	-0.0778	0.002 (10)	0.013 (22)		-0.003 (29)		0.017 (16)
5	0.23	-0.15867	-0.08778	0.001 (9)	0.011 (20)		0.016 (13)		0.016 (13)
5	0.24	-0.16867	-0.09778	-0.002 (8)	0.013 (19)		. (0)		0.008 (11)
4	0.25	-0.17867	-0.10778	-0.006 (7)	0.012 (17)				0.008 (9)
4	0.26	-0.18867	-0.11778	-0.006 (6)	0.008 (14)				0.01 (7)
4	0.27	-0.19867	-0.12778	0 (4)	0.007 (13)				0.007 (6)
4	0.28	-0.20867	-0.13778	0.002 (3)	0.005 (11)				0.005 (4)
4	0.29	-0.21867	-0.14778	0.001 (1)	0.004 (10)				0.004 (3)
4	0.3	-0.22867	-0.15778	. (0)	0.003 (9)				0.005 (1)
4	0.31	-0.23867	-0.16778		0.006 (7)				. (0)
4	0.32	-0.24867	-0.17778		0.005 (5)				
4	0.33	-0.25867	-0.18778		-0.003 (3)				

eTable 2: 3-year Absolute risk reduction ranges by quartile of risk: Risk quartile ARR ranges are calculated by applying intervention-quartile specific hazard ratios to the quartiles of baseline 3-year predicted probability.

Total Points from nomogram	Risk quartile	Untreated event rate range - untreated	Absolute risk reduction (range) Metformin		Lifestyle	
< 106	1	< 0.095	-0.4%	-0.7%	3.5%	6.7%
106-120	2	0.095 - 0.151	2.0%	3.2%	5.2%	8.3%
120-137	3	0.151 – 0.27	2.7%	4.9%	8.6%	15.4%
> 137	4	> 0.27	15.1%	28.0%	17.8%	33.0%

Supplement Works Cited

1. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB, Sr. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Archives of internal medicine.* May 28 2007;167(10):1068-1074.

2. Mann DM, Bertoni AG, Shimbo D, et al. Comparative validity of 3 diabetes mellitus risk prediction scoring models in a multiethnic US cohort: the Multi-Ethnic Study of Atherosclerosis. *American journal of epidemiology.* May 1 2010;171(9):980-988.

3. Dorresteijn JAN, Visseren FLJ, Ridker PM, et al. Estimating treatment effects for individual patients based on the results of randomised clinical trials. *Brit Med J.* Oct 3 2011;343.

4. Vickers AJ, Kattan MW, Daniel S. Method for evaluating prediction models that apply the results of randomized trials to individual patients. *Trials.* 2007;8:14.

5. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Medical decision making : an international journal of the Society for Medical Decision Making.* Nov-Dec 2006;26(6):565-574.

6. Harrell J, Frank E. rms: Regression Modeling Strategies. R package version 4.0-0. 2013; http://CRAN.R-project.org/package=rms, 2013.

7. R: A language and environment for statistical computing. 2013; http://www.R-project.org/.