

An Electronic Health Record Compatible Model to Predict Personalized Treatment Effects from the Diabetes Prevention Program: A Cross-Evidence Synthesis Approach Using Clinical Trial and Real World Data

David M Kent¹ MD, MS; Jason Nelson¹ MPH; Anastassios Pittas² MD, MS; Francis Colangelo³ MD, MS; Carolyn Koenig⁴ MD; David van Klaveren^{1,5} PhD; Elizabeth Ciemins⁶ PhD; John Cuddeback⁶ MD, PhD

¹Predictive Analytics and Comparative Effectiveness Center, Tufts Medical Center, Boston, MA

²Division of Endocrinology, Tufts Medical Center, Boston, MA

³Allegheny Health Network, Pittsburgh, PA

⁴Mercy Health, St. Louis, MO

⁵Department of Public Health, Erasmus MC University Medical Center, Rotterdam, the Netherlands

⁶AMGA (American Medical Group Association), Alexandria, VA

Running title: Predicting Individualized Benefits of Diabetes Prevention

Word count:

Abstract: 324

Main text: 3626

Figures: 4

Tables: 3

Corresponding Author:

David Kent, MD, MS

Director, Predictive Analytics and Comparative Effectiveness (PACE) Center

Institute for Clinical Research and Health Policy Studies

Tufts Medical Center

800 Washington St, Box 63

Boston, MA 02111

dkent1@tuftsmedicalcenter.org

1 **Abstract**

2 Background: An intensive lifestyle modification program or metformin pharmacotherapy reduced the
3 risk of developing diabetes in patients at high risk, but are not widely used in the 88 million American
4 adults with prediabetes.

5
6 Objective: Develop an electronic health record (EHR)-based risk tool that provides point-of-care
7 estimates of diabetes risk to support targeting interventions to patients most likely to benefit.

8
9 Design: Cross-design synthesis: risk prediction model developed and validated in large observational
10 database, treatment effect estimates from risk-based reanalysis of clinical trial data.

11
12 Setting: Outpatient clinics in US.

13
14 Patients: Risk model development cohort: 1.1 million patients with prediabetes from the OptumLabs
15 Data Warehouse (OLDW); validation cohort: distinct sample of 1.1 million patients in OLDW.
16 Randomized clinical trial cohort: 3081 people from the Diabetes Prevention Program (DPP) study.

17
18 Interventions: Randomization in the DPP: 1) an intensive program of lifestyle modification; 2) standard
19 lifestyle recommendations plus 850 mg metformin twice daily; or 3) standard lifestyle recommendations
20 plus placebo twice daily.

21
22 Results: Eleven variables reliably obtainable from the EHR were used to predict diabetes risk. This model
23 validated well in the OLDW (c -statistic = 0.76; observed 3-year diabetes rate was 1.8% in lowest-risk
24 quarter and 19.6% in highest-risk quarter). In the DPP, the hazard ratio for lifestyle modification was
25 constant across all levels of risk (HR = 0.43, 95% CI 0.35 – 0.53); while the HR for metformin was highly
26 risk-dependent (HR HR = 1.1 [95% CI: 0.61 - 2.0] in the lowest-risk quarter vs. HR=0.45 [95% CI: 0.35 -
27 0.59] in the highest risk quarter). Fifty-three percent of the benefits of population-wide dissemination of
28 the DPP lifestyle modification, and 76% of the benefits of population-wide metformin therapy can be
29 obtained targeting the highest risk quarter of patients.

30
31 Limitations: Differences in variable definitions and in missingness across observational and trial settings
32 may introduce estimation error in risk-based treatment effects.

33
34 Conclusion: An EHR-compatible risk model might support targeted diabetes prevention to more
35 efficiently realize the benefits of the DPP interventions.

36 **Introduction**

37 The Diabetes Prevention Program (DPP) Study showed that either an intensive program of lifestyle
38 modification or pharmacotherapy with metformin substantially reduced the risk of developing type 2
39 diabetes in patients at high risk, compared to “usual care.”¹ The findings have broad implications, as
40 “prediabetes” affects approximately 88 million US adults in the US.²

41 Strenuous calls to address the epidemic of diabetes with prevention^{3,4} have been counter-balanced
42 by concerns about the over-medicalization of prediabetes.⁵ Almost two decades after the publication of
43 the DPP Study, it remains unclear how best to implement these interventions in such an overwhelmingly
44 large, and mostly undiagnosed, population. A 2015 study examining a national sample of over 17,000
45 working-age adults with prediabetes found that only 3.7% were receiving metformin.⁶ Similarly,
46 widespread use of the intensive lifestyle intervention remains largely unrealized despite evidence that
47 rigorous diet and physical activity promotion reduces diabetes risk in the community setting.⁷

48 Yet, prediabetes is itself a heterogeneous condition. We previously showed that even among
49 patients enrolled in the DPP Study itself, the risk of developing diabetes within 3 years varies widely and
50 is highly skewed.⁸ Some trial participants were estimated to have a 1–2% risk, others 90%.
51 Unsurprisingly, the degree of benefit from metformin or from the lifestyle intervention was also
52 distributed unevenly.

53 This prior proof-of-concept work had several limitations. Notably, the risk distribution within the
54 DPP trial participants may differ from that of patients seen in routine practice, particularly since the
55 American Diabetes Association (ADA) has subsequently broadened its definition of prediabetes to
56 include a still more heterogeneous population.⁹ Further, the application of prediction methods to data
57 routinely collected in the electronic health record (EHR) provides a promising means to overcome some
58 of the major barriers to the use of risk models.^{10,11} For example, in addition to requiring manual
59 ascertainment of variables, the previously reported DPP-based model required waist circumference and
60 waist-to-hip ratio measurements that are not difficult to ascertain in routine practice. Herein, we
61 describe development of a clinical prediction model using a hybrid approach that makes use of routinely
62 collected EHR data to predict the risk of diabetes onset and clinical trial data to estimate unbiased risk-
63 based effects of preventive interventions.

64

65 **Methods**

66 *Overview*

67 We sought to develop and validate a diabetes risk prediction model using data elements readily
68 available in the EHR for dissemination across healthcare systems as an EHR-embedded tool, to facilitate
69 ease of use. The tool provides clinicians and their patients with an individualized risk of developing
70 diabetes and the estimated benefit of applying a DPP treatment strategy—either an intensive lifestyle
71 program or pharmacotherapy with metformin (the combination of both was not tested in the DPP
72 Study).

73

74 *Data sources and Participants*

75 The model was developed and validated using EHR data from the OptumLabs Data Warehouse (OLDW).
76 The OptumLabs EHR database is a geographically diverse sample of the US population with longitudinal
77 clinical data on over 33 million lives with at least one clinic visit during the study period. Using a
78 retrospective observational cohort design, we geographically stratified the database by US Census
79 Region into a development cohort of 1.1 million patients (North East, South, and West) and a separate
80 validation cohort of 1.1 million patients (Midwest).

81 Eligibility criteria included age between 25 and 75 upon an “index” office or clinic encounter (“index
82 visit” defined by CPT/HCPCS codes, see Appendix Table 1) between January 1, 2012, and December 31,
83 2016, at which time they met lab-based criteria for the diagnosis of prediabetes. Prediabetes was
84 defined by current American Diabetes Association (ADA) criteria, i.e., having no diagnosis of type 1 or
85 type 2 diabetes on the problem list and one of the following within 12 months prior to the visit:
86 hemoglobin A1c between 5.7 and 6.4% inclusive and/or fasting glucose (FG) between 100 and 125
87 mg/dL inclusive. Since labeling of fasting status may be incomplete, a glucose drawn at the same time as
88 a lipid panel or triglycerides was considered as fasting. We did not use the 2-hour post glucose load
89 criterion as it is rarely used in clinical practice for prediabetes. Patients were excluded if they had
90 random (non-fasting) glucose greater than or equal to 200 mg/dL on two occasions within a 3-month
91 period prior to the index visit. Women with documented pregnancy within 24 months of the index visit
92 were also excluded. To ascertain development of diabetes, patients also had to have some clinical
93 activity 3 years after the index visit. Eligibility criteria are detailed in Appendix Table 1.

94 The DPP dataset was used to estimate treatment effect for metformin or the intensive lifestyle
95 modification program. The design, rationale, outcomes, and loss to follow-up of the DPP have been
96 described in detail elsewhere^{1,12}. Briefly, inclusion criteria included a body mass index (BMI) of 24 or
97 higher (22 or higher in Asians) and a fasting plasma glucose concentration of 95 to 125 mg/dL inclusive
98 (impaired fasting glucose) and a concentration of 140 to 199 mg/dL inclusive two hours after a 75 g oral

99 glucose load (impaired glucose tolerance). We note these criteria differ from the ADA's current
100 diagnostic criteria for prediabetes we used for the OLDW model; the ADA definition imposes no BMI
101 requirement.¹³ The DPP participants were randomized to: 1) standard lifestyle recommendations plus
102 850 mg of metformin twice daily; 2) an intensive program of lifestyle modification that included 16
103 lessons with a case manager and set goals of at least a 7 percent weight loss and at least 150 minutes of
104 physical activity per week; or 3) standard lifestyle recommendations plus placebo twice daily. After a
105 median follow-up period of 2.8 (range 1.8–4.6) years, progression to diabetes was reduced by 58% (95%
106 confidence interval, 47% to 66%) in the lifestyle modification arm and 31% (17% to 43%) in the
107 metformin arm, both compared with the placebo arm¹. The NIDDK Data Repository, from which we
108 obtained data, includes 3081 of the 3234 DPP participants (95% of full population), as some local
109 institutional review boards declined to participate in data distribution.

110

111 *Outcome*

112 For the OLDW cohort, the time to event outcome was defined as the time to the first patient encounter
113 after the index visit with documented evidence of type 2 diabetes by any of the following criteria:¹⁴
114 diagnosis codes ICD-9 250.x0 or 250.x2 or ICD-10 E11.xx; pharmacotherapy or procedure for type 2
115 diabetes (as detailed in Appendix Table 1); A1c greater than 6.4%; fasting glucose (or presumed fasting,
116 as noted above) greater than 125 mg/dL; 2-hour OGTT post-load glucose greater than 199 mg/dL. Lab-
117 based criteria required confirmation by an additional lab in the diabetes range or by another method
118 (i.e., diagnosis or medication). Follow-up time for patients who did not meet the outcome definition was
119 censored at the first occurrence of the last observed encounter or end of study period.

120

121 *Candidate predictors*

122 A priori risk model predictors were identified by a systematic review conducted by Collins et al.¹⁵ We
123 selected the following 11 independent variables that were included in at least 3 prior diabetes risk
124 models and were judged to be easily and reliably obtainable in EHR data: age, gender, race, smoking
125 status, BMI, presence or absence of a diagnosis of hypertension, systolic blood pressure, HDL
126 cholesterol, triglycerides, fasting glucose, hemoglobin A1c (HgbA1c). Four variables included in 3 prior
127 models were not considered based on the difficulty of ascertaining them in EHR data: physical activity,
128 waist circumference, waist-to-hip ratio, and family history of diabetes.

129

130 *Missing data*

131 Missing data is a common limitation when working with EHR data.¹⁶ While multiple imputation may
132 improve estimates of parameter effects under a missing-at-random assumption, it does not provide a
133 practical means to cope with missingness in actual patients for whom a prediction needs to be made.
134 Thus, we used missing indicator variables to capture the predictive effects of missingness under the
135 assumption that future and prior missingness are similarly informative. For each predictor, an additional
136 dichotomous variable indicated the presence of missing values. For continuous variables (e.g., BMI,
137 HgbA1c), the missing value of the original variable was replaced by a fixed constant (the median) prior to
138 model estimation, and the missing indicator variable appropriately adjusted for the “missing variable
139 effect.” For categorical variables (e.g., race, smoking status), an additional level was added to define the
140 missing category.

141

142 *Model development*

143 We used multivariable Cox proportional hazards regression to estimate the predicted probability of
144 developing type 2 diabetes. We included two a priori interactions, race*BMI and race*HgbA1c, based
145 upon clinical judgment and the literature.^{17,18} Model performance was assessed for discrimination and
146 calibration. A bootstrap resampling procedure with 500 samples was used to internally validate the
147 model, estimate optimism-corrected discrimination, and assess calibration.

148

149 *Model validation*

150 Using the equation derived in the development cohort we calculated the predicted probability of
151 developing type 2 diabetes for patients in the validation cohort. Model performance upon external
152 validation was assessed for discrimination using Harrell’s measure of concordance for censored
153 response variable and calibration.¹⁹

154

155 *Estimating risk-specific treatment effects*

156 To estimate the risk-based treatment effect for metformin pharmacotherapy or the DPP lifestyle
157 modification, we performed a risk-based heterogeneity of treatment effect analysis on the DPP.²⁰ The
158 applicability of the OLDW model to the DPP data was anticipated to be limited by: differences between
159 predictor variable definitions and measurement within a trial context vs. EHR data, differences in the
160 pattern of missingness between these contexts (i.e., there was essentially no data missingness in the
161 DPP), differences in patient enrollment in the two settings and differences in outcome definition and
162 ascertainment.²¹ Thus, we refit the OLDW model to the DPP, using the same variables and interaction

163 terms. Consistent with methodological recommendations^{22,23}, all 3 DPP arms were used, since research
164 has shown that overfitting to a control arm can induce spurious heterogeneity of treatment effects.²⁴⁻²⁶
165 The treatment effect was then estimated by incorporating this linear predictor into a Cox proportional
166 hazards model with the following terms: treatment (metformin or DPP lifestyle modification), the linear
167 predictor of risk from the refitted model, and (potentially) an interaction between these to account for
168 important changes in relative risk reduction across different levels of baseline risk. Based on a previous
169 analysis,⁸ we anticipated a risk-by-treatment interaction with metformin pharmacotherapy and a
170 consistent relative effect with the DPP lifestyle modification, but we examined interactions for both
171 treatment arms. We also performed a sensitivity analysis, examining the risk-by-treatment interactions,
172 stratifying the DPP by the OLDW model without any refitting, and examining the distribution of
173 predicted effects using this model.

174

175 *Incorporation of Decision Support in Electronic Health Record*

176 In order to facilitate use in clinical decision making, based on patient and provider focus groups and
177 interviews, we implemented the model in two different ways: 1) a hard coded calculation in an Allscripts
178 EHR; 2) a cloud-hosted SMART on FHIR²⁷ app that can be incorporated into any EHR, leveraging
179 interoperability standards recently promulgated by the US office of the National Coordinator of Health
180 Information (ONC).

181

182 *Role of the Funding Source*

183 This study was funded by a Patient-Centered Outcomes Research Institute (PCORI), who had no role in
184 the conduct of this research or the decision to publish the results.

185

186 *IRB Approval*

187 This study was reviewed and approved by the Tufts Health Sciences Institutional Review Board prior to
188 accessing the deidentified data from the DPP and OLDW datasets.

189

190 **Results**

191 Figure 1 shows the development of the derivation and validation OLDW datasets. Approximately 1.1
192 million people with prediabetes from the Northeast, South and West were included in the derivation
193 cohort, and a similar number from the Midwest were included in the validation cohort. Characteristics
194 of these cohorts are shown in Table 1.

195

196 *Model development and validation: Risk stratification*

197 The coefficients for each of the variable and interaction terms included in the model are shown in Table
198 2. The optimism-corrected *c*-statistic on the derivation sample was 0.73. When the model was tested on
199 the validation cohort, the *c*-statistic was slightly higher, 0.76. Calibration on the validation cohort was
200 very good (Figure 2). Among the 268,959 patients in the lowest-risk quartile, the predicted diabetes rate
201 was 3.1% (95% confidence interval, 3.0 to 3.2%), while the observed rate was 1.8% (1.7 to 1.9%); among
202 the 268,958 patients in highest-risk quartile, the predicted diabetes rate was 19.2% (18.6 to 19.9%),
203 while the observed rate was 19.6% (19.4 to 19.8%).

204

205 *Calculation of relative treatment effects in the DPP Study*

206 Prior work demonstrated a consistent relative treatment effect across risk groups with the DPP lifestyle
207 modification and an increasing relative effect with progressively higher risk for metformin
208 pharmacotherapy.⁸ Using the OLDW model refit to the DPP data (Appendix Table 2; *c*-statistic 0.719),
209 we confirmed the absence of a treatment-by-risk interaction for lifestyle modification (*p* for interaction
210 = 0.68); thus, we applied a constant relative risk reduction in the prediction model (HR = 0.43; 95% CI:
211 0.35 – 0.53) to estimate the diabetes outcome with lifestyle modification. We also confirmed the
212 presence of a treatment-by-risk interaction with metformin pharmacotherapy (*p* for interaction= 0.003;
213 using the continuous risk on the logit scale): low-risk patients had outcomes with metformin that were
214 similar to usual care (in lowest risk quarter, observed HR = 1.1; 95% CI: 0.61 to 2.0), and high-risk
215 patients have outcomes with metformin that were similar to the DPP lifestyle modification (in highest
216 risk quarter, observed HR = 0.45; 95% CI: 0.35 to 0.59). Figure 3 shows observed and predicted benefits
217 across quartiles for the DPP, for both lifestyle and metformin. A look-up table showing the relative risk
218 reduction with metformin for each level of risk is shown in Appendix Table 3, truncated at a low value of
219 0% (no harm or benefit) and a high value of 60%.

220

221 *Distribution of risks and benefits in OLDW*

222 The overall average 3-year predicted risk of developing diabetes for patients in the validation OLDW
223 cohort was 9.0%, 3.9%, and 6.0%, with usual care, the DPP lifestyle diabetes and metformin respectively.
224 Predictions for the median patient in each quartile using the final model are shown in Table 2. For
225 lifestyle modification, 53% of the total preventable cases of diabetes could be prevented by treating the
226 25% of patients at highest risk; 76% by treating the 50% at highest risk and 91% by treating the 75% at

227 highest risk. For metformin therapy, 73% of the total preventable cases could be prevented by treating
228 the 25% of patients at highest risk; 93% by treating the 50% at highest risk, and 100% by treating the
229 75% at highest risk.

230

231 *Sensitivity analysis*

232 Direct application of the OLDW model (not refit) on the DPP showed a moderately diminished
233 discrimination (c -statistic = 0.68). There was no risk-by-treatment interaction with lifestyle ($p = 0.69$).

234 The risk-by-treatment interaction with metformin was qualitatively similar to that with the refit model
235 ($p = 0.08$), and the distribution of predicted benefits with this model was also similar. For lifestyle
236 modification, 53% of the total cases of preventable diabetes could be prevented by treating the 25% of
237 patients at highest risk; 76% by treating the 50% at highest risk. For metformin therapy, 65% of the total
238 cases of preventable diabetes could be prevented by treating the 25% of patients at highest risk; 86% by
239 treating the 50% at highest risk.

240

241 *Implementation of the final model*

242 Figure 4 shows the user interface of the SMART app in an EHR. Predictions are generated automatically
243 based on the data available and retrieved from the patient's record, using appropriate indicators in the
244 model for missingness where necessary.

245

246 **Discussion**

247 We present an EHR-compatible model to predict diabetes onset, using 11 variables routinely
248 collected in clinical practice. A major strength of the risk model is that it was derived on the OLDW,
249 which reflects people with prediabetes defined by the most commonly used ADA criteria, from
250 heterogeneous EHRs and more than 30 US healthcare systems. The risk model derived in 3 US Census
251 regions performed very well in a geographically distinct cohort. Compatible risk-specific estimates of
252 treatment effect were then obtained directly from the DPP. By prioritizing care based on the risk of
253 diabetes, this "hybrid" model might help optimize the efficiency of diabetes prevention: Treating just
254 the highest-risk half of people with prediabetes would capture 77% of the benefit of population-wide
255 lifestyle modification or 93% of the benefit of population-wide metformin pharmacotherapy. This is
256 important because lifestyle programs are resource-intensive and require a high level of commitment
257 from the patient. Pharmacotherapy is not without adverse effects, and over-treatment should be
258 avoided, especially in low-risk patients who do not appear to benefit.

259 The issue of how to address prediabetes has grown in importance as broader diabetes screening
260 has been recommended and promoted.^{13,28} For every patient with diabetes identified, screening
261 identifies 6 patients with prediabetes; health systems are thus confronted with a growing number of
262 patients who have prediabetes, without the capacity to treat everybody, reserving limited resources to
263 improving cardiometabolic control for patients with diabetes. While the ADA has lowered the A1c and
264 FG thresholds to define prediabetes,^{9,29} some have argued that the value of medicalizing prediabetes
265 and defining an ever-growing proportion of the population as diseased is of dubious value.⁵ Most
266 patients who are classified as prediabetic do not develop diabetes even in a decade, and risks of
267 developing end organ damage are low for those developing diabetes later in life.³⁰ Risk stratification
268 offers an approach that promises more focused resources specifically on those who are likely to benefit.
269 While our prior research results provided proof-of-concept that risk stratification could support
270 providers and health systems prioritize these patients,⁸ the present EHR-compatible model is designed
271 to be used at point of care, and it has been incorporated into the EHRs at several locations in the US.

272 A longstanding concern regarding limitations of randomized clinical trial results is that they
273 might not be applicable when there is non-random selection into the trial and when treatment effects
274 are heterogeneous³¹. Here, for example, we found that the “real world” at-risk population was at
275 substantially lower overall risk than patients included in the DPP, and that treatment effects were risk-
276 dependent. The lower overall risk in the OLDW cohort is the result of multiple factors, including: 1)
277 different inclusion criteria for the DPP (including a high BMI and elevated 2-hour glucose after a 75-gram
278 glucose load); 2) differences in the distribution of risk variables; 3) different outcome ascertainment,
279 which is substantially more rigorous in the trial setting. Cross-design synthesis has been proposed as a
280 means of addressing the potential problems of external validity of trial evidence by combining the
281 strengths of both designs—observational designs to capture the full range of patients and randomized
282 trials for unbiased treatment effects^{32,33}.

283 While several different methods for cross-design synthesis have been proposed^{34,35}, all
284 approaches depend on the ability to adjust results based on patient characteristics across designs. A
285 seldom-discussed barrier is that variable definitions and ascertainment can differ considerably between
286 clinical trial data and routinely collected observational data. Our approach was designed to address
287 these barriers in a pragmatic way, by estimating risk-specific treatment effects in the clinical trial using
288 the same set of variables as used in the observational risk model. This approach was driven in part by
289 our novel aim, to predict effects in patients in clinical care, based on individual patient characteristics,
290 rather than estimating average treatment effects in broad target populations.

291 A related issue that has received limited attention is how to deploy clinical prediction models in
292 an EHR. There is a proliferation of clinical prediction models; use of routinely collected EHR data to
293 automatically generate individual patient predictions is an appealing approach to disseminate these into
294 the clinic. However, most published clinical prediction models are developed on research cohorts or
295 clinical trials. Predictor variables collected in a trial are not consistently and rigorously captured in the
296 EHR. Recent work has highlighted that heterogeneity in predictor measurement across different settings
297 can substantially degrade model performance.^{21,36} Finally, use of trial or registry data cannot yield a
298 model robust to missing values in the EHR database used for clinical prediction, since the pattern of
299 missingness present across research and EHR environments is expected to differ. The usual approaches
300 addressing potential bias arising from missingness (e.g., multiple imputation) are not designed to cope
301 with missingness in variables used to generate predictions. These issues guided our decision to derive
302 separate models in the EHR and trial setting, using a common set of variables that were well ascertained
303 in both settings.

304

305 *Limitations*

306 The methods we used for “cross-walking” between the two very different types of data (trial
307 and EHR real world data) potentially introduce estimation error. Ideally, individualized treatment effects
308 would be estimated on databases that combine the advantages of these different data sources:
309 unbiased effect estimates through randomization; meticulous outcome ascertainment; consistency of
310 predictors across derivation and implementation populations, and large, heterogeneous populations.
311 Improving the quality of data collection in routine care and integrating randomized trials into routine
312 care³⁷⁻³⁹ may narrow the gap between trial and “real-world” data. Despite these limitations, we
313 obtained qualitatively consistent risk-stratified results in the DPP regardless of which risk model was
314 used: consistency of relative treatment effects of lifestyle modification across all levels of risk and
315 heterogeneous relative treatment effects with metformin, with much stronger relative effects in higher-
316 risk patients.

317

318 *Conclusion*

319 While the number of people in the US who have prediabetes and qualify for diabetes prevention
320 programs could potentially overwhelm health care systems, these patients have substantial variation in
321 their risk of developing diabetes and in their likelihood of benefiting from prevention therapies.
322 Incorporation of a tool into the EHR to support automated risk stratification of patients in routine

323 clinical care—by predicting individualized benefits—can support shared decision-making and prioritize
324 those patients who are most likely to benefit, where capacity might be limited.

325

326 **Funding/Acknowledgements**

327 This work was supported by a Patient-Centered Outcomes Research Institute (PCORI) contract (DI-
328 1604-35234). David Kent is the guarantor for this manuscript.

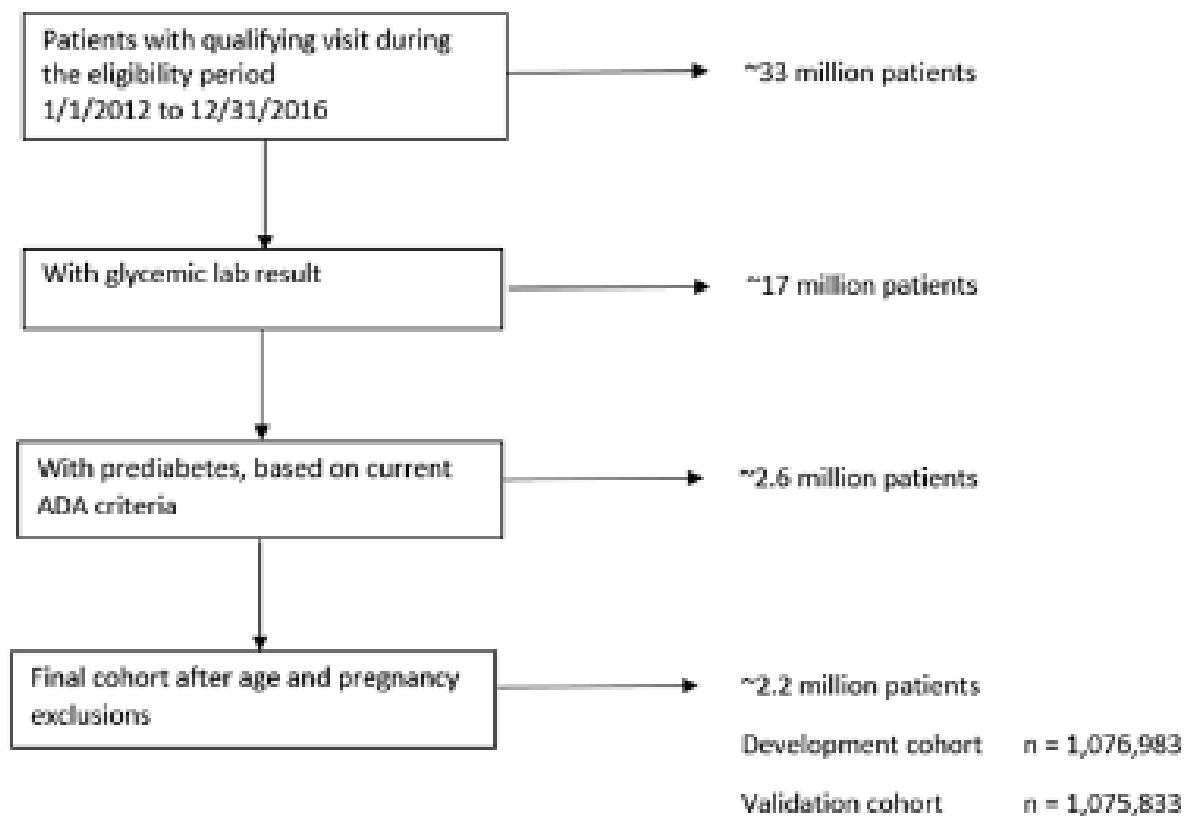


Figure 1. CONSORT Diagram for OLDW derivation and validation cohort

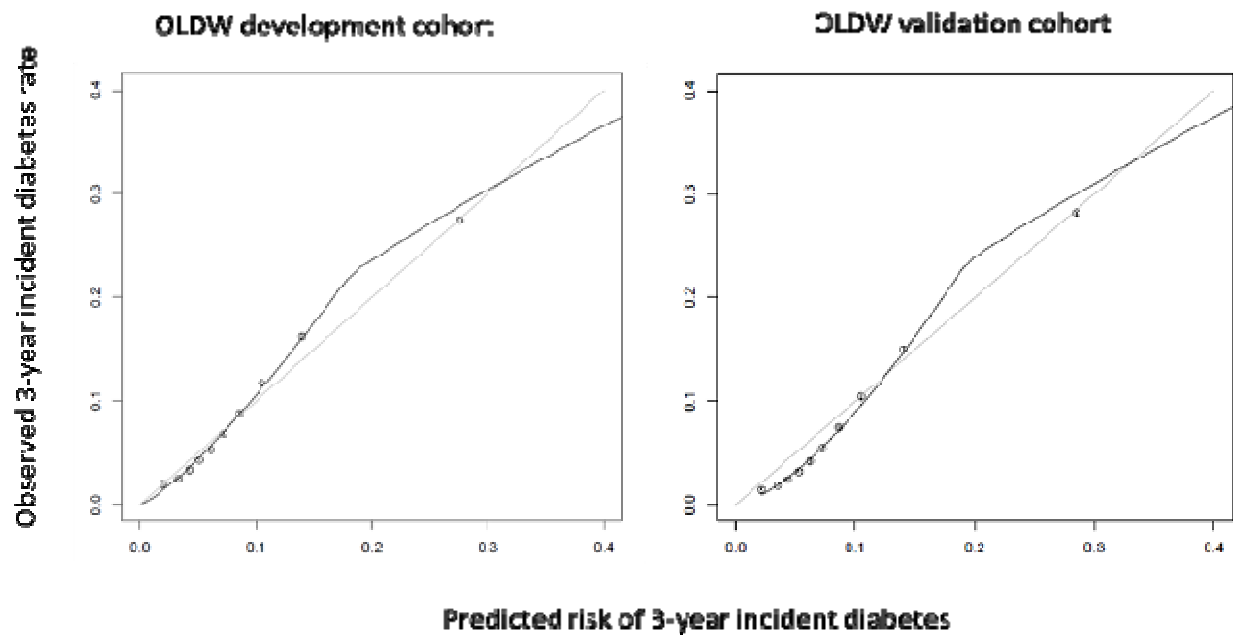


Figure 2. Calibration Curves

The figure on the left depicts the observed versus the predicted 3-year rate of developing diabetes in the 1.1 million patients in the derivation cohort (the Northeast, South and West regions) divided into equal sized tenths. The figure on the risk depicts the observed versus the predicted 3-year rate of developing diabetes in the 1.1 million patients in the validation cohort (the Midwest).

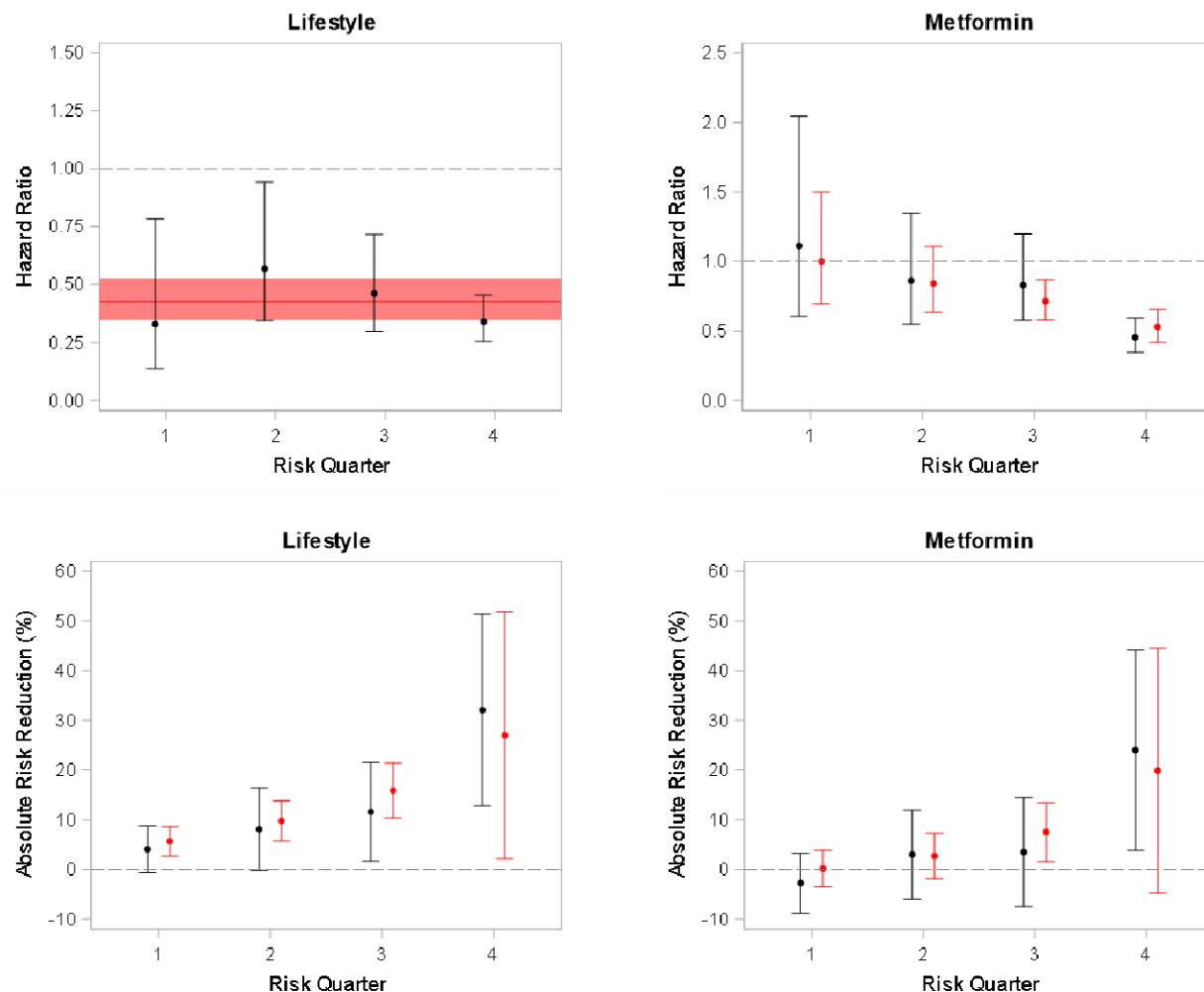


Figure 3. Observed and predicted treatment effects in the DPP Study Across Risk Groups

Black dot and bar is observed treatment effect. Red dot and bar is predicted treatment effect.

Figure 3 depicts the observed treatment effects (black dots) in patients in the DPP Study when patients are stratified into quarters based on their predicted risk for the DPP lifestyle modification intervention (left) and for metformin (right). Predicted effects across risk groups are shown in red. The top set of graphs displays relative effects and shows a consistency of effects across risk groups for lifestyle modification but heterogeneous treatment effects for metformin ($p=0.003$). The bottom graphs show effects on the absolute risk difference scale, which shows increasing benefits for higher risk patients for both interventions.

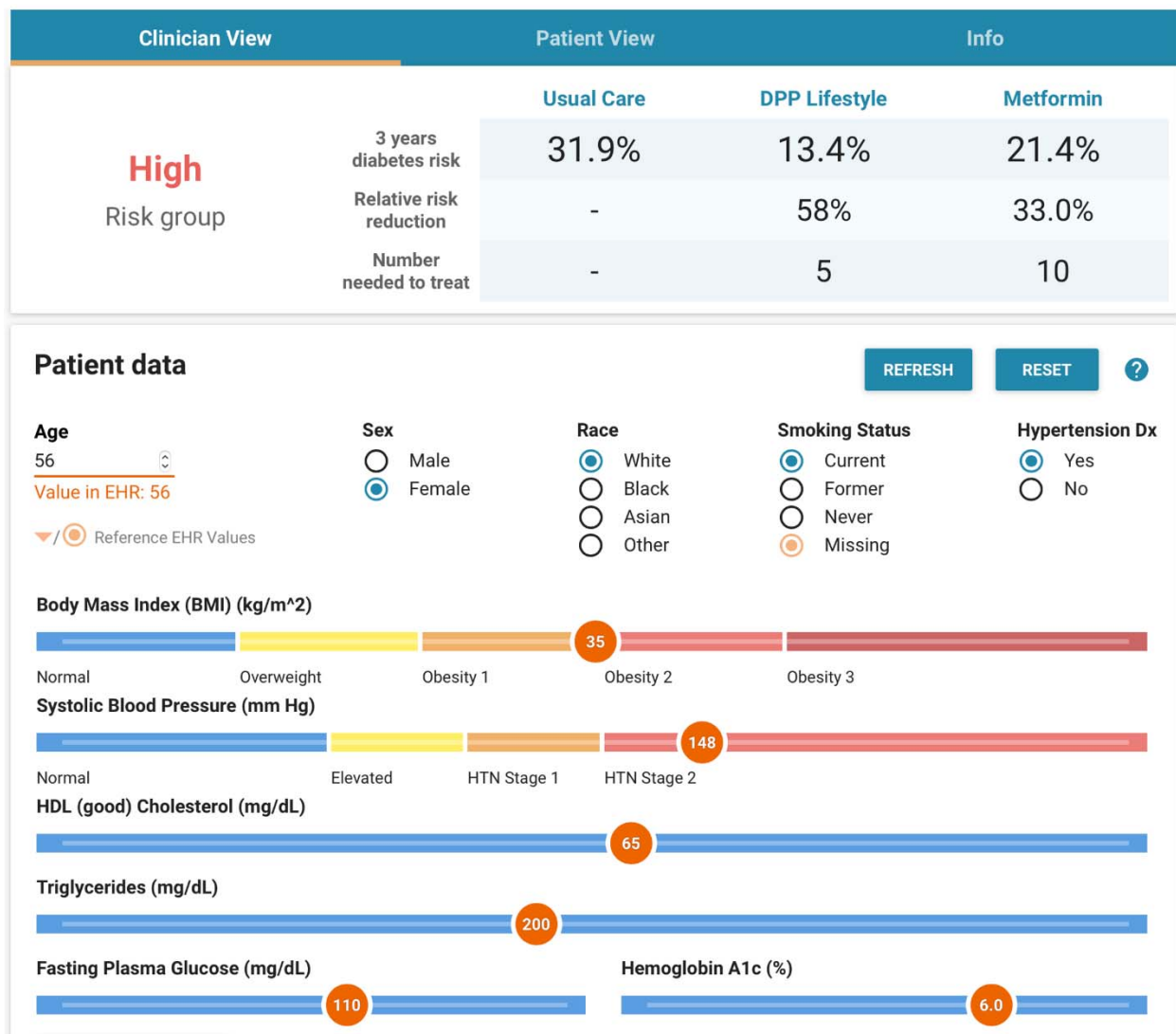


Figure 4. User interface for decision support tool

Table 1. Cohort Characteristics

	Missing	OLDW cohort			DPP
		Overall	Development	Validation	n= 3081
		n= 2,152,816	n= 1,076,983	n= 1,075,833	
Age, years, mean (SD)	0.0%	54.9 (11.7)	55.1 (11.9)	55 (11.5)	50.6 (9.0)
Female, %	0.1%	50.3	51.3	49.1	66.6
Race, %					
White	8.2%	86.5	84.3	88.9	57.4
Black	8.2%	10.2	10.8	9.1	20.9
Other nonwhite race (Optum = Asian)	8.2%	3.4	4.9	1.9	5.2
Smoking status, %	15.6%				
Current smoker		23.3	20	26.4	9.0
Never smoked		48	53.2	42.9	35.2
Former smoker		28.8	26.8	30.7	55.8
Height, cm, mean (SD)	15.9%	170.1 (10.1)	169.5 (10.1)	170.7 (10)	166.8 (9.2)
Body mass index, mean (SD)	12.2%	31.1 (7)	30.8 (6.7)	31.8 (6.9)	33.5 (5.8)
Diagnosis of hypertension, n (%)	NA	44.5	44.4	45	27.1
Systolic blood pressure, mmHg, mean (SD)	9.0%	127.4 (14.9)	127.6 (15.2)	127.3 (14.7)	124.2 (14.7)
HDL cholesterol, mg/dl, mean (SD)	12.3%	50.9 (14.7)	51.3 (14.9)	50.6 (14.5)	45.6 (11.8)
Triglycerides, mg/dl, mean (SD)	12.6%	138.3 (72.8)	136.9 (72.8)	139.7 (72.7)	162.9 (93.5)
Hemoglobin A1c, %, mean (SD)	54.7%	5.8 (0.3)	5.8 (0.3)	5.8 (0.3)	5.9 (0.5)
Fasting plasma glucose, mg/dl, mean (SD)	3.8%	103.7 (10.8)	103 (11.1)	104.5 (10.4)	107.2 (7.7)
FPG (fasting), mg/dL	86.3%	103.3 (9.2)	101.3 (10.5)	105.3 (7.3)	
FPG (random), mg/dL	13.0%	103.7 (11.4)	103.1 (11.4)	104.4 (11.2)	

Table 2. Final model for incident diabetes

	HR	Lower	Upper		HR	Lower	Upper
Age, per 10 years	1.08	1.08	1.08	<i>adjustments for missing data</i>			
Female	1.21	1.19	1.23	Race(missing) vs White	0.16	0.07	0.38
Black vs white	2.73	1.32	5.64	Smoking(missing) vs never	1.08	1.06	1.11
Asian vs white	0.01	0.00	0.02	A1c(missing)	0.75	0.74	0.77
Current smoker vs never	1.22	1.19	1.24	FPG(missing)	1.03	0.99	1.07
Former smoker vs never	1.11	1.09	1.13	Triglycerides(missing)	1.08	1.03	1.12
HTN	1.23	1.21	1.25	BMI(missing)	1.22	1.18	1.26
A1c, per 0.1%	1.24	1.18	1.29	SBP(missing)	1.22	1.17	1.26
FPG, per 10 mg/dL	1.29	1.29	1.29	HDL(missing)	1.23	1.17	1.28
Triglycerides, per 10 mg/dL	1.01	1.01	1.02	AA*BMI(missing)	0.97	0.91	1.03
BMI, per 5 units	1.24	1.24	1.24	AA*A1c(missing)	1.56	1.48	1.64
SBP, per 20 mmHG	1.05	1.05	1.05	Asian*BMI(missing)	0.77	0.70	0.84
HDL, per 10 mg/dL	0.85	0.85	0.85	Asian*A1c(missing)	2.03	1.85	2.23
Black *BMI	0.98	0.98	0.99	Race(missing)*BMI	0.99	0.99	1.00
Black *A1c	0.95	0.84	1.07	Race(missing)*BMI(missing)	0.82	0.77	0.87
Asian*BMI	1.00	0.99	1.01	Race(missing)*A1c	1.40	1.22	1.61
Asian*A1c	2.32	1.91	2.83	Race(missing)*A1c(missing)	1.46	1.37	1.55

Baseline hazard at 5 years (S0):

1 year = 0.02470

2 years = 0.04757

3 years = 0.07044

Bibliography

1. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *The New England journal of medicine*. 2002;346(6):393-403.
2. Centers for Disease Control and Prevention (CDC). Prediabetes - Your Chance to Prevent Type 2 Diabetes. <https://www.cdc.gov/diabetes/basics/prediabetes.html>. Published 2020. Accessed 16 Oct 2020.
3. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus--present and future perspectives. *Nat Rev Endocrinol*. 2011;8(4):228-236.
4. Herman WH, Zimmet P. Type 2 Diabetes: An Epidemic Requiring Global Attention and Urgent Action. *Diabetes care*. 2012;35(5):943-944.
5. Yudkin JS, Montori VM. The epidemic of pre-diabetes: the medicine and the politics. *BMJ (Clinical research ed)*. 2014;349:g4485.
6. Moin T, Li J, Duru OK, et al. Metformin Prescription for Insured Adults With Prediabetes From 2010 to 2012: A Retrospective Cohort Study. *Annals of internal medicine*. 2015;162(8):542-548.
7. Balk EM, Earley A, Raman D, Avendano EA, Pittas AG, Remington PL. Combined Diet and Physical Activity Promotion Programs to Prevent Type 2 Diabetes Among Persons at Increased Risk: A Systematic Review for the Community Preventive Services Task Force. *Ann Intern Med*. 2015;163(6):437-451.
8. Sussman JB, Kent DM, Nelson JP, Hayward RA. Improving diabetes prevention with benefit based tailored treatment: risk based reanalysis of Diabetes Prevention Program. *BMJ (Clinical research ed)*. 2015;350:h454.
9. American Diabetes Association (ADA). Diagnosis and classification of diabetes mellitus. *Diabetes care*. 2010;33:s62-69.
10. Watson J, Hutrya CA, Clancy SM, et al. Overcoming barriers to the adoption and implementation of predictive modeling and machine learning in clinical care: what can we learn from US academic medical centers? *JAMIA Open*. 2020;3(2):167-172.
11. Wallace E, Johansen ME. Clinical Prediction Rules: Challenges, Barriers, and Promise. *Annals of family medicine*. 2018;16(5):390-392.
12. Diabetes Prevention Program Research Group. Design and methods for a clinical trial in the prevention of type 2 diabetes. *Diabetes Care*. 1999;22(4):623-634.
13. American Diabetes Association (ADA). Standards of Medical Care in diabetes--2014. *Diabetes care*. 2014;37 Suppl 1:S14-80.
14. National Committee for Quality Assurance (NCQA). *National Committee for Quality Assurance (NCQA) Healthcare Effectiveness Data and Information Set (HEDIS) Comprehensive Diabetes Care*. Washington, D.C.2015.
15. Collins GS, Mallett S, Omar O, Yu LM. Developing Risk Prediction Models for Type 2 Diabetes: A Systematic Review of Methodology and Reporting. *BMC medicine*. 2011;9(103).
16. Wells BJ, Chagin KM, Nowacki AS, Kattan MW. Strategies for Handling Missing Data in Electronic Health Record Derived Data. In: *EGEMS (Wash DC)*. Vol 1. Washington, D.C.: EGEMS; 2013:1035.
17. Beck RW, Riddlesworth TD, Ruedy K, et al. Continuous Glucose Monitoring Versus Usual Care in Patients With Type 2 Diabetes Receiving Multiple Daily Insulin Injections: A Randomized Trial *Ann Intern Med*. 2017;167(6):365-374.
18. Zhu Y, Sidell MA, Arterburn D, et al. Racial/Ethnic Disparities in the Prevalence of Diabetes and Prediabetes by BMI: Patient Outcomes Research To Advance Learning (PORTAL) Multisite Cohort of Adults in the U.S. *Diabetes care*. 2019;42(12):2211-2219.

19. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer-Verlag New York; 2001.
20. Kent DM, Steyerberg EW, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ (Clinical research ed)*. 2018;363:k4245.
21. Luijken K, Groenwold RHH, Van Calster B, Steyerberg EW, van Smeden M. Impact of Predictor Measurement Heterogeneity Across Settings on the Performance of Prediction Models: A Measurement Error Perspective. *Statistics in medicine*. 2019;38(18):3444-3459.
22. Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. *Ann Intern Med*. 2020;172(1):35-45.
23. Kent DM, van Klaveren D, Paulus JK, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement: Explanation and Elaboration. *Ann Intern Med*. 2020;172(1):W1-W25.
24. Abadie A, Chingos M, West M. Endogenous Stratification in Randomized Experiments. *Review of Economics and Statistics*. 2018;100(4):567-580.
25. Burke JF, Hayward RA, Nelson JP, Kent DM. Using internally developed risk models to assess heterogeneity in treatment effects in clinical trials. *Circulation Cardiovascular quality and outcomes*. 2014;7(1):163-169.
26. van Klaveren D, Balan TA, Steyerberg EW, Kent DM. Models with interactions overestimated heterogeneity of treatment effects and were prone to treatment mistargeting. *Journal of clinical epidemiology*. 2019;114:72-83.
27. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association : JAMIA*. 2016;23(5):899-908.
28. Siu AL. Screening for Abnormal Blood Glucose and Type 2 Diabetes Mellitus: U.S. Preventive Services Task Force Recommendation Statement. *Annals of internal medicine*. 2015;163(11):861-868.
29. Genuth S, Alberti KG, Bennett P, et al. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes care*. 2003;26:3160-3170.
30. Vijan S, Sussman JB, Yudkin JS, Hayward RA. Effect of patients' risks and preferences on health gains with plasma glucose level lowering in type 2 diabetes mellitus. *JAMA internal medicine*. 2014;174(8):1227-1234.
31. Longford NT. Selection Bias and Treatment Heterogeneity in Clinical Trials. *Statistics in Medicine*. 1999;18(12):1467-1474.
32. Droitcour J, Silberman G, Chelimsky E. Cross-design Synthesis: A New Form of Meta-analysis for Combining Results from Randomized Clinical Trials and Medical-practice Databases. *International Journal of Technology Assessment in Health Care*. 1993;9(3):440-449.
33. Kaizar EE. Estimating treatment effect via simple cross design synthesis. *Statistics in Medicine*. 2011;30(25):2986-3009.
34. Cole SR, Stuart EA. Generalizing Evidence From Randomized Clinical Trials to Target Populations. *American journal of epidemiology*. 2010;172(1):107-115.
35. Varadhan R, Henderson NC, Weiss CO. Cross-design synthesis for extending the applicability of trial evidence when treatment effect is heterogeneous: Part I. Methodology. *Communications in Statistics: Case Studies, Data Analysis and Applications*. 2017;2(3-4):112-126.
36. Luijken K, Wynants L, van Smeden M, et al. Changing Predictor Measurement Procedures Affected the Performance of Prediction Models in Clinical Examples. *Journal of clinical epidemiology*. 2020;119:7-18.
37. Vickers AJ, Scardino PT. The clinically-integrated randomized trial: proposed novel method for conducting large trials at low cost. In: *Trials*. Vol 10.2009:14.

38. Simon KC, Tideman S, Hillman L, et al. Design and Implementation of Pragmatic Clinical Trials Using the Electronic Medical Record and an Adaptive Design. *JAMIA open*. 2018;1(1):99-106.
39. van Staa TP, Dyson L, McCann G, et al. The Opportunities and Challenges of Pragmatic Point-Of-Care Randomised Trials Using Routinely Collected Electronic Records: Evaluations of Two Exemplar Trials. *Health technology assessment (Winchester, England)*. 2014;18(43):1-146.