

Using polygenic risk scores for prioritising individuals at greatest need of a CVD risk assessment

Ryan Chung^{1,2}, Zhe Xu^{1,2}, Matthew Arnold^{1,2}, Samantha Ip^{1,2,3}, Hannah Harrison³, Jessica Barrett⁴, Lisa Pennells^{1,2}, Lois G. Kim^{1,2,5}, Emanuele DiAngelantonio^{1,2,5,6,7,8}, Ellie Paige^{9,10}, Scott C. Ritchie^{1,2,6,11}, Michael Inouye^{1,2,6,7,10,12}, Juliet A. Usher-Smith¹³, Angela M. Wood*^{1,2,5,6,7,14}

¹ British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

² Heart and Lung Research Institute, University of Cambridge, Cambridge, UK

³ Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

⁴ Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge, UK

⁵ National Institute for Health and Care Research Blood and Transplant Research Unit in Donor Health and Behaviour, University of Cambridge, Cambridge, UK

⁶ British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK

⁷ Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK

⁸ Health Data Science Research Centre, Human Technopole, Milan, Italy 20157

⁹ National Centre for Epidemiology and Population Health, Australian National University, Canberra, Australia

¹⁰ The George Institute for Global Health, UNSW Sydney, Australia

¹¹ Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

¹² Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia

¹³ Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

31 ¹⁴ Cambridge Centre of Artificial Intelligence in Medicine

32

33 *Corresponding author: amw79@medschl.cam.ac.uk, Cardiovascular Epidemiology Unit,

34 Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

35

36 Abstract Word count = 283

37 Manuscript Word count = 4547

38

39 4 Main tables

40 2 Main figures

41

42 **Supplementary Material**

43 16 Supplementary Tables

44 6 Supplementary Figures

45 Web Appendix 1-3

46

47 **ABSTRACT**

48

49 **Background:** To provide quantitative evidence of the use of polygenic risk scores (PRS) for
50 systematically identifying individuals for invitation for full formal cardiovascular disease
51 (CVD) risk assessment.

52
53

54 **Methods:** 108,685 participants aged 40-69, with measured biomarkers, linked primary care
55 records and genetic data in UK Biobank were used for model derivation and population
56 health modelling. Prioritisation tools using age, PRS for coronary artery disease and stroke,
57 and conventional risk factors for CVD available within longitudinal primary care records
58 were derived using sex-specific Cox models. Rescaling to account for the healthy cohort
59 effect, we modelled the implications of initiating guideline-recommended statin therapy
60 after prioritising individuals for invitation to a formal CVD risk assessment.

61

62 **Results:** 1,838 CVD events were observed over median follow up of 8.2 years. If primary
63 care records were used to prioritise individuals for formal risk assessment using age- and
64 sex-specific thresholds corresponding to 5% false negative rates then we would capture 65%
65 and 43% events amongst men and women respectively. The numbers of men and women
66 needed to be screened to prevent one CVD event (NNS) are 74 and 140 respectively. In
67 contrast, adding PRS to both prioritisation and formal assessments, and selecting thresholds
68 to capture the same number of events resulted in a NNS of 60 for men and 90 for women.

69

70 **Conclusion:** The use of PRS together with primary care records to prioritise individuals at
71 highest risk of a CVD event for a formal CVD risk assessment can more efficiently prioritise
72 those who need interventions the most than using primary care records alone. This could
73 lead to better allocation of resources by reducing the number of formal risk assessments in
74 primary care while still preventing the same number CVD events.

75

76

77

78 INTRODUCTION

79

80 Cardiovascular disease (CVD) remains a major cause of morbidity and mortality worldwide.¹
81 Identifying individuals at a high risk of CVD in order to manage and implement interventions
82 to reduce risk of CVD remains an important aim.^{2,3} Prediction tools utilising the risk factor
83 levels of individuals to estimate a 5 or 10 year risk of CVD have been developed to aid
84 clinical decision making and are recommended by healthcare guidelines across the world.^{3–}
85 ¹⁰ However, recent studies have debated the clinical value and cost effectiveness of national
86 risk assessment programmes.^{11–17} In line with this, recent guidelines have made
87 recommendations to better utilise existing primary care records to improve the
88 stratification of high-risk individuals prior to formal CVD risk assessments¹⁸. However, few
89 strategies or tools to systematically identify such individuals have been recommended.
90 Proposals have also been recommended to prioritise individuals using CVD-based polygenic
91 risk scores (PRS); such PRS have been shown to be independent of other CVD risk factors,
92 offering improved stratification with high concordance between categories of polygenic risk
93 and future CVD risk across the life course, and to improve discriminatory performance when
94 used to supplement existing CVD risk scores.^{19–21} However, no studies have quantified the
95 impact PRS would have for prioritisation.

96

97 Therefore, to investigate the benefits of PRS to systematically prioritise individuals at high
98 risk of CVD, we compare systematically prioritising individuals using a PRS based
99 prioritisation tool against current guidelines recommendations of using a prioritisation tool
100 based on longitudinal primary care records.^{22–26}

101

102

103

104 **METHODS**

105

106 **UK Biobank data source**

107 Data from UK Biobank (UKB) were used to derive each CVD risk tool (prioritisation and
108 formal assessment) and to model the implications of prioritising individuals for formal
109 assessment. UKB is a prospective cohort study with detailed baseline information, genetic
110 data and linked primary care record data available for 177,359 individuals in England
111 recruited between 2006 and 2010²⁷. Genetic data was sequenced using a genome wide
112 array of approximately 826,000 markers with imputation to approximately 96 million
113 markers.²⁷ Primary care data was provided from the The Phoenix Partnership, Egton Medical
114 Information Systems and Vision GP system suppliers²⁸. Data were linked with secondary
115 care admissions from Hospital Episode Statistics (HES) and mortality records from the Office
116 for National Statistics (ONS). For this study, primary care records were restricted to those
117 measured between the 1st April 2004, the introduction of the Quality and Outcomes
118 Framework (QOF) and UKB baseline survey. To assess the impact of PRS as a prioritisation
119 tool and compare with primary care records, analysis was restricted to individuals with
120 complete genetic data necessary for calculating the PRS, at least one primary care record
121 before baseline survey and without prior CVD or statin initiation. Individuals contributing to
122 the PRS derivation were also excluded.

123

124 **CPRD data source**

125 Data from the Clinical Practice Research Datalink (CPRD) were used for estimating the
126 average risk factor values and rescaling of 10-year risks (described in later sections.) The
127 CPRD database is a large UK primary care database containing primary care records²⁸ with
128 linked information from HES and mortality records from the ONS. The most recent 5 years
129 available primary care records were extracted for 870,486 individuals who were still alive
130 and without prior CVD on the 1st January 2014 and had no statins throughout follow-up until
131 31st May 2019, the end of data availability (**Supplementary figure 1**).

132

133 **Outcomes**

134 CVD was defined as the first ever incident of fatal or non-fatal events of coronary heart
135 disease (including angina and myocardial infarction) ischaemic heart disease and stroke
136 (code lists provided in **Supplementary table 1**), appearing in the linked HES and ONS
137 databases during follow up.

138

139 **Risk factors**

140 Two PRS for coronary artery disease (CAD) and stroke, constructed using a meta-score
141 approach and external summary statistics from large genome wide association studies^{20,29},
142 were used as independent variables. Conventional risk factors (as those in the QRISK2
143 scores⁴) were selected: age, sex, ethnicity, Townsend score, smoking status (current
144 smoker), history of diabetes (type 1 or type 2 or history of diabetes medication), family
145 history of CVD, history of chronic kidney disease (stages 4 and 5), history of atrial fibrillation
146 status, history of blood pressure treatment, history of rheumatoid arthritis, total and high
147 density lipoprotein (HDL) cholesterol, systolic blood pressure (SBP), body mass index (BMI),
148 and age interactions with Townsend score, history of diabetes, family history of CVD, history
149 of atrial fibrillation, history of blood pressure treatment, SBP and BMI.

150

151 **Statistical modelling**

152 Sex-specific Cox models were used to derive three different prioritisation tools for
153 estimating 10-year CVD prioritisation risk. First, we derived a prioritisation tool with linear
154 predictors of baseline age, CAD PRS²⁰ and stroke PRS²⁹. Age interactions were considered
155 but were not statistically significant at the 5% level. Second, we derived a prioritisation tool
156 with predictors utilising longitudinal primary care records. To handle missing values, the tool
157 was derived in two stages: in the first stage, we used sex-specific multivariate mixed effects
158 regression models on longitudinal risk factor measurements for SBP, total and HDL
159 cholesterol and BMI to estimate current risk factor values (**Web appendix 1**); in the second
160 stage, we derived sex-specific Cox models with the estimated current risk factor values for
161 SBP, total and HDL cholesterol and BMI, and the most recent primary care measurements
162 for the remaining QRISK2 risk factors. Third, we derived a prioritisation tool with both PRS
163 and primary care records, using the two-stage approach described above with the addition
164 of linear predictors for the CAD PRS and stroke PRS in the second stage Cox models. For

165 each of these three tools, the model is used to identify individuals crossing a minimum 10-
166 year risk threshold to be invited for a formal assessment.

167

168 Sex-specific Cox models were used to derive two formal risk assessment models for
169 predicting 10-year formal assessment CVD risk using risk factor measurements observed at
170 UKB baseline survey. First, we re-derived a model based on QRISK2 predictors and second,
171 we derived a model based on QRISK2 predictors enhanced with the CAD PRS and stroke PRS.

172

173 All models were validated using 10-fold cross validation and prognostic ability was
174 quantified using Harrell's C-index to measure discrimination and the net reclassification
175 improvement (NRI).

176

177 **Population Health Modelling**

178 Population health modelling was conducted to compare the population health impact of 1)
179 prioritising using a primary care records-based tool followed by a formal assessment with
180 conventional risk factors, 2) prioritising using a PRS and age-based tool followed by a formal
181 assessment with conventional risk factors and PRS and 3) prioritising using both PRS and
182 primary care records, followed by a formal assessment with conventional risk factors and
183 PRS. (**Figure 1**). Due to UKB being a cohort of healthier individuals than the UK primary care
184 population, we rescaled all estimated 10-year CVD risks so that the distribution of risks
185 estimated were using age-group- and sex- specific level risk factors obtained from CPRD and
186 the published QRISK2 score to better reflect CVD risk assessment programme in the general
187 population. (**Web appendix 2, supplementary table 2**). Details of the rescaling method has
188 been described elsewhere.^{30,31}

189

190 A hypothetical population of 100,000 individuals (50,000 men and women) from the United
191 Kingdom was created; the population age structure was obtained using data from the ONS
192 in 2015³² and the number of expected CVD events was calculated using age-group and sex-
193 specific incidence rates from CPRD (**Supplementary table 3**). A policy of statin initiation for
194 individuals at $\geq 10\%$ predicted 10-year formal assessment CVD risk as currently
195 recommended by National Institute for Health and Care Excellence (NICE) guidelines and a
196 20% reduction in CVD risk were assumed.^{33,34} The population health impact for each of the

197 three prioritisation tools was modelled using age- and sex-specific prioritisation thresholds
198 in two ways. First, we selected prioritisation thresholds to limit the formal CVD risk
199 assessment false negative rate to 5%. Second, we selected prioritisation thresholds for the
200 PRS based prioritisation tools to identify the same number of events that would have been
201 identified if prioritising with primary care records only (with prioritisation thresholds
202 corresponding to 5% false negative rates). We determined the thresholds by varying the
203 false negative rates and chose the rate that best matched the number of events identified.

204

205 Summary metrics were estimated for: the number needed to screen (NNS) to prevent one
206 CVD event, the number of CVD events identified and the number needed to invite (NNI) to
207 prevent one CVD event. We assumed 100% statin compliance and a 50% invitation uptake
208 of a formal assessment if inviting all individuals^{35,36}. We further assumed an increased
209 invitation uptake of 55% if individuals were prioritised for an invitation to a formal
210 assessment.

211

212 In sensitivity analyses, we repeated population-health analyses including all individuals,
213 including those without a primary care record for any one of SBP, HDL, total cholesterol or
214 BMI, where those without a record were all invited for formal assessment (**Supplementary**
215 **table 3**). We also repeated analyses assuming that only 50% of individuals treated with
216 statins were compliant with treatment.

217

218 Analyses were conducted in R x64 3.6.1 (R Foundation for Statistical Computing, Vienna,
219 Austria). This study follows the RECORD statement (Web Appendix 3)³⁷

220

221 RESULTS

222

223 Population characteristics

224 For our primary analysis, we identified a subset of 108,685 (60%) individuals in UKB with
225 genetic data and a primary care record for at least one of SBP, HDL, total cholesterol and
226 BMI, necessary for the in-person comparison of using either polygenic risk scores or primary
227 care records as a prioritisation method (**Supplementary figure 2**). All individuals had
228 complete information for the conventional risk factors necessary to calculate a 10-year
229 formal CVD risk at baseline survey.

230

231 The mean age at baseline was 56.2 years (SD 8.0) for men and 56.1 years (SD 7.8) for
232 women. During mean of follow-up of 8.2 years, there were 1,838 incident cardiovascular
233 events (**Table 1**). Compared to the measurements observed at the UKB baseline survey, the
234 measurements recorded in primary care records were lower for SBP and total cholesterol
235 and although similar for current smoking status and history of diabetes, were less
236 concordant for the remaining disease statuses. The mean time between the first available
237 primary care record and baseline survey was 3.8 years (95% CI: 2.3, 4.5) for men and 3.8
238 years (95% CI: 2.8, 4.7) for women.

239

240 Model performance and comparison

241 Hazard ratios (HRs) in the prioritisation tools and formal assessment models, for the same
242 predictors, were similar (**Supplementary table 4-5**). HRs for the CAD and stroke PRS were
243 higher among men than in women and remained consistent with and without the inclusion
244 of conventional risk factors.

245

246 All models had good discriminatory performance with higher performance in women. The C-
247 index for the PRS + age prioritisation tool in men (C-index = 0.663, 95% CI: 0.649, 0.678) was
248 slightly lower than the primary care records prioritisation tool (C-index = 0.684, 95% CI:
249 0.670, 0.699) (**Table 2**), however the difference in performance was greater in women. The
250 greatest discriminatory performance was observed in the model using both conventional
251 risk factors with PRS in both men (C-index = 0.716, 95% CI: 0.702, 0.730) and in women (C-
252 index = 0.742, 95% CI: 0.722, 0.762). Using conventional risk factors with PRS also improved

253 the classification of high and low risk individuals compared to using conventional risk factors
254 only in both men (NRI = 0.0262, 95% CI: 0.0072, 0.0458) and in women (NRI = 0.0265, 95%
255 CI: 0.0065, 0.0502) (**Supplementary table 6**).

256

257 The estimated 10-year risks between the primary care records only prioritisation tool and
258 the formal assessment model using conventional risk factors were highly correlated
259 (correlation coefficient = 0.75 for men and 0.80 for women). In contrast, the estimated 10-
260 year risks between the PRS + age prioritisation tool and the formal assessment model using
261 conventional risk factors and PRS were less highly correlated (correlation coefficients = 0.67
262 for men and women), and the estimated 10-year risks between the PRS and primary care
263 records based prioritisation tool and the formal assessment model using conventional risk
264 factors with PRS were more highly correlated (correlation coefficients = 0.82 for men and
265 women) (**Table 3**). Rescaled 10-year risk estimates between all models were similar
266 (**Supplementary figure 3**).

267

268 **Population health modelling**

269 In a representative population of 100,000 individuals aged 40 to 69, we estimated that there
270 would be 3,573 men and 1,808 women who would experience a CVD event over the next 10
271 years. If conventional risk factors were used as a formal risk assessment model on the whole
272 population, then 2,426 (67.9%) of those men and 801 (44.3%) of those women would have
273 been identified as being at high risk, i.e. conventional risk factors \geq 10% 10-year CVD risk
274 (**Figure 2, supplementary table 7**). Assuming 50% of individuals accepted the invitation for a
275 formal assessment and statin therapy would be initiated on all individuals with a 10-year
276 formal CVD assessment risk greater than 10%, and no other preventive interventions
277 implemented, the NNS to prevent one CVD event in men and women would be 103 (95% CI:
278 100, 107) and 312 (95% CI: 288, 334), and the NNI to prevent one CVD event in men and
279 women would be 206 (95% CI: 199, 213) and 624 (95% CI: 576, 668) respectively.

280

281 If the primary care records-based prioritisation tool, using age- and sex-specific
282 prioritisation thresholds corresponding to 5% false negative rates, was used to prioritise
283 formal assessment (with conventional risk factors) in the population, then 2,335 (65.3%)
284 men and 785 (43.4%) women with CVD events over the next 10 years would be classified at

285 high risk (**Figure 2, supplementary table 7**). The NNS to prevent one event would reduce to
286 74 (95% CI: 72, 77) in men and 140 (95% CI: 130, 150) in women (28.2% and 55.1%
287 reduction respectively). The NNI to prevent one event would be 135 (95% CI: 130, 141) in
288 men and 255 (95% CI: 235, 274) in women (34.5% and 59.1% reduction respectively)

289
290 If conventional risk factors enhanced with PRS was used as a formal risk assessment model
291 on the whole population, then 2,457 (68.8%) of those men and 844 (46.7%) of those women
292 would have been identified as being at high risk (**Figure 2, supplementary table 8**). The NNS
293 to prevent one CVD event in men and women would be 102 (95% CI: 98, 106) and 296 (95%
294 CI: 273, 315), and the NNI to prevent one CVD event in men and women would be 204 (95%
295 CI: 197, 211) and 592 (95% CI: 545, 631) respectively.

296
297 If the PRS + age prioritisation tool, using age- and sex-specific prioritisation thresholds
298 corresponding to 5% false negative rates, was used to prioritise formal risk assessment (with
299 the same conventional risk factors with PRS) in the population, then 78.8% of men and
300 74.8% of women would be prioritised and, amongst them, 2,356 (65.9%) men and 813
301 (45.0%) women with CVD events over the next 10 years would be classified at high risk
302 (**Figure 2, supplementary table 8**). This equates to a 3.7%-4.1% reduction compared to a
303 formal risk assessment using conventional risk factors with PRS on the whole population.
304 However, the NNS to prevent one event would also reduce to 84 (95% CI: 61, 66) in men
305 and 230 (95% CI: 117, 136) in women (17.8% and 22.4% reduction respectively). The NNI to
306 prevent one event would be 152 (95% CI: 146, 158) in men and 418 (95% CI: 384, 447) in
307 women (25.5% and 29.4% reduction respectively).

308
309 If the PRS and primary care records-based prioritisation tool, using age- and sex-specific
310 prioritisation thresholds corresponding to 5% false negative rates, was used to prioritise
311 formal assessment (with conventional risk factors with PRS) in the population, then 2,367
312 (66.3%) men and 825 (45.6%) women would be classified as high risk. (**Figure 2,**
313 **Supplementary table 9**). The NNS to prevent one event would reduce to 63 (95% CI: 60, 66)
314 in men and 127 (95% CI: 117, 136) in women (38.2% and 57.1% reduction respectively). The
315 NNI to prevent one event would be 115 (95% CI: 111, 120) in men and 232 (213, 248) in
316 women (43.6% and 60.8% reduction respectively).

317 Changing the prioritisation thresholds so that the total number of events identified after
318 prioritisation was 63.5% and 43.4% of all events amongst men and women respectively (the
319 total number identified when the primary care records-based prioritisation tool was used
320 with conventional risk factors), prioritising using PRS resulted in a NNS of 82 (95% CI: 79, 85)
321 in men and 223 (95% CI: 205, 240) in women and a NNI of 149 (95% CI: 143, 155) in men and
322 223 (95% CI: 205, 240) in women. Prioritising using PRS and primary care records resulted in
323 a NNS of 58 (95% CI: 55, 60) in men and 90 (95% CI: 83, 97) in women and a NNI of 105 (95%
324 CI: 101, 110) in men and 163 (95% CI: 151, 176) in women (**Table 4, supplementary figure 4**).
325 Compared to prioritising using primary care records, the reductions in the NNS and NNI
326 when prioritising using PRS and primary care records, were consistently greater in younger
327 individuals and were statistically significant at the 5% level for all age groups except in
328 women aged 40-49.

329

330 **Sensitivity Analysis**

331 In sensitivity analyses including all individuals (i.e., including 15,324 individuals without a
332 primary care record for any one of SBP, total cholesterol, HDL cholesterol or BMI)
333 (**Supplementary table 10**), we found comparable results for the PRS-based prioritisation
334 tool and the primary care-based prioritisation tool in men and women. As expected, we
335 observed an increase in the NNS in the prioritisation tools derived with primary care records,
336 especially amongst the lower age group (**Supplementary table 11-13, supplementary figure**
337 **5**).

338

339 In sensitivity analyses assuming only half of those who accepted the invitation for a formal
340 assessment and were deemed high risk complied with statin uptake, the NNS doubled for all
341 prioritisation tools, with the NNS to prevent one event increasing to 143 in men and 805 in
342 women when using the PRS based prioritisation tool, and 131 in men and 373 in women
343 when using the primary care records-based prioritisation tool (**Supplementary table 14-16,**
344 **supplementary figure 6**).

345

346 DISCUSSION

347

348 This study has rigorously assessed the impact of using PRS both alone and in combination
349 with traditional risk factors for systematically prioritising individuals for a formal CVD risk
350 assessment, and compared the efficiency and effectiveness against current
351 recommendations of using existing data on CVD risk factors within primary care records.
352 First, we found that adding PRS to both a prioritisation tool and formal CVD risk assessment
353 model improves their correlation, which subsequently leads to higher efficiency and
354 effectiveness of a prioritisation tool, especially amongst younger individuals. Consequently,
355 PRS in combination with primary care records reduces the number of men and women
356 needed to be screened to prevent one CVD event (NNS) by around 20% and 35%
357 respectively, in comparison to using primary care records alone and identifying the same
358 number of events. In contrast, using the PRS alone and in place of primary care records in a
359 prioritisation tool leads to larger NNS. These results support the use of PRS together with
360 primary care records to prioritise individuals at highest risk of a CVD event for a formal CVD
361 risk assessment, which could lead to better allocation of resources by reducing the number
362 of formal risk assessments in primary care.

363

364 This study has provided a comparison of prioritisation tools using longitudinal primary care
365 records and/or PRS within the UK population aged between 40 and 69 years who are
366 currently invited for a National Health Service (NHS) Health Check to assess their individual
367 risk of CVD. We have demonstrated the benefits of PRS not only by measuring model
368 discrimination, but also by evaluating the health impact if implemented within this
369 population. Compared with previous studies which have generally focussed on the role of
370 PRS in a formal CVD risk assessment model^{20,21,29,38}, our study has uniquely assessed its role
371 in a prioritisation tool, in conjunction with a CVD risk model. We have also shown that if PRS
372 were widely available, the inclusion of PRS in a prioritisation tool could improve the
373 effectiveness of a prioritisation tool especially in younger individuals by reducing the
374 reliance on primary care records.

375

376 The benefits in prioritising a subgroup of those individuals at low absolute risk to increase
377 efficiency echoes other studies, which have also shown that selecting a smaller proportion
378 of younger, low-risk individuals can lead to dramatically reduced costs whilst resulting in
379 more Quality Adjusted Life Years (QALYs) gained.³⁹ Using a prioritisation tool could also
380 efficiently help reduce the concerning backlog in health checks caused by the COVID-19
381 pandemic⁴⁰, where the number of people invited to health checks in England declined by
382 82% between the end of 2019 and 2020,⁴¹⁻⁴³ while still preventing nearly the same number
383 of CVD events. Whilst the addition of PRS has the potential to prioritise individuals earlier
384 for a formal CVD assessment, further extensions include using PRS to identify individuals at
385 high risk of other common chronic diseases, including diabetes, dementia and kidney
386 disease.⁴⁴⁻⁴⁸

387

388

389 **Strengths and Limitations**

390 Our study has several strengths. This study is the first of our knowledge to directly compare
391 how using different data types for a prioritisation tool can impact on the CVD risk
392 assessment programme in England. This was possible due to the unique data linkage of
393 primary care records along with a baseline survey in UK Biobank. We derived the PRS based
394 prioritisation tools using two current and well documented PRS that have been shown to
395 improve model performance independent of traditional CVD risk factors. We also took
396 advantage of the sporadically observed longitudinal primary care records when deriving the
397 primary care records-based prioritisation tools, by estimating current risk factor values using
398 a multivariate mixed model. Whilst QRISK2, which replaces missing non-recorded values
399 with age, sex and ethnicity-specific population average values, could have been used in our
400 study for the prioritisation tool, we chose to optimise the available longitudinal data in
401 primary care records to reduce possible over-inflation of the information from PRS. Another
402 strength of this study is the use of 10-fold cross validation to correct for over optimism that
403 may exist in our analyses as we derived and conducted the population health modelling in
404 the same individuals. Further, we used rescaling methods to adjust the 10-year risk
405 estimates for all of the models to minimise the healthy selection bias when deriving models
406 in UK Biobank.

407

408 However, several potential limitations exist. First, whilst we used primary care records that
409 were no more than six and a half years old before baseline, the mean risk factor levels
410 between primary care records and at the UKB baseline differed within the same individuals
411 which could lead to a different distribution of 10-year risk estimates. This may also weaken
412 the correlations between the prioritisation tool and formal risk assessment models reported.
413 Second, we determined the number of events identified in the population health modelling
414 by calculating the model's sensitivity in UKB and translating to a hypothetical population;
415 due to the low number of events in UKB, the sensitivity of each model may be limited in
416 accuracy, especially in younger age groups with fewer events. Third, PRS for cardiovascular
417 disease are still under active development and, while we utilise two extensively studied and
418 validated PRS, there are likely more powerful PRS soon to be available⁴⁹. Finally, the age-
419 range of the population health modelling was limited to between 40 and 69 years old due to
420 the use of UK Biobank. This restricts the population health modelling and in particular limits
421 the ability to investigate the early prioritisation capabilities of PRS (which are fixed at
422 conception).

423

424 **Conclusion**

425 Population health guidelines in England recommend individuals at higher estimated risk of
426 CVD be prioritised for formal risk assessment. Our results show that incorporating PRS
427 improves the correlation between prioritisation tools and formal CVD risk assessment
428 models. In particular, the use of PRS together with primary care records to prioritise
429 individuals at highest risk of a CVD event for a formal CVD risk assessment has the ability to
430 efficiently prioritise those who need interventions the most, which could lead to better
431 allocation of resources by reducing the number of formal risk assessments in primary care.

432

433 **DECLARATIONS**

434 **Acknowledgements**

435 CPRD uses data provided by patients and collected by the NHS as part of their care and
436 support. This work was supported by Health Data Research UK, which is funded by the UK
437 Medical Research Council, Engineering and Physical Sciences Research Council, Economic
438 and Social Research Council, Department of Health and Social Care (England), Chief Scientist

439 Office of the Scottish Government Health and Social Care Directorates, Health and Social
440 Care Research and Development Division (Welsh Government), Public Health Agency
441 (Northern Ireland), British Heart Foundation and Wellcome. For the purpose of open access,
442 the author has applied a Creative Commons Attribution (CC BY) licence to any Author
443 Accepted Manuscript version arising from this submission. This work was performed using
444 resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by
445 the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided
446 by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences
447 Research Council (capital grant EP/P020259/1), and DiRAC funding from the Science and
448 Technology Facilities Council (www.dirac.ac.uk).

449

450 **Funding**

451 This work was supported by core funding from the: British Heart Foundation
452 (RG/13/13/30194; RG/18/13/33946), BHF Cambridge Centre of Research Excellence
453 (RE/13/6/30180) and NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) [*].

454 *The views expressed are those of the author(s) and not necessarily those of the NIHR,
455 NHSBT or the Department of Health and Social Care.

456

457 This work was funded by the Medical Research Council (MR/K014811/1). The study funders
458 played no role in the design, analysis or interpretation of the study. R.C. is funded by a BHF
459 PhD studentship (FS/18/56/34177). Z.X. is funded by the Chinese Scholarship Council. M.A.
460 was funded by a British Heart Foundation Programme Grant (RG/18/13/33946). S.I. was
461 funded by a BHF-Turing Cardiovascular Data Science Award (BCDSA\100005) and is funded
462 by a University College London FB Cancer Research UK Award (C18081/A31373). H.H. is
463 funded by an International Alliance for Cancer Early Detection Project Award
464 (ACEDFR3_0620I135PR007). J.B. was funded by a Medical Research Council fellowship
465 (MR/L501566/1) and unit programme (MC_UU_00002/5). L.P. is funded by a British Heart
466 Foundation Programme Grant (RG/18/13/33946). L.G.K. was funded by the NIHR BTRU in
467 Donor Health and Genomics (NIHR BTRU-2014–10024) and is funded by the NIHR BTRU in
468 Donor Health and Behaviour (NIHR203337) [*]. M.I. is supported by the Munz Chair of
469 Cardiovascular Prediction and Prevention and the NIHR Cambridge Biomedical Research
470 Centre (BRC-1215-20014) and EU Horizon 2020 (No 101016775 INTERVENE). M.I. was also

471 supported by the UK Economic and Social Research 878 Council (ES/T013192/1). J.A.U. is
472 funded by an NIHR Advanced Fellowship (NIHR300861). A.M.W. is part of the
473 BigData@Heart Consortium, funded by the Innovative Medicines Initiative-2 Joint
474 Undertaking under grant agreement No 116074. A.M.W. is supported by the BHF-Turing
475 Cardiovascular Data Science Award (BCDSA\100005).

476

477 **Author contributions**

478 R.C., J.A.U. and A.M.W. designed the study. R.C., Z.X., M.A. and A.M.W. were involved in
479 data preparation. R.C., Z.X., M.A., L.P., M.I. and A.M.W were involved with the methodology.
480 R.C., Z.X. and A.M.W. conducted the statistical analysis. R.C. and A.M.W. wrote the first
481 version of the manuscript. All authors reviewed and edited the manuscript.

482

483 **Ethics approval**

484 This research has been conducted using the UK Biobank Resource under Application
485 Number 26865. Data from the Clinical Practice Research Datalink (CPRD) were obtained
486 under licence from the UK Medicines and Healthcare products Regulatory Agency (protocol
487 162RMn2).

488

489 **Data availability**

490 All data files are available from the UK Biobank and CPRD databases.

491

492 **Transparency**

493 The lead author (RC) affirms that the manuscript is an honest, accurate, and transparent
494 account of the study being reported; that no important aspects of the study have been
495 omitted; and that any discrepancies from the study as planned (and, if relevant, registered)
496 have been explained.

497

498 **Conflict of interest**

499 During the drafting of the manuscript, M.A. became an employee of AstraZeneca.

500

501

502

- 503 1. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, Ahmed M, Aksut B,
504 Alam T, Alam K, et al. Global, Regional, and National Burden of Cardiovascular Diseases for
505 10 Causes, 1990 to 2015. *J. Am. Coll. Cardiol.* [Internet]. 2017 [cited 2019 Aug 12];70:1–25.
506 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28527533>
- 507 2. Wallace ML, Ricco JA, Barrett B. Screening strategies for cardiovascular disease in
508 asymptomatic adults [Internet]. *Prim. Care - Clin. Off. Pract.* 2014 [cited 2021 Jan
509 15];41:371–397. Available from:
510 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4042912/>
- 511 3. Siontis GCM, Tzoulaki I, Siontis KC, Ioannidis JPA. Comparisons of established risk
512 prediction models for cardiovascular disease: Systematic review. *BMJ* [Internet]. 2012
513 [cited 2021 Jan 15];344. Available from: <https://pubmed.ncbi.nlm.nih.gov/22628003/>
- 514 4. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, Brindle P.
515 Predicting cardiovascular risk in England and Wales: prospective derivation and
516 validation of QRISK2. *BMJ* [Internet]. 2008 [cited 2021 Nov 23];336:1475–1482.
517 Available from: <https://pubmed.ncbi.nlm.nih.gov/18573856/>
- 518 5. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk
519 prediction algorithms to estimate future risk of cardiovascular disease: Prospective
520 cohort study. *BMJ* [Internet]. 2017 [cited 2021 Jan 18];357. Available from:
521 <http://www.bmj.com/>
- 522 6. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, Cooney MT, Corrà
523 U, Cosyns B, Deaton C, et al. 2016 European guidelines on cardiovascular disease
524 prevention in clinical practice. The Sixth Joint Task Force of the European Society of
525 Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical
526 Practice (constituted by representatives of 10 societies and by invited experts.
527 Developed with the special contribution of the European Association for
528 Cardiovascular Prevention & Rehabilitation. *G. Ital. Cardiol. (Rome)*. 2017;18:547–612.
- 529 7. Anderson TJ, Gr J, Pearson GJ, Barry AR, Couture P, Dawes M, Francis GA, Genest J,
530 Grover S, Gupta M, et al. 2016 Canadian Cardiovascular Society Guidelines for the
531 Management of Dyslipidemia for the Prevention of Cardiovascular Disease in the
532 Adult. *Can. J. Cardiol.* [Internet]. 2016 [cited 2020 May 13];32:1263–1282. Available
533 from: <http://dx.doi.org/10.1016/j.cjca.2016.07.5100828-282X/Ó>
- 534 8. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ,

- 535 Himmelfarb CD, Khera A, Lloyd-Jones D, McEvoy JW, et al. 2019 ACC/AHA Guideline
536 on the Primary Prevention of Cardiovascular Disease: A Report of the American
537 College of Cardiology/American Heart Association Task Force on Clinical Practice
538 Guidelines. *Circulation*. 2019;140:e596–e646.
- 539 9. D’Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB.
540 General cardiovascular risk profile for use in primary care: The Framingham heart
541 study. *Circulation* [Internet]. 2008 [cited 2021 Jan 19];117:743–753. Available from:
542 <https://pubmed.ncbi.nlm.nih.gov/18212285/>
- 543 10. Deanfield J, Sattar N, Simpson I, Wood D, Bradbury K, Fox K, Boon N, Winocour P,
544 Feher M, Doherty P, et al. Joint British Societies’ consensus recommendations for the
545 prevention of cardiovascular disease (JBS3) [Internet]. *Heart*. 2014 [cited 2020 Jul
546 1];100:ii1–ii67. Available from: https://heart.bmj.com/content/100/Suppl_2/ii1
- 547 11. Larsen LB, Sondergaard J, Thomsen JL, Halling A, S nderlund AL, Christensen JR,
548 Thilsing T. Step-wise approach to prevention of chronic diseases in the Danish
549 primary care sector with the use of a personal digital health profile and targeted
550 follow-up- A n assessment of attendance. *BMC Public Health* [Internet]. 2019 [cited
551 2020 Oct 20];19:1092. Available from:
552 <https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-019-7419-4>
- 553 12. Krogsb ll LT, J rgensen KJ, Gr nh j Larsen C, G tzsche PC. General health checks in
554 adults for reducing morbidity and mortality from disease: Cochrane systematic review
555 and meta-analysis. *BMJ* [Internet]. 2012 [cited 2021 Jan 17];345. Available from:
556 <http://www.bmj.com/content/345/bmj.e7191?tab=related#webextra>
- 557 13. Kypridemos C, Allen K, Hickey GL, Guzman-Castillo M, Bandosz P, Buchan I, Capewell S,
558 O’Flaherty M. Cardiovascular screening to reduce the burden from cardiovascular
559 disease: Microsimulation study to quantify policy options. *BMJ* [Internet]. 2016 [cited
560 2021 Jan 17];353. Available from: <http://dx.doi.org/10.1136/bmj.i2793>
- 561 14. Si S, Moss JR, Sullivan TR, Newton SS, Stocks NP. Effectiveness of general practice-
562 based health checks: A systematic review and meta-analysis [Internet]. *Br. J. Gen.*
563 *Pract*. 2014 [cited 2020 Oct 27];64:e47–e53. Available from:
564 <https://bjgp.org/content/64/618/e47>
- 565 15. Capewell S, McCartney M, Holland W. Invited debate: NHS Health Checks--a naked
566 emperor? *J. Public Health (Bangkok)*. [Internet]. 2015 [cited 2021 Jan 18];37:187–192.

- 567 Available from: [https://academic.oup.com/jpubhealth/article-](https://academic.oup.com/jpubhealth/article-lookup/doi/10.1093/pubmed/fdv063)
568 [lookup/doi/10.1093/pubmed/fdv063](https://academic.oup.com/jpubhealth/article-lookup/doi/10.1093/pubmed/fdv063)
- 569 16. Forster AS, Burgess C, Dodhia H, Fuller F, Miller J, McDermott L, Gulliford MC. Do
570 health checks improve risk factor detection in primary care? Matched cohort study
571 using electronic health records. *J. Public Health (Bangkok)*. [Internet]. 2016 [cited
572 2021 Jan 18];38:552–559. Available from:
573 <https://academic.oup.com/jpubhealth/article-lookup/doi/10.1093/pubmed/fdv119>
- 574 17. Robson J, Dostal I, Madurasinghe V, Sheikh A, Hull S, Boomla K, Griffiths C, Eldridge S.
575 NHS Health Check comorbidity and management: An observational matched study in
576 primary care. *Br. J. Gen. Pract.* [Internet]. 2017 [cited 2021 Jan 18];67:e86–e93.
577 Available from: <https://doi.org/10.3399/bjgp16X688837>
- 578 18. National Institute for Health and Care Excellence (NICE). Lipid modification:
579 Cardiovascular risk assessment and the modification of blood lipids for the primary
580 and secondary prevention of cardiovascular disease (clinical guideline CG181). 2014;
- 581 19. Abraham G, Havulinna AS, Bhalala OG, Byars SG, De Livera AM, Yetukuri L, Tikkanen E,
582 Perola M, Schunkert H, Sijbrands EJ, et al. Genomic prediction of coronary heart
583 disease. *Eur. Heart J.* [Internet]. 2016 [cited 2019 Aug 12];37:3267–3278. Available
584 from: <http://www.ncbi.nlm.nih.gov/pubmed/27655226>
- 585 20. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, Lai FY,
586 Kaptoge S, Brozynska M, Wang T, et al. Genomic Risk Prediction of Coronary Artery
587 Disease in 480,000 Adults: Implications for Primary Prevention. *J. Am. Coll. Cardiol.*
588 2018;72:1883–1893.
- 589 21. Sun L, Pennells L, Kaptoge S, Nelson CP, Ritchie SC, Abraham G, Arnold M, Bell S,
590 Bolton T, Burgess S, et al. Polygenic risk scores in cardiovascular risk prediction: A
591 cohort study and modelling analyses. *PLOS Med.* [Internet]. 2021 [cited 2021 Jan
592 17];18:e1003498. Available from: <https://dx.plos.org/10.1371/journal.pmed.1003498>
- 593 22. Paige E, Barrett J, Pennells L, Sweeting M, Willeit P, Angelantonio E Di, Gudnason V,
594 Nordestgaard BG, Psaty BM, Goldbourt U, et al. Use of Repeated Blood Pressure and
595 Cholesterol Measurements to Improve Cardiovascular Disease Risk Prediction: An
596 Individual-Participant-Data Meta-Analysis. *Am. J. Epidemiol.* 2017;186:899.
- 597 23. Paige E, Barrett J, Stevens D, Keogh RH, Sweeting MJ, Nazareth I, Petersen I, Wood
598 AM. Landmark Models for Optimizing the Use of Repeated Measurements of Risk

- 599 Factors in Electronic Health Records to Predict Future Disease Risk. *Am. J. Epidemiol.*
600 [Internet]. 2018 [cited 2021 Nov 23];187:1530–1538. Available from:
601 <https://academic.oup.com/aje/article/187/7/1530/4952104>
- 602 24. Arruda-Olson AM, Afzal N, Priya Mallipeddi V, Said A, Moussa Pacha H, Moon S,
603 Chaudhry AP, Scott CG, Bailey KR, Rooke TW, et al. Leveraging the Electronic Health
604 Record to Create an Automated Real-Time Prognostic Tool for Peripheral Arterial
605 Disease. *J. Am. Heart Assoc.* [Internet]. 2018 [cited 2020 May 26];7:e009680.
606 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30571601>
- 607 25. Barrett JK, Sweeting MJ, Wood AM. Dynamic Risk Prediction for Cardiovascular
608 Disease: An Illustration Using the ARIC Study. In: Handbook of Statistics. Elsevier B.V.;
609 2017. p. 47–65.
- 610 26. Sweeting MJ, Barrett JK, Thompson SG, Wood AM. The use of repeated blood
611 pressure measures for cardiovascular risk prediction: a comparison of statistical
612 models in the ARIC study. *Stat. Med.* [Internet]. 2017 [cited 2019 Dec 27];36:4514–
613 4528. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27730661>
- 614 27. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D,
615 Delaneau O, O’Connell J, et al. The UK Biobank resource with deep phenotyping and
616 genomic data. *Nature* [Internet]. 2018 [cited 2022 Jul 22];562:203–209. Available
617 from: <https://pubmed.ncbi.nlm.nih.gov/30305743/>
- 618 28. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Staa T van, Smeeth L. Data
619 Resource Profile: Clinical Practice Research Datalink (CPRD). *Int. J. Epidemiol.*
620 2015;44:827–836.
- 621 29. Abraham G, Malik R, Yonova-Doing E, Salim A, Wang T, Danesh J, Butterworth AS,
622 Howson JMM, Inouye M, Dichgans M. Genomic risk score offers predictive
623 performance comparable to clinical risk factors for ischaemic stroke. *Nat. Commun.*
624 2019 101 [Internet]. 2019 [cited 2022 Jan 17];10:1–10. Available from:
625 <https://www.nature.com/articles/s41467-019-13848-1>
- 626 30. Pennells L, Kaptoge S, Wood A, Sweeting M, Zhao X, White I, Burgess S, Willeit P,
627 Bolton T, Moons KGM, et al. Equalization of four cardiovascular risk algorithms after
628 systematic recalibration: Individual-participant meta-analysis of 86 prospective
629 studies. *Eur. Heart J.* 2019;40:621–631.
- 630 31. Kaptoge S, Pennells L, De Bacquer D, Cooney MT, Kavousi M, Stevens G, Riley LM,

- 631 Savin S, Khan T, Altay S, et al. World Health Organization cardiovascular disease risk
632 charts: revised models to estimate risk in 21 global regions. *Lancet Glob. Heal.*
633 2019;7:e1332–e1345.
- 634 32. Office for National Statistics. Estimates of the population for the UK, England and
635 Wales, Scotland and Northern Ireland [Internet]. [cited 2022 Jan 17];Available from:
636 [https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/p](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesScotlandandNorthernIreland)
637 [opulationestimates/datasets/populationestimatesforukenglandandwalesScotlandand](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesScotlandandNorthernIreland)
638 [northernireland](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesScotlandandNorthernIreland)
- 639 33. Mihaylova B, Emberson J, Blackwell L, Keech A, Simes J, Barnes EH, Voysey M, Gray A,
640 Collins R, Baigent C, et al. The effects of lowering LDL cholesterol with statin therapy
641 in people at low risk of vascular disease: Meta-analysis of individual data from 27
642 randomised trials. *Lancet* [Internet]. 2012 [cited 2021 Nov 23];380:581–590. Available
643 from: <http://www.thelancet.com/article/S0140673612603675/fulltext>
- 644 34. Cook NR, Ridker P. Further insight into the cardiovascular risk calculator: the roles of
645 statins, revascularizations, and underascertainment in the Women’s Health Study.
646 *JAMA Intern. Med.* [Internet]. 2014 [cited 2022 Feb 17];174:1964–1971. Available
647 from: <https://pubmed.ncbi.nlm.nih.gov/25285455/>
- 648 35. Patel R, Barnard S, Thompson K, Lagord C, Clegg E, Worrall R, Evans T, Carter S,
649 Flowers J, Roberts D, et al. Evaluation of the uptake and delivery of the NHS Health
650 Check programme in England, using primary care data from 9.5 million people: A
651 cross-sectional study. *BMJ Open* [Internet]. 2020 [cited 2021 Jan 14];10:42963.
652 Available from: <http://bmjopen.bmj.com/>
- 653 36. Martin A, Saunders CL, Harte E, Griffin SJ, MacLure C, Mant J, Meads C, Walter FM,
654 Usher-Smith JA. Delivery and impact of the NHS Health Check in the first 8 years: A
655 systematic review. *Br. J. Gen. Pract.* [Internet]. 2018 [cited 2020 Oct 27];68:e449–
656 e459. Available from: </pmc/articles/PMC6014431/?report=abstract>
- 657 37. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Peteresen I, Sørensen HT,
658 von Elm E, Langan SM. The REporting of studies Conducted using Observational
659 Routinely-collected health Data (RECORD) Statement. *PLOS Med.* [Internet]. 2015
660 [cited 2021 Nov 23];12:e1001885. Available from:
661 <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001885>
- 662 38. Riveros-Mckay F, Weale ME, Moore R, Selzam S, Krapohl E, Sivley RM, Tarran WA,

- 663 Sørensen P, Lachapelle AS, Griffiths JA, et al. Integrated Polygenic Tool Substantially
664 Enhances Coronary Artery Disease Prediction. *Circ. Genomic Precis. Med.* [Internet].
665 2021 [cited 2022 Aug 23];14:E003304. Available from:
666 <https://www.ahajournals.org/doi/abs/10.1161/CIRCGEN.120.003304>
- 667 39. Crossan C, Lord J, Ryan R, Nherera L, Marshall T. Cost effectiveness of case-finding
668 strategies for primary prevention of cardiovascular disease: A modelling study. *Br. J.*
669 *Gen. Pract.* [Internet]. 2017 [cited 2021 Apr 7];67:e67–e77. Available from:
670 <https://bjgp.org/content/67/654/e67>
- 671 40. Public Health England. NHS Health Check - Data - OHID [Internet]. [cited 2022 Feb
672 15];Available from: [https://fingertips.phe.org.uk/profile/nhs-health-check-](https://fingertips.phe.org.uk/profile/nhs-health-check-detailed/data#page/4/gid/1938132726/pat/159/par/K02000001/ati/15/are/E92000001/iid/91733/age/219/sex/4/cat/-1/ctp/-1/yrr/1/cid/4/tbm/1)
673 [detailed/data#page/4/gid/1938132726/pat/159/par/K02000001/ati/15/are/E9200000](https://fingertips.phe.org.uk/profile/nhs-health-check-detailed/data#page/4/gid/1938132726/pat/159/par/K02000001/ati/15/are/E92000001/iid/91733/age/219/sex/4/cat/-1/ctp/-1/yrr/1/cid/4/tbm/1)
674 [1/iid/91733/age/219/sex/4/cat/-1/ctp/-1/yrr/1/cid/4/tbm/1](https://fingertips.phe.org.uk/profile/nhs-health-check-detailed/data#page/4/gid/1938132726/pat/159/par/K02000001/ati/15/are/E92000001/iid/91733/age/219/sex/4/cat/-1/ctp/-1/yrr/1/cid/4/tbm/1)
- 675 41. Health and Social Care Committee. Clearing the backlog caused by the pandemic
676 [Internet]. [cited 2022 Feb 15];Available from:
677 <https://publications.parliament.uk/pa/cm5802/cmselect/cmhealth/599/report.html>
- 678 42. Iacobucci G. Covid-19: Government must fund extra NHS capacity to tackle backlog,
679 BMA urges. *BMJ* [Internet]. 2020 [cited 2022 Feb 15];370:m3587. Available from:
680 <https://www.bmj.com/content/370/bmj.m3587>
- 681 43. Levene LS, Seidu S, Greenhalgh T, Khunti K. Pandemic threatens primary care for long
682 term conditions. *BMJ* [Internet]. 2020 [cited 2022 Feb 15];371. Available from:
683 <https://www.bmj.com/content/371/bmj.m3793>
- 684 44. Padilla-Martínez F, Collin F, Kwasniewski M, Kretowski A. Systematic Review of
685 Polygenic Risk Scores for Type 1 and Type 2 Diabetes. *Int. J. Mol. Sci.* [Internet]. 2020
686 [cited 2022 Aug 23];21. Available from: [/pmc/articles/PMC7084489/](https://pubmed.ncbi.nlm.nih.gov/3484489/)
- 687 45. Najar J, van der Lee SJ, Joas E, Wetterberg H, Hardy J, Guerreiro R, Bras J, Waern M,
688 Kern S, Zetterberg H, et al. Polygenic risk scores for Alzheimer’s disease are related to
689 dementia risk in APOE ϵ 4 negatives. *Alzheimer’s Dement. Diagnosis, Assess. Dis.*
690 *Monit.* [Internet]. 2021 [cited 2022 Aug 23];13. Available from:
691 [/pmc/articles/PMC7821873/](https://pubmed.ncbi.nlm.nih.gov/3484489/)
- 692 46. Chaudhury S, Brookes KJ, Patel T, Fallows A, Guetta-Baranes T, Turton JC, Guerreiro R,
693 Bras J, Hardy J, Francis PT, et al. Alzheimer’s disease polygenic risk score as a
694 predictor of conversion from mild-cognitive impairment. *Transl. Psychiatry* 2019 91

- 695 [Internet]. 2019 [cited 2022 Aug 23];9:1–7. Available from:
696 <https://www.nature.com/articles/s41398-019-0485-7>
- 697 47. Leonenko G, Baker E, Stevenson-Hoare J, Sierksma A, Fiers M, Williams J, de Strooper
698 B, Escott-Price V. Identifying individuals with high risk of Alzheimer’s disease using
699 polygenic risk scores. *Nat. Commun.* 2021 121 [Internet]. 2021 [cited 2022 Aug
700 23];12:1–10. Available from: <https://www.nature.com/articles/s41467-021-24082-z>
- 701 48. Hirohama D, Susztak K. From mapping kidney function to mechanism and prediction.
702 *Nat. Rev. Nephrol.* 2021 182 [Internet]. 2021 [cited 2022 Aug 23];18:76–77. Available
703 from: <https://www.nature.com/articles/s41581-021-00512-5>
- 704 49. Mishra A, Malik R, Hachiya T, Jürgenson T, Namba S, Posner DC, Kamanu FK, Koido M,
705 Le Grand Q, Shi M, et al. Stroke genetics informs drug discovery and risk prediction
706 across ancestries. *Nat.* 2022 [Internet]. 2022 [cited 2022 Oct 6];1–15. Available from:
707 <https://www.nature.com/articles/s41586-022-05165-3>
708

Table 1: Key characteristics of individuals in UK Biobank baseline survey and linked primary care records

Characteristic	Men, N = 44,184 (41%)		Women, N = 64,501 (59%)	
CVD events, N	1230		608	
Follow up duration: years, median (5 th , 95 th percentile)	8.1 (6.1, 10.8)		8.2 (6.8, 10.9)	
Duration between first primary care record and baseline visit: years, median (5 th , 95 th percentile)	3.6 (0.9, 5.5)		3.9 (1.1, 5.6)	
	Primary care records	Baseline	Primary care records	Baseline
Age, mean (SD) ^a	-	56.2 (8.0)	-	56.1 (7.8)
Coronary artery disease PRS, mean (SD)	-	-1.15 (0.46)	-	-1.13 (0.46)
Stroke PRS, mean (SD)	-	1.55 (0.22)	-	1.56 (0.23)
Ethnicity — White, N (%) ^a	-	42,283 (95.7%)	-	61,977 (96.1%)
Townsend, mean (SD) ^a	-	-1.5 (3.0)	-	-1.5 (2.9)
Systolic blood pressure: mmHg, mean (SD) ^b	135.3 (7.82)	141.0 (17.3)	130.7 (9.62)	134.9 (19.1)
Number of historical records, mean	3.8	-	4.6	-
Total cholesterol: mmol/litre, mean (SD) ^b	5.48 (0.47)	5.79 (1.01)	5.71 (0.50)	6.03 (1.08)
Number of historical records, mean	2.0	-	2.0	-
HDL cholesterol: mmol/litre, mean (SD) ^b	1.35 (0.19)	1.30 (0.31)	1.68 (0.24)	1.61 (0.37)
Number of historical records, mean	1.8	-	1.9	-
BMI: kg/m ² , mean (SD) ^b	27.2 (3.1)	27.5 (4.1)	26.6 (4.1)	26.8 (5.0)
Number of historical records, mean	2.0	-	2.3	-
Current smoker, N (%)	4,472 (10.1%)	5,233 (11.8%)	4,911 (7.61%)	5,511 (8.5%)
History of diabetes, N (%)	466 (1.05%)	630 (1.4%)	412 (0.64%)	459 (0.7%)
Blood pressure-lowering medication prescriptions, N (%)	6,396 (14.5%)	5,529 (12.51%)	9,737 (15.1%)	7,643 (11.85%)
Family history, N (%) ^a	1,568 (3.55%)	-	2,494 (3.87%)	-
Chronic kidney disease (4/5), N (%) ^a	57 (0.13%)	-	79 (0.12%)	-
Rheumatoid arthritis, N (%)	146 (0.33%)	381 (0.86%)	336 (0.52%)	989 (1.53%)
Atrial fibrillation, N (%)	336 (0.33%)	123 (0.28%)	1,749 (2.7%)	89 (0.14%)

Abbreviations: BMI, body mass index; CVD, cardiovascular disease; HDL cholesterol, high-density lipoprotein cholesterol; PRS, polygenic risk score; SD, standard deviation.

^a Risk factor values in both baseline and primary care records if one was missing.

^b Risk factor values for primary care records estimated using multivariate mixed effects model.

Table 2: C indices of prioritisation tools and formal CVD risk assessment tools in UK Biobank

Model	C-index (95% confidence interval)		
	All individuals	Men	Women
Prioritisation tool			
Primary care records only	0.730 (0.719, 0.741)	0.684 (0.670, 0.699)	0.734 (0.715, 0.754)
PRS + age	0.663 (0.652, 0.675)	0.663 (0.649, 0.678)	0.686 (0.665, 0.707)
PRS + primary care records	0.740 (0.730, 0.751)	0.704 (0.691, 0.718)	0.738 (0.718, 0.758)
Formal risk assessment tool			
Conventional risk factors	0.730 (0.719, 0.740)	0.700 (0.686, 0.714)	0.739 (0.720, 0.759)
Conventional risk factors +PRS	0.738 (0.727, 0.749)	0.716 (0.702, 0.730)	0.742 (0.722, 0.762)

Abbreviations: PRS, polygenic risk score.

C indices and 95% confidence intervals from each model for the prediction of 10-year cardiovascular disease by sex and for the combined population in UK Biobank after 10-fold cross validation.

Table 3: Correlation of predicted 10-year risks between prioritisation tools and formal assessment tools by sex in the derivation dataset

Men	Primary care records only prioritisation tool	PRS + age prioritisation tool	PRS + primary care records prioritisation tool
Conventional risk factor formal assessment tool	0.75	-	-
Conventional risk factor + PRS formal assessment tool	-	0.67	0.82
Women			
Conventional risk factor formal assessment tool	0.80	-	-
Conventional risk factor + PRS formal assessment tool	-	0.67	0.82

Abbreviations: PRS, polygenic risk score.

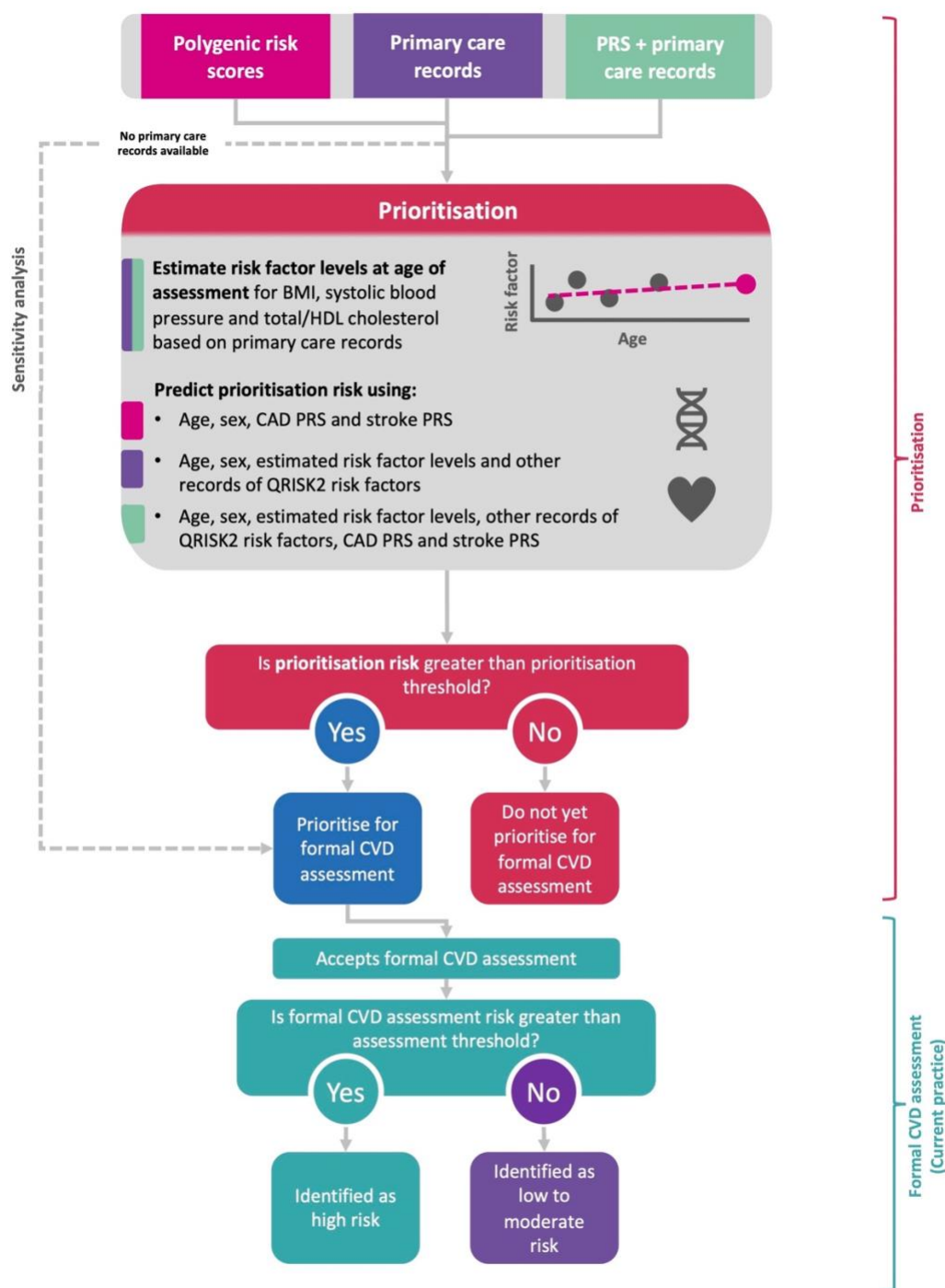
Table 4: Number needed to invite and screen to prevent one event and number of events identified after prioritisation and formal assessment in a hypothetical population of 100,000 individuals in England, with prioritisation thresholds selected to identify the same number of events if prioritising with primary care records with prioritisation thresholds controlling the false negative rate to 5%.

		Prioritisation using primary care records followed by conventional risk factors				Prioritisation using PRS + age, followed by conventional risk factors + PRS				Prioritisation using PRS and primary care records, followed by conventional risk factors + PRS			
Age group	Participants	Participants prioritised (%)	NNI (95% CI)	NNS (95% CI)	Number of events identified as high risk (%)	Participants prioritised (%)	NNI (95% CI)	NNS (95% CI)	Number of events identified as high risk (%)	Participants prioritised (%)	NNI (95% CI)	NNS (95% CI)	Number of events identified as high risk (%)
Men													
40-49	18253	10126 (55.5%)	765 (498.9, 932.8)	421 (274.4, 513.0)	120 (24.8%)	13016 (71.3%)	727 (530.4, 856.5)	400 (291.7, 471.1)	163 (33.6%)	7208 (39.5%)	369 (265.8, 438.4)	203 (146.2, 241.1)	163 (33.6%)
50-59	17391	12134 (69.8%)	170 (150.6, 184.2)	93 (82.8, 101.3)	651 (52.5%)	13100 (75.3%)	179 (162.6, 193.3)	98 (89.4, 106.3)	666 (53.7%)	9878 (56.8%)	133 (118.4, 144.4)	73 (65.1, 79.4)	654 (52.7%)
60-69	14356	12517 (87.2%)	73 (70.4, 74.9)	40 (38.7, 41.2)	1564 (84.7%)	12196 (84.9%)	74 (70.8, 76.1)	40 (39.0, 41.9)	1506 (81.5%)	11064 (77.1%)	65 (62.1, 66.9)	36 (34.2, 36.8)	1519 (82.3%)
Total	50000	34777 (69.6%)	135 (130.4, 140.5)	74 (71.7, 77.3)	2335 (65.3%)	38313 (76.6%)	149 (143.0, 154.9)	82 (78.7, 85.2)	2335 (65.4%)	28150 (56.3%)	105 (100.5, 109.5)	58 (55.3, 60.2)	2336 (65.4%)
Women													
40-49	18107	3233 (17.9%)	1092 (0.0, 1596.1)	601 (0.0, 877.9)	27 (10.0%)	10139 (56.0%)	3550 (0.0, 5198.6)	1952 (0.0, 2859.2)	27 (10.0%)	1748 (9.7%)	506 (0.0, 731.1)	279 (0.0, 402.1)	31 (11.7%)
50-59	17282	8329 (48.2%)	572 (389.4, 701.7)	314 (214.2, 385.9)	132 (23.0%)	12436 (72.0%)	742 (507.0, 891.7)	408 (278.8, 490.4)	156 (27.1%)	4683 (27.1%)	306 (198.7, 373.0)	168 (109.3, 205.2)	139 (24.1%)
60-69	14611	10459 (71.6%)	152 (138.7, 162.1)	84 (76.3, 89.1)	626 (65.1%)	11357 (77.7%)	178 (163.2, 189.7)	98 (89.8, 104.3)	603 (62.7%)	7714 (52.8%)	114 (105.2, 121.2)	63 (57.9, 66.7)	616 (64.0%)
Total	50000	22021 (44.0%)	255 (235.4, 273.5)	140 (129.5, 150.4)	785 (43.4%)	33932 (67.9%)	406 (372.3, 436.0)	223 (204.8, 239.8)	786 (43.5%)	14145 (28.3%)	163 (150.5, 175.5)	90 (82.8, 96.5)	786 (43.5%)

Abbreviations: NNS, number needed to screen; PRS, polygenic risk score.

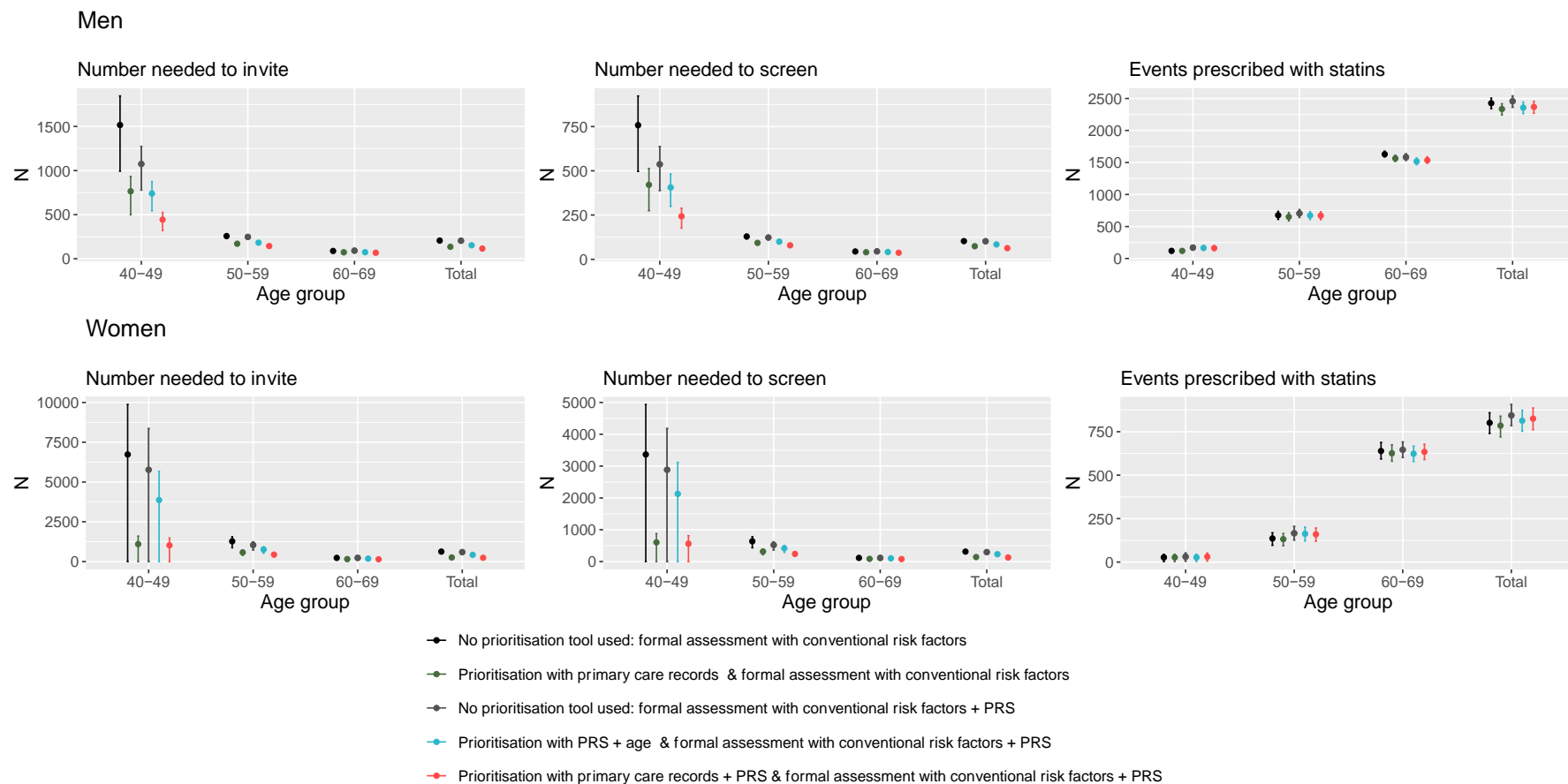
Age structure of hypothetical population extrapolated from Office for National Statistics, England, United Kingdom 2015. Expected events at 10 years based on extrapolation of incidence rates from CPRD, 2014-2019. Age group and sex specific prioritisation thresholds when prioritising using primary care records were defined as the level such that the expected false negative rate is controlled to be 5%. Age group and sex specific prioritisation thresholds when prioritising using PRS or PRS and primary care records tool chose thresholds that resulted in a similar number of events identified if prioritising using primary care records. Prioritisation thresholds for the PRS prioritisation tool was equivalent to a 5.9% and 6.5% false negative rate for men and women respectively, and for the PRS and primary care records prioritisation tool was equivalent to an 8.1% and 13.0% false negative rate respectively. NNI and NNS assumes 100% statin compliance. NNI assumes a 50% invitation uptake if assessing without using prioritisation tool, and a 55% invitation uptake if assessing with using prioritisation tool.

Figure 1: Flow chart of the implementation of a prioritisation tool for formal cardiovascular disease assessments



Abbreviations: BMI, body mass index; CAD, coronary artery disease; CVD; cardiovascular disease; HDL, high density lipoprotein; PRS, polygenic risk score.

Figure 2: Number needed to invite, number needed to screen and number of events identified after prioritising for a formal CVD assessment, in a hypothetical population of 100,000 individuals in England.



Abbreviations: NNS, number needed to screen; NNI, number needed to invite; PRS, polygenic risk score.

95% confidence intervals are represented by vertical lines. Age group and sex specific prioritisation thresholds were defined as the level such that the expected false negative rate is controlled to be 5%. NNI and NNS assumes 100% statin compliance, and half of all individuals invited for formal assessment attend.