

Development and validation of a tool to stratify the treatment effect of low-dose aspirin in patients with cardiovascular disease: VISTA (Vascular Intervention Stratification Tool for Aspirin)

Authors

Yekai Zhou BSc^{1*}, Joseph Edgar Blais BScPharm, PhD^{2*}, Esther Wai Yin Chan BPharm(Hons), PhD^{2,3}, Tak-Wah Lam PhD¹, Kai Hang Yiu MD, PhD⁴, Hung Fat Tse MD, PhD⁴, Celine SL Chui PhD^{3,5,6#}, Ruibang Luo PhD^{1#}

***Co-first authors**

#Corresponding authors

¹ Department of Computer Science, The University of Hong Kong, Hong Kong Special Administration Region, China

² Centre for Safe Medication Practice and Research, Department of Pharmacology and Pharmacy, The University of Hong Kong, Hong Kong Special Administrative Region, China

³ Laboratory of Data Discovery for Health (D24H), Hong Kong Science Park, Hong Kong Science and Technology Park, Hong Kong Special Administrative Region, China

⁴ Department of Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administration Region, China

⁵ School of Nursing, The University of Hong Kong, 5/F Academic Building, 3 Sassoon Road, Pokfulam, Hong Kong Special Administrative Region, China

⁶ School of Public Health, The University of Hong Kong, Hong Kong Special Administrative Region, China

Short Title: Predicting low-dose aspirin treatment effect

Corresponding Authors

Prof Celine Sze Ling Chui, 5/F Academic Building, 3 Sassoon Road, Pokfulam, Hong Kong Special Administration Region, China. Email: cslchui@hku.hk

Prof Ruibang Luo, Rm 301 Chow Yei Ching Building, The University of Hong Kong, Pokfulam Road, Pokfulam, Hong Kong Special Administration Region, China. Email: rbluo@cs.hku.hk

Manuscript word count: 3911

Abstract

Background

Aspirin resistance, as determined by measuring platelet aggregation function, has been investigated as a proxy outcome for clinical treatment failure with low-dose aspirin (occurrence of cardiovascular events despite regular aspirin intake). However, among adults with cardiovascular disease (CVD), there is no method to predict aspirin treatment failure using routinely available clinical data. We aimed to develop and internally validate the Vascular Intervention Stratification Tool for Aspirin (VISTA), a model that predicts inter-individual variability in clinical response to low-dose aspirin.

Methods

We used electronic health records of the Hong Kong Hospital Authority to identify derivation (n=48,743) and validation (n=322,731) cohorts consisting of individuals diagnosed with CVD between January 1, 2015 to December 31, 2018. The composite outcome of recurrent CVD event included the diagnosis of coronary heart disease, ischemic stroke, and peripheral artery disease after low-dose aspirin initiation (≤ 100 mg). Cox proportional hazards regression with the least absolute shrinkage and selection operator regularization was used to identify the most strongly associated and relevant risk factors for aspirin treatment failure. One-year hazard ratio (HR) was estimated across different risk categories for aspirin treated vs. untreated individuals.

Results

The derivation cohort included 1,623 individuals who initiated low-dose aspirin after their CVD diagnosis. Among 109 variables available, six were selected as model inputs: atrial fibrillation, dyslipidemia, hyperglycemia, polypharmacy, neutrophilia, and elevated serum creatine kinase. In the model validation cohort, we identified 22,192 individuals who initiated low-dose aspirin and 3,747 individuals without aspirin, other antiplatelets, or anticoagulants. Results of the model validation demonstrated a strong graded association between the number of VISTA risk factors and the one-year risk of CVD. Compared to untreated individuals, low-dose aspirin use with no VISTA risk factors had the lowest HR for CVD (0.68, 95% CI of 0.57 to 0.81). For low-dose aspirin user with 1-2 VISTA risk factors, HRs was 0.87 (0.81 to 0.93). The presence of 3-6 VISTA risk factors was associated with aspirin treatment failure (HR 0.99; 95% CI of 0.88 to 1.12), which occurred in approximately 20% of patients in our validation cohort.

Conclusions

VISTA can predict the heterogeneity of low-dose aspirin's treatment effect against recurrent CVD. VISTA could be used to stratify patients based on six readily available risk factors and inform patients and clinicians about the potential benefits of aspirin therapy and the potential for alternate antiplatelet treatments.

Keywords: Low-dose aspirin, aspirin treatment failure, aspirin resistance, stratified medicine, heterogeneity of treatment effect, secondary prevention, risk prediction model, electronic health records

Clinical perspective

What is new?

- We developed VISTA (Vascular Intervention Stratification Tool for Aspirin), the first tool that enables the stratification of low-dose aspirin treatment effect for the secondary prevention of cardiovascular disease (CVD). By utilizing six easily accessible clinical risk factors (atrial fibrillation, dyslipidemia, hyperglycemia, polypharmacy, neutrophilia, and elevated serum creatine kinase), VISTA allows for the assessment of aspirin suitability before prescription.
- VISTA can differentiate patients who are likely to benefit significantly from aspirin treatment (0 risk factors, hazard ratio [HR] of 0.68) from those who may experience aspirin treatment failure (3-6 risk factors, HR of 0.99).

What are the clinical implications?

- Estimated from a high-quality contemporary validation cohort, there are approximately 20% of patients with CVD who may experience aspirin treatment failure.
- By utilizing VISTA, healthcare providers can personalize aspirin treatment, optimizing its effectiveness and minimizing the potential for treatment failure. This tool empowers clinicians to make more accurate and tailored decisions in prescribing low-dose aspirin for secondary prevention of cardiovascular disease, ultimately leading to improved patient outcomes.

1 Introduction

2

3 Cardiovascular diseases (CVD), such as coronary heart disease and stroke, are the leading
4 cause of death worldwide, resulting in an estimated 18.6 million fatalities in 2019.^{1,2} Aspirin,
5 or acetylsalicylic acid, has long been recognized as a crucial intervention to improve CVD
6 outcomes due to its antiplatelet properties, particularly in secondary prevention of CVD.^{3,4}
7 However, the occurrence of occlusive CVD events despite regular aspirin intake, a
8 phenomenon known as aspirin treatment failure, poses a significant clinical challenge.⁵⁻¹⁰
9 Identifying patients who may have limited benefit from CVD prevention with aspirin but are
10 exposed to an elevated risk of major organ bleeding is essential for appropriate aspirin
11 prescription and selection of suitable antiplatelet treatments for CVD prevention.¹¹
12 Although methods exist to indirectly query aspirin effectiveness via platelet aggregation
13 function tests, there are currently no solutions available to predict aspirin treatment failure
14 in clinical settings.^{5,12-15}

15

16 Existing methods for evaluating aspirin effectiveness involve assays that test a laboratory
17 phenomenon called aspirin resistance.⁵ These assays, including the platelet function
18 analyzer (PFA-100), whole blood aggregometry (WBA), and VerifyNow aspirin assay
19 (VerifyNow), quantify and compare the platelet aggregation function before and after the
20 administration of aspirin.^{5,12-15} However, these assays have limitations when it comes to
21 aspirin prescriptions. First, they serve only as risk indicators, similar to age and lipid levels;
22 they cannot measure the treatment effect of aspirin.^{16,17} Second, these assays utilize
23 different techniques, resulting in poor agreement even among the same group of
24 patients.¹⁸⁻²⁵ This lack of consistency adds to the difficulty for clinicians in selecting the most
25 appropriate assay. Third, these assays are not endorsed by clinical guidelines and are not
26 typically available in general hospital settings, limiting their use as routine tests for aspirin
27 prescription reference.^{26,27} Therefore, there is an urgent need for a method based on easily
28 accessible patient characteristics that can directly estimate the treatment effect of aspirin in
29 CVD prevention.

30

31 In this study, we present the development and validation of the Vascular Intervention
32 Stratification Tool for Aspirin (VISTA), which was designed to stratify the treatment effect of
33 aspirin for the prevention of recurrent CVD events. VISTA is based on six easily accessible
34 patient characteristics that can be assessed prior to aspirin prescription in actual clinical
35 settings. We conducted a large, comprehensive cohort study in Hong Kong, encompassing a
36 wide range of potential risk factors. By employing advanced big data approaches, we
37 analyzed this extensive dataset to identify the most significant risk factors contributing to
38 the variation in aspirin treatment effects. The validity of the identified risk factors was then
39 confirmed in a separate validation cohort, which demonstrated their accuracy in stratifying
40 the treatment effects of aspirin. Our method provides a straightforward estimation of
41 aspirin's treatment effect, serving as a tool to assess its potential benefits for individual
42 patients. Leveraging VISTA's capabilities, clinicians gain early access to this information,
43 enabling informed decisions about aspirin therapy suitability.

44

45 Methods

46 *Data source*

47 Two contemporary cohorts were identified for model derivation and validation from the
48 electronic health records of the Hong Kong Hospital Authority (HA) with rigorous inclusion
49 criteria (Figure 1). The HA is a statutory body and the largest public healthcare provider of
50 Hong Kong. It provides government subsidized primary, secondary, and tertiary care for all
51 residents who want it, capturing over 70% of all hospitalizations in Hong Kong.²⁸ Previous
52 studies demonstrated the high validity of the data source, with a positive predictive value of
53 85% for myocardial infarction (MI) and 91% for stroke.²⁹ Because patients are provided
54 outpatient medications at very low out-of-pocket cost, dispensing records for low-dose
55 aspirin are recorded in the HA database. This makes it suitable to investigate the chronic
56 effects of low-dose aspirin, which may not be feasible in many other health systems where
57 patients purchase aspirin as an over-the-counter medication without prescription or
58 recording.³⁰

59

60 Study cohorts and eligibility criteria

61 We included patients who 1) had their first CVD diagnosis recorded between 1 January 2015
62 and 31 December 2018, 2) did not have an aspirin prescription dispensed before the index
63 CVD event, and 3) had a prescription record for low-dose (≤ 100 mg) aspirin daily until the
64 end of study follow-up, or did not have any prescription record for aspirin at any dose until
65 the end point of their study (as the control group to compare with the treated patients to
66 calculate the treatment effect of aspirin). The cohort entry date was the date of their first
67 diagnosis of CVD in any inpatient or outpatient setting. Patients were followed until the
68 earliest occurrence of the composite recurrent CVD outcome, date of registered death, one
69 year after their index CVD diagnosis, or the study end date (31 December 2019). We did not
70 include patients with CVD diagnosis within one year before the study end date to ensure the
71 patients would have sufficient exposure to aspirin therapy. Daily aspirin users whose most
72 frequently visited healthcare facility within the study period was on Hong Kong Island were
73 categorized as the Hong Kong Island cohort for model derivation. Patients with their most
74 frequently visited healthcare facility within the study period was on Kowloon and New
75 Territories were categorized as the Kowloon and New Territories cohort for model
76 validation. The daily aspirin users were in the treatment group and the non-aspirin users in
77 the untreated (control) group.

78

79 *Exposure, outcomes and risk factor variables*

80 The included patients had a CVD diagnosis (coronary heart disease, ischemic stroke, and
81 peripheral artery disease) between January 1, 2015, and December 31, 2018. Patients were
82 not prescribed low-dose (≤ 100 mg) aspirin before the CVD event. The control group
83 patients were not treated with low-dose aspirin before their censored date. The control
84 group patients received sufficient daily low-dose aspirin treatment from the date of CVD
85 event to their censored date.

86

87 The outcome was the diagnosis of CVD defined by the International Classification of
88 Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes. The outcome was a
89 composite of coronary heart disease, ischaemic or haemorrhagic stroke, and peripheral
90 artery disease (Supplementary Table 1).

91

92 We compiled an extensive list of over 100 variables to ensure the selection of
93 representative risk factors. The list encompassed clinical laboratory tests, disease and

94 medication history, family history of disease, demographic factors, and healthcare utilization
95 (Table 2). Diagnoses and procedures were defined by ICD-9-CM codes (Supplementary Table
96 2), and medication history was defined by the British National Formulary (BNF) sections
97 (Supplementary Table 3).

98

99 *Statistical analysis*

100 A robust feature selection pipeline was applied to identify the risk variables for model
101 derivation.^{31,32} Multivariate imputation with chained equations (MICE) was used to generate
102 one imputed dataset to replace the missing values of clinical laboratory tests.³³ MICE is a
103 principled method for dealing with missing data and is extremely reliable on high-
104 dimensional datasets with various missing patterns.³⁴ For better statistical reliability and
105 clinical utility, risk variables with missing rates below 20% (e.g., clinical laboratory tests) and
106 an event rate above 1% (e.g., disease and medication history) were passed for feature
107 selection. We employed a Cox proportional hazards model (CPH) with the least absolute
108 shrinkage and selection operator (LASSO) regularization to shortlist statistically significant (p
109 value<0.05) risk variables.^{35,36} CPH is the most widely used multivariate statistical model for
110 survival analysis.^{37,38} Its regression coefficients can be interpreted as hazard ratios which can
111 be easily understood by clinicians for better decision-making. LASSO is a robust feature
112 selection method. It selects the most representative but independent set of risk variables,
113 which is reliable when downstream manual prioritization is required. The final list of risk
114 variables was also determined based on current clinical evidence to ensure the final set of
115 risk variables are comprehensive and relevant to CVD prognosis. The identified continuous
116 risk variables were converted to categorical form based on clinical diagnosis criteria. Each
117 risk variable was weighted as one point, and the risk category was defined as the total
118 number of risk factors for each patient.

119

120 Patients in the validation cohort were assigned risk categories based on the identified risk
121 variables. One-year CVD risk was calculated for patients within different risk categories in
122 the treatment and control groups respectively using the Kaplan-Meier method.³⁹ The
123 treatment effect and the confidence interval across different risk categories were calculated
124 as the coefficient in CPH, which was trained on the composite of treatment and control
125 group patients with treatment as a covariate in each risk category. All analyses were
126 conducted using Python (version 3.9.1), with its add-on package lifelines.⁴⁰ This study report
127 is in accordance with the TRIPOD statement.⁴¹ Ethical approval for this study was granted by
128 the Institutional Review Board of The University of Hong Kong/HA Hong Kong West Cluster
129 (UW24-124).

130

131 **Results**

132 *Study cohorts*

133 For the derivation cohort, we initially identified 48,743 patients aged 18 or above with lipid
134 test records and diagnosis of CVD at hospitals in the Hong Kong West Cluster. We excluded
135 39,262 patients whose first CVD event occurred before January 1, 2015, or after December
136 31, 2018. We excluded 3,169 patients with low dose aspirin treatment before their first CVD.
137 We excluded 4,546 patients with no daily aspirin treatment after CVD. We excluded 140
138 patients with anticoagulant or other antiplatelet treatments before or after CVD. We
139 excluded 3 patients with the first CVD diagnosis of hemorrhagic stroke. Overall, 1,623

140 patients (selected out of 48,743 patients) who took aspirin daily were included in the Hong
141 Kong Island cohort for model derivation.

142

143 For the validation cohort, we initially identified a cohort of 322,731 patients aged 35 or
144 above with blood pressure records and diagnosis of CVD by the Hospital Authority. Similar
145 exclusion criteria were applied to the validation cohort. A total of 245,188 patients were
146 excluded who had their first CVD event before January 1, 2015 or after December 31, 2018.
147 A total of 28,902 of patients were excluded who had low dose aspirin treatment before their
148 first CVD. A total of 8,372 patients were excluded who had received aspirin treatment after
149 CVD, but not daily. A total of 580 patients were excluded because of immediate CVD death.
150 To ensure no overlap with the derivation cohort, we excluded 6,845 patients with most
151 frequently visited healthcare utilization on the Hong Kong Island. We excluded 1,619
152 patients with anticoagulant or other antiplatelet treatments before or after CVD. We
153 excluded 5,286 patients with the first CVD diagnosis of hemorrhagic stroke. Overall, 25,939
154 patients (selected out of 322,731 patients) were included in the Kowloon and New
155 Territories cohort for model validation; 22,192 daily aspirin users were assigned to the
156 treatment group and 3,747 non aspirin users were assigned to the control group.

157

158 *Baseline characteristics*

159 The baseline characteristics were similar across the cohorts and subgroups, to ensure the
160 reliability of our selection criteria and support the subsequent model derivation and
161 validation (Table 1). The median age at the cohort entry date was 67 to 75, and 35–48% of
162 the patients were female. A considerable proportion of the patients had comorbidities: 12-
163 23% with dyslipidemia, 30-40% with hypertension, 19-20% with diabetes, 3-10% with atrial
164 fibrillation, 3-5% with chronic kidney disease, and 6-7% with congestive heart failure.
165 Regarding medication history, 18-21% of patients had received statin therapy, 15-22% of
166 patients had taken antidiabetic drugs, and 42-65% of patients were on antihypertensive
167 drugs. The median number of concurrent medications was 6-7.

168

169 *Model development*

170 The complete list of all available variables and the details of the feature selection process
171 are shown in Table 2. There were initially 109 risk variables under consideration. We
172 excluded 38 of them with a high missing rate (>20%) or low event rate (<1%) from
173 subsequent LASSO regression for better statistical reliability and clinical utility. The LASSO
174 regression on the remaining variables identified 7 statistically significant (p value < 0.05)
175 variables. We shortlisted 4 of them based on clinical evidence. We replaced the use of
176 antiplatelet drugs with fasting glucose level as a measure of diabetic inclination due to the
177 higher data completeness of the latter. We added atrial fibrillation to the final list
178 considering its established role in suboptimal aspirin effectiveness.^{42,43} We translated the
179 selected numerical variables, namely neutrophil count, creatine kinase, glucose levels, and
180 concurrent medication, into categorical variables based on specific diagnostic criteria.
181 Neutrophil count was categorized as neutrophilia (neutrophil count > $7.5 \times 10^9/L$), creatine
182 kinase was categorized as elevated CK (defined as CK > 400 IU/L), glucose levels were
183 categorized as hyperglycemia (glucose > 7 mmol/L), and concurrent medication was
184 categorized as polypharmacy (concurrent medication > 4).

185

186 The identified risk factors encompassed a comprehensive spectrum of direct and indirect
187 causes that contributed aspirin suboptimal effects, which were supported by the literature.
188 First, while aspirin is more effective in atherosclerotic CVD, its effectiveness in other types of
189 CVD may be limited.⁴⁴ Aspirin has limited efficacy for stroke prevention in patients with
190 atrial fibrillation.⁴³ Neutrophilia may serve as an indicator of inflammation signaling arteritis
191 which compromise the efficacy of aspirin.^{44,45} Elevated CK, an indicator of muscle damage, is
192 a possible sign of cardiomyopathy or myonecrosis, which could potentially impact the
193 response to aspirin treatment.^{46,47} Specific comorbidities and prevalent drug use may also
194 affect the treatment effect of aspirin. Polypharmacy can lead to drug interactions that
195 interfere with the effect of aspirin. Hyperglycemia and dyslipidemia may hinder the
196 intended antiplatelet effects of aspirin by increasing the production of prostaglandin F2-like
197 compounds.⁴⁸⁻⁵⁰ The final list of the 6 risk factors was then utilized in the CPH. The
198 parameter estimates for each of the 6 predictor variables showed statistical significance and
199 were of similar magnitude (Table 3).

200

201 To facilitate practical implementation, we adopted a pragmatic approach by categorizing the
202 suboptimal aspirin treatment effect based on the number of risk factors present in an
203 individual. This risk stratification strategy allows for a straightforward assessment of the
204 likelihood and extent of suboptimal aspirin response based on the number and nature of
205 identified risk factors.

206

207 *Model validation*

208 We compared the outcomes in the treatment and control group across different risk
209 categories with different number of risk factors in the validation cohort (Figure 2). For the
210 ease of clinical use and model stability, we defined 3 risk categories regarding the total
211 number of 0, 1-2, and 3-6 risk factors (out of 6) for aspirin treatment effect stratification.
212 Around 80% of patients (79.7% in the treatment group and 78.7% in the control group) were
213 in the risk categories with less than three risk factors. The results showed that patients had
214 similar baseline one-year CVD recurrence risk regardless of the risk category (from 43.5%
215 with 0 risk factors to 52.4% with 3-6 risk factors in the control group). In contrast, the
216 recurrence risk with aspirin treatment showed strong graded association with the risk
217 categories (from 34.0% with 0 risk factors to 54.5% with 3-6 risk factors in the aspirin
218 treatment group). The stable baseline risk and the graded risk after treatment showed the
219 robustness of the risk factors that we identified. As the number of risk factors increased,
220 there was a gradual approaching tendency in the CVD risk between the treatment and
221 control group. This indicated that aspirin treatment might be more effective in individuals
222 with fewer risk factors than those with multiple risk factors.

223

224 The HR of aspirin treatment, estimated based on both groups, exhibited a similar trend
225 (Table 4). Specifically, patients with no risk factors (14.2% of patients) had an aspirin
226 treatment HR of 0.68. This suggested that aspirin treatment significantly reduced the risk of
227 CVD events in this low-risk subgroup. For patients with 1-2 risk factors, which constituted
228 64.8% of the cohort, the aspirin treatment HR is 0.87. Although the effect size was slightly
229 attenuated compared to the no-risk factor group, these individuals still benefited from
230 aspirin treatment as indicated by the reduced HR. In contrast, the remaining 21.0% of
231 patients with 3-6 risk factors exhibited an aspirin treatment HR of 0.99. This group of high-

232 risk patients may be referred to as had aspirin treatment failure, since aspirin therapy did
233 not provide an additional protective effect against CVD events.

234

235 *Website design*

236 A website employed VISTA at www.bio8.cs.hku.hk/vista.html (Supplementary Information 1)
237 was deployed for demonstration, and to enhance utility and clinician-patient discussions.

238

239 **Discussion**

240 Our study represents the first investigation, to the best of our knowledge, into the impact of
241 patient characteristics on the effectiveness of low-dose aspirin for preventing recurrent CVD
242 events in secondary prevention. We conducted rigorous statistical testing on a large pool of
243 variables and identified six relevant risk factors, which encompass potential mechanisms
244 underlying suboptimal treatment effects of aspirin. The results of our model validation
245 further demonstrated the strong performance of our model in stratifying aspirin treatment
246 effects. By leveraging VISTA's capabilities, clinicians gain early access to the potential
247 benefits of aspirin therapy as a treatment option for each patient, allowing them to make
248 informed decisions regarding its suitability. The tool enables clinicians to assess a patient's
249 risk of treatment failure and proactively adjust the medication regimen by considering
250 alternative therapeutic approaches. For instance, if a patient exhibits a high risk of
251 treatment failure, the clinician can opt to rely on other medications, such as setting lower
252 targets for blood pressure and LDL cholesterol with a higher intensity of antihypertensive
253 and lipid-lowering agents. By optimizing the combination of medications based on individual
254 patient characteristics, this decision support tool has the potential to enhance treatment
255 efficacy and improve outcomes in the secondary prevention of CVD.

256

257 Optimizing the prescription of low-dose aspirin for CVD prevention requires a
258 comprehensive evaluation of both the potential bleeding risk and the treatment effect.
259 While a study has attempted on this objective, it only predicted the bleeding risk, without
260 assessing the treatment benefit.⁵¹ In scenarios where the predicted treatment benefit is
261 unknown, clinicians may only rely on the bleeding risk as a reference. If the bleeding risk is
262 not high, aspirin may be considered as a viable option, even though the benefit remains
263 uncertain. Conversely, if the estimated benefit from aspirin is modest, regardless of how low
264 the bleeding risk is, it is advisable to avoid prescribing aspirin and instead prioritize other
265 medications to achieve the targeted treatment objective. Our study addresses this limitation
266 by directly assessing the treatment effect of aspirin, offering a more comprehensive
267 perspective. Moreover, the bleeding risk study identified 20 risk factors that were
268 associated with the bleeding risk. However, only one of them overlapped with our identified
269 risk factors (diabetes), while the other 19 of them were not associated with treatment
270 failure. This disparity suggests that the treatment benefit and bleeding risk operate through
271 distinct mechanisms and systems. It emphasizes the significance of our study in prompting
272 clinicians to consider the benefit-risk scenario differently when prescribing aspirin.

273

274 Aspirin treatment failure has long been recognized as important consideration in clinical
275 practice. Existing methods for assessing aspirin's antiplatelet effects rely on specific
276 laboratory tests, such as PFA-100, WBA, and VerifyNow, which measure individual platelet
277 function to predict the effectiveness of aspirin. Several case studies have shown that aspirin
278 resistance, as indicated by these assays, is associated with a higher rate of CVD events.^{16,52}

279 However, these tests are not readily available in general hospital settings and have modest
280 correlation and agreement. For instance, in a study of 201 patients with stable coronary
281 artery disease receiving >80mg/day of aspirin, the prevalence of aspirin resistance varied
282 greatly: 7% from VerifyNow, 18% from WBA, and 60% from PFA-100.¹⁸ Also, they do not
283 directly predict the effectiveness of aspirin in preventing CVD, limiting their practical
284 usefulness for clinicians' reference. In contrast, our study proposes a practical method that
285 utilizes easily accessible variables to predict the treatment effect of aspirin on CVD
286 prevention. According to the validation results, approximately 80% of patients are expected
287 to experience a considerable treatment effect from aspirin, with hazard ratios ranging from
288 0.67 to 0.83. However, approximately 20% of patients may have treatment failure (HR close
289 to 1.00) from aspirin if two or more risk factors are present. The findings from our research
290 provide valuable guidance for clinicians in the decision-making process of prescribing aspirin.
291 By offering a straightforward estimation of aspirin's treatment effect, our method can be
292 used as a tool to assess the potential benefits of aspirin as a treatment option for each
293 patient.

294
295 Our work introduced a novel lens to investigate treatment recommendation methods in the
296 era of big data. Traditional approaches to evaluate treatment effects rely on randomized
297 controlled trials (RCTs). RCTs face practical constraints, cost considerations, and data
298 management complexities, which often lead to the collection of a limited number of patient
299 characteristics. While RCTs are reliable in yielding population-level outcomes, stratifying
300 individual benefits based on these limited covariates is challenging. In our research, we
301 demonstrated the potential of high-quality retrospective cohort studies with a
302 comprehensive set of covariates in treatment effect stratification. We utilized a robust
303 feature selection process to identify key risk factors in aspirin efficacy variation and applied
304 rigorous validation to estimate the variation of the treatment effects across various risk
305 groups. In the era of big data and with advances in data curation, there is an opportunity to
306 develop more robust treatment effect stratification that complements the population-level
307 results obtained from RCTs. Our approach has proved to be robust in addressing this task
308 and can be extended to analyze the efficacy of various therapies. By leveraging the vast
309 amount of available data, we can gain valuable insights into individual treatment responses
310 and enhance personalized medicine approaches.

311
312 There are limitations in our study. First, both our derivation and validation cohorts were
313 developed using data from Hong Kong. Although Hong Kong is a multi-ethnic city, its
314 population is predominantly Chinese. Consequently, the generalizability of our model's
315 performance to other ethnicities and regions may be uncertain. Future research should aim
316 to apply our model to validation cohorts comprising diverse ethnic populations to assess its
317 performance across different racial groups. Second, our estimation of treatment effects
318 across risk groups was based on a retrospective cohort study design. This introduces the
319 possibility of selection bias in the selection of the treatment and control groups. To further
320 validate our model, it would be beneficial to conduct more rigorous investigations, such as
321 prospective clinical studies that incorporate the collection of the identified key risk factors.
322 This independent validation of each risk factor and evaluation of the performance of risk
323 group treatment stratification would provide stronger evidence for the effectiveness and
324 reliability of our model.

325

326 **Conclusions**

327 The variability in aspirin treatment effects poses a significant challenge in optimizing
328 therapeutic outcomes for patients with CVD. VISTA can be helpful in the secondary
329 prevention of CVD which can identify patients who may be at risk of treatment failure thus
330 help clinicians with treatment planning. We identified six key risk factors, history of atrial
331 fibrillation, dyslipidemia, hyperglycemia, polypharmacy, neutrophilia, and elevated CK, that
332 contribute to the variation in aspirin treatment effects. The big-data approach employed in
333 our study also holds promise across various therapeutic areas beyond aspirin therapy. This
334 research represents a significant step towards advancing personalized medicine in aspirin
335 therapy, ultimately enhancing therapeutic effectiveness and improving CVD prevention on
336 an individualized basis.

337
338 **Acknowledgements**

339 Nil.

340

341 **Sources of funding**

342 Nil.

343

344 **Disclosures**

345 EWYC reports honorarium from Hospital Authority; and grants from Research Grants Council
346 (RGC, Hong Kong), Research Fund Secretariat of the Food and Health Bureau, National
347 Natural Science Fund of China, Wellcome Trust, Bayer, Bristol-Myers Squibb, Pfizer, Janssen,
348 Amgen, Takeda, and Narcotics Division of the Security Bureau of the Hong Kong Special
349 Administrative Region, outside the submitted work. CSLC has received grants from the Food
350 and Health Bureau of the Hong Kong Government, Hong Kong Research Grant Council, Hong
351 Kong Innovation and Technology Commission, Pfizer, IQVIA, MSD, and Amgen; and personal
352 fees from PrimeVigilance; outside the submitted work. RL has received grants from Hong
353 Kong Research Grant Council, Hong Kong Innovation and Technology Commission; and
354 research donations from Oxford Nanopore Technologies, outside the submitted work. All
355 other authors declare no competing interests.

356

357 **References**

- 358 1. Roth GA, Mensah GA, Johnson CO, et al. Global Burden of Cardiovascular Diseases
359 and Risk Factors, 1990-2019: Update From the GBD 2019 Study. *J Am Coll Cardiol* 2020;
360 **76**(25): 2982-3021.
- 361 2. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and
362 injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global
363 Burden of Disease Study 2019. *Lancet* 2020; **396**(10258): 1204-22.
- 364 3. Collaborative overview of randomised trials of antiplatelet therapy--I: Prevention of
365 death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various
366 categories of patients. Antiplatelet Trialists' Collaboration. *BMJ* 1994; **308**(6921): 81-106.
- 367 4. Antithrombotic Trialists C. Collaborative meta-analysis of randomised trials of
368 antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk
369 patients. *BMJ* 2002; **324**(7329): 71-86.
- 370 5. Bhatt DL, Topol EJ. Scientific and therapeutic advances in antiplatelet therapy. *Nat*
371 *Rev Drug Discov* 2003; **2**(1): 15-28.

- 372 6. Neergaard-Petersen S, Ajjan R, Hvas AM, et al. Fibrin clot structure and platelet
373 aggregation in patients with aspirin treatment failure. *PLoS One* 2013; **8**(8): e71150.
- 374 7. Chu JW, Wong CK, Chambers J, Wout JV, Herbison P, Tang EW. Aspirin resistance
375 determined from a bed-side test in patients suspected to have acute coronary syndrome
376 portends a worse 6 months outcome. *QJM* 2010; **103**(6): 405-12.
- 377 8. Airee A, Draper HM, Finks SW. Aspirin resistance: disparities and clinical implications.
378 *Pharmacotherapy* 2008; **28**(8): 999-1018.
- 379 9. Patrono C, Garcia Rodriguez LA, Landolfi R, Baigent C. Low-dose aspirin for the
380 prevention of atherothrombosis. *N Engl J Med* 2005; **353**(22): 2373-83.
- 381 10. Grundmann K, Jaschonek K, Kleine B, Dichgans J, Topka H. Aspirin non-responder
382 status in patients with recurrent cerebral ischemic attacks. *J Neurol* 2003; **250**(1): 63-6.
- 383 11. Antithrombotic Trialists C, Baigent C, Blackwell L, et al. Aspirin in the primary and
384 secondary prevention of vascular disease: collaborative meta-analysis of individual
385 participant data from randomised trials. *Lancet* 2009; **373**(9678): 1849-60.
- 386 12. Harrison P. Progress in the assessment of platelet function. *Br J Haematol* 2000;
387 **111**(3): 733-44.
- 388 13. Kundu SK, Heilmann EJ, Sio R, Garcia C, Davidson RM, Ostgaard RA. Description of an
389 in vitro platelet function analyzer--PFA-100. *Semin Thromb Hemost* 1995; **21** Suppl 2: 106-12.
- 390 14. Rand ML, Leung R, Packham MA. Platelet function assays. *Transfus Apher Sci* 2003;
391 **28**(3): 307-17.
- 392 15. Patrono C. Aspirin resistance: definition, mechanisms and clinical read-outs. *J*
393 *Thromb Haemost* 2003; **1**(8): 1710-3.
- 394 16. Gum PA, Kottke-Marchant K, Welsh PA, White J, Topol EJ. A prospective, blinded
395 determination of the natural history of aspirin resistance among stable patients with
396 cardiovascular disease. *J Am Coll Cardiol* 2003; **41**(6): 961-5.
- 397 17. Chen WH, Cheng X, Lee PY, et al. Aspirin resistance and adverse clinical events in
398 patients with coronary artery disease. *Am J Med* 2007; **120**(7): 631-5.
- 399 18. Lordkipanidze M, Pharand C, Schampaert E, Turgeon J, Palisaitis DA, Diodati JG. A
400 comparison of six major platelet function tests to determine the prevalence of aspirin
401 resistance in patients with stable coronary artery disease. *Eur Heart J* 2007; **28**(14): 1702-8.
- 402 19. Breet NJ, van Werkum JW, Bouman HJ, Kelder JC, Ten Berg JM, Hackeng CM. High
403 on-aspirin platelet reactivity as measured with aggregation-based, cyclooxygenase-1
404 inhibition sensitive platelet function tests is associated with the occurrence of
405 atherothrombotic events. *J Thromb Haemost* 2010; **8**(10): 2140-8.
- 406 20. Grove EL, Hvas AM, Johnsen HL, et al. A comparison of platelet function tests and
407 thromboxane metabolites to evaluate aspirin response in healthy individuals and patients
408 with coronary artery disease. *Thromb Haemost* 2010; **103**(6): 1245-53.
- 409 21. Cattaneo M. Laboratory detection of 'aspirin resistance': what test should we use (if
410 any)? *Eur Heart J* 2007; **28**(14): 1673-5.
- 411 22. Lev EI. Aspirin resistance transient laboratory finding or important clinical entity? *J*
412 *Am Coll Cardiol* 2009; **53**(8): 678-80.
- 413 23. Muir AR, McMullin MF, Patterson C, McKeown PP. Assessment of aspirin resistance
414 varies on a temporal basis in patients with ischaemic heart disease. *Heart* 2009; **95**(15):
415 1225-9.
- 416 24. Santilli F, Rocca B, De Cristofaro R, et al. Platelet cyclooxygenase inhibition by low-
417 dose aspirin is not reflected consistently by platelet function assays: implications for aspirin
418 "resistance". *J Am Coll Cardiol* 2009; **53**(8): 667-77.

- 419 25. Shen H, Herzog W, Drolet M, et al. Aspirin Resistance in healthy drug-naive men
420 versus women (from the Heredity and Phenotype Intervention Heart Study). *Am J Cardiol*
421 2009; **104**(4): 606-12.
- 422 26. Michelson AD, Cattaneo M, Eikelboom JW, et al. Aspirin resistance: position paper of
423 the Working Group on Aspirin Resistance. *J Thromb Haemost* 2005; **3**(6): 1309-11.
- 424 27. Wang TH, Bhatt DL, Topol EJ. Aspirin and clopidogrel resistance: an emerging clinical
425 entity. *Eur Heart J* 2006; **27**(6): 647-54.
- 426 28. Census and Statistics Department. Hong Kong Special Administrative Region.
427 Thematic Household Survey Report No. 68. 2018.
428 https://www.censtatd.gov.hk/en/data/stat_report/product/B1130201/att/B11302682019X
429 XXXB0100.pdf.
- 430 29. Wong AY, Root A, Douglas IJ, et al. Cardiovascular outcomes associated with use of
431 clarithromycin: population based study. *bmj* 2016; **352**.
- 432 30. Shami JJP, Yan VKC, Wei Y, et al. Low-dose aspirin does not lower the risk of
433 colorectal cancer in patients with type 2 diabetes taking metformin. *J Intern Med* 2023;
434 **293**(3): 371-83.
- 435 31. Zhou Y, Lin CJ, Yu Q, et al. Development and validation of risk prediction model for
436 recurrent cardiovascular events among Chinese: the Personalized CARdiovascular Disease
437 risk Assessment for Chinese model. *European Heart Journal - Digital Health* 2024.
- 438 32. Zhou Y, Lin CJX, Yu Q, et al. Primary Prevention Cardiovascular Disease Risk
439 Prediction Model for Contemporary Chinese (1° P-CARDIAC): Model Derivation and
440 Validation Using a Hybrid Statistical and Machine-Learning Approach.
- 441 33. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained
442 Equations in R. *Journal of Statistical Software* 2011; **45**(3): 1 - 67.
- 443 34. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations:
444 what is it and how does it work? *Int J Methods Psychiatr Res* 2011; **20**(1): 40-9.
- 445 35. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal*
446 *Statistical Society: Series B (Methodological)* 1996; **58**(1): 267-88.
- 447 36. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society:*
448 *Series B (Methodological)* 1972; **34**(2): 187-202.
- 449 37. Deo SV, Deo V, Sundaram V. Survival analysis—part 2: Cox proportional hazards
450 model. *Indian journal of thoracic and cardiovascular surgery* 2021; **37**(2): 229-33.
- 451 38. Harrell FE. Regression modeling strategies: with applications to linear models,
452 logistic regression, and survival analysis: Springer; 2001.
- 453 39. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations.
454 *Journal of the American Statistical Association* 1958; **53**: 457-81.
- 455 40. Davidson-Pilon C. lifelines: survival analysis in Python. *Journal of Open Source*
456 *Software* 2019; **4**(40): 1317.
- 457 41. Transparent reporting of a multivariable prediction model for individual prognosis or
458 diagnosis (TRIPOD): The TRIPODstatement. The UK EQUATOR Centre.
- 459 42. Warfarin versus aspirin for prevention of thromboembolism in atrial fibrillation:
460 Stroke Prevention in Atrial Fibrillation II Study. *Lancet* 1994; **343**(8899): 687-91.
- 461 43. Lip GY. The role of aspirin for stroke prevention in atrial fibrillation. *Nat Rev Cardiol*
462 2011; **8**(10): 602-6.
- 463 44. Sanderson S, Emery J, Baglin T, Kinmonth AL. Narrative review: aspirin resistance and
464 its clinical implications. *Ann Intern Med* 2005; **142**(5): 370-80.

- 465 45. Arruda-Olson AM, Reeder GS, Bell MR, Weston SA, Roger VL. Neutrophilia predicts
466 death and heart failure after myocardial infarction: a community-based study. *Circ*
467 *Cardiovasc Qual Outcomes* 2009; **2**(6): 656-62.
- 468 46. Savonitto S, Granger CB, Ardissino D, et al. The prognostic value of creatine kinase
469 elevations extends across the whole spectrum of acute coronary syndromes. *J Am Coll*
470 *Cardiol* 2002; **39**(1): 22-9.
- 471 47. Chen W-H, Lee P-Y, Ng W, Tse H-F, Lau C-P. Aspirin resistance is associated with a
472 high incidence of myonecrosis after non-urgent percutaneous coronary intervention despite
473 clopidogrel pretreatment. *Journal of the American College of Cardiology* 2004; **43**(6): 1122-6.
- 474 48. Cambria-Kiely JA, Gandhi PJ. Possible mechanisms of aspirin resistance. *J Thromb*
475 *Thrombolysis* 2002; **13**(1): 49-56.
- 476 49. Davi G, Ciabattini G, Consoli A, et al. In vivo formation of 8-iso-prostaglandin f2alpha
477 and platelet activation in diabetes mellitus: effects of improved metabolic control and
478 vitamin E supplementation. *Circulation* 1999; **99**(2): 224-9.
- 479 50. Parker WAE, Sagar R, Kurdee Z, et al. A randomised controlled trial to assess the
480 antithrombotic effects of aspirin in type 1 diabetes: role of dosing and glycaemic control.
481 *Cardiovasc Diabetol* 2021; **20**(1): 238.
- 482 51. Selak V, Jackson R, Poppe K, et al. Predicting Bleeding Risk to Guide Aspirin Use for
483 the Primary Prevention of Cardiovascular Disease: A Cohort Study. *Ann Intern Med* 2019;
484 **170**(6): 357-68.
- 485 52. Eikelboom JW, Hirsh J, Weitz JI, Johnston M, Yi Q, Yusuf S. Aspirin-resistant
486 thromboxane biosynthesis and the risk of myocardial infarction, stroke, or cardiovascular
487 death in patients at high risk for cardiovascular events. *Circulation* 2002; **105**(14): 1650-5.
- 488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510

Table 1. Baseline characteristics of the derivation and validation cohorts

	Derivation cohort: Hong Kong Island (n=1,623)		Validation cohort: Kowloon and New Territories			
			Treatment group (n=22,192)		Control group (n=3,747)	
Demographic factors						
Age	67	(58-78)	68	(59-79)	75	(63-85)
Female sex	561	(35%)	8,404	(38%)	1,807	(48%)
Male sex	1,062	(65%)	13,788	(62%)	1,940	(52%)
Disease history						
Atrial fibrillation	74	(5%)	708	(3%)	390	(10%)
Dyslipidemia	215	(13%)	5,180	(23%)	431	(12%)
Diabetes	315	(19%)	4,404	(20%)	725	(19%)
Hypertension	492	(30%)	7,694	(35%)	1,488	(40%)
Chronic kidney disease	43	(3%)	563	(3%)	179	(5%)
Congestive heart failure	109	(7%)	1,379	(6%)	274	(7%)
Medication history						
Statins	288	(18%)	4,638	(21%)	784	(21%)
Antidiabetic drugs	242	(15%)	4,487	(20%)	841	(22%)
Antihypertensive	680	(42%)	12,556	(57%)	2453	(65%)
Count of medication	6	(4-8)	6	(5-9)	7	(5-10)
Clinical laboratory tests						
Neutrophil, 10 ⁹ /L	4.8	(3.6-6.7, 0%)	5.4	(4.0-7.7, 1%)	6.2	(4.3-9.8, 1%)
Creatine kinase, IU/L	140	(83-319, 14%)	136	(82-304, 10%)	112	(63-219, 11%)
Glucose (fasting), mmol/L	5.7	(5.2-6.9, 1%)	5.8	(5.2-7.0, 1%)	5.7	(5.1-7.0, 9%)

All data in n (%), median (interquartile range), or median (interquartile range, proportion of missing data).

511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534

Table 2. Summary of included 109 variables assessed in the model derivation cohort

Categories (number of variables)	Risk variables
Demographic factors (2)	age, sex
Family history of disease (2)	diabetes, cardiovascular disease
Healthcare utilization (3)	outpatient visits per year*, accident and emergency visits per year, inpatient visits per year
Clinical laboratory tests (39)	neutrophil*^, creatine kinase (total)*^, monocyte*, glucose (fasting)^, aspartate transaminase, alanine aminotransferase, low-density lipoprotein cholesterol, hemoglobin A1c, prothrombin time, potassium (serum), estimated glomerular filtration rate, triglycerides, basophil, albumin, international normalized ratio, bilirubin (total), total cholesterol, lymphocyte, creatinine (serum), platelet, red blood cell, high-density lipoprotein cholesterol, body mass index, calcium (serum), white blood cell, alkaline phosphatase, sodium (serum), eosinophil, hemoglobin, thyroid stimulating hormone, diastolic blood pressure, systolic blood pressure, erythrocyte sedimentation rate, C-reactive protein, bicarbonate (serum), blood pH, free thyroxine, arterial partial pressure of oxygen, troponin-I
Medication history (24)	antidiabetic drugs*, count of medication*^, statins, antihypertensive drugs, non-steroidal anti-inflammatory drugs, corticosteroids, proton-pump inhibitors, H2-receptor antagonists, nicotine replacement therapy, antiarrhythmic drugs, antithyroid drugs, psychotropic drugs, nitrates, thyroid hormones, fibrates, other non-statin lipid-modifying drugs, cardiac glycosides, omega-3 fatty acids, PCSK9 inhibitors, cholesterol absorption inhibitors, bile acid sequestrants, testosterone, oestrogen, niacin
Disease history (39)	dyslipidemia*^, pacemaker implantation, atrial fibrillation^, hypertension, diabetes, congestive heart failure, thyroid disease, arrhythmia and conduction disorders, obesity, oxygen therapy/ventilator/intubation, asthma, injury and poisoning, severe mental illness, dementia, liver disease, chronic obstructive pulmonary disease, cancer, renal disease, chronic kidney disease, dialysis, Huntington's disease, erectile dysfunction, systemic lupus erythematosus, Parkinson's disease, nephrotic syndrome, Creutzfeldt-Jakob disease, Down's syndrome, defibrillator insertion, smoker, alcohol user, heart transplantation, rheumatoid arthritis, hypothyroidism, cardioversion, muscle pain or myopathy or rhabdomyolysis, migraine, cardiac wall/valve/shunt replacement/repairment, cardiomyopathy, major organ bleeding

Strikethrough: excluded from LASSO regression due to high missing rate (>20%) or low event rate (<1%).

*Significant variables (p-value<0.05) in LASSO regression. ^Selected variables for final modeling. PCSK9 = Proprotein convertase subtilisin/kexin type 9.

535
536
537
538
539
540
541
542
543

Table 3. Multivariate stratification model for aspirin treatment effect

Variable	HR (95%CI)	p-value	points
Atrial fibrillation	1.73 (1.33-2.24)	<0.05	1
Dyslipidemia	1.19 (1.00-1.42)	0.06	1
Hyperglycemia	1.19 (1.03-1.38)	<0.05	1
Polypharmacy	1.23 (1.08-1.41)	<0.05	1
Neutrophilia	1.45 (1.23-1.71)	<0.05	1
Elevated CK	1.70 (1.46-1.99)	<0.05	1

HR = hazard ratio. CI = confidence interval. Hyperglycemia = glucose > 7 (mmol/L).

Polypharmacy = concurrent medication > 4. Neutrophilia = neutrophil > 7.5 (10⁹/L).

Elevated CK = creatine kinase > 400 (IU/L).

544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582

Table 4. Treatment effect stratification for low-dose aspirin compared to no aspirin treatment after the

diagnosis of CVD in the validation cohort (n=25,939)

Number of VISTA risk factors	One-year aspirin treatment effect (HR)	Number of patients	Proportion of patients
0	0.677 (0.569-0.806)	4,403	14.2%
1-2	0.866 (0.810-0.927)	16,799	64.8%
3-6	0.991 (0.876-1.121)	5459	21.0%

583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622

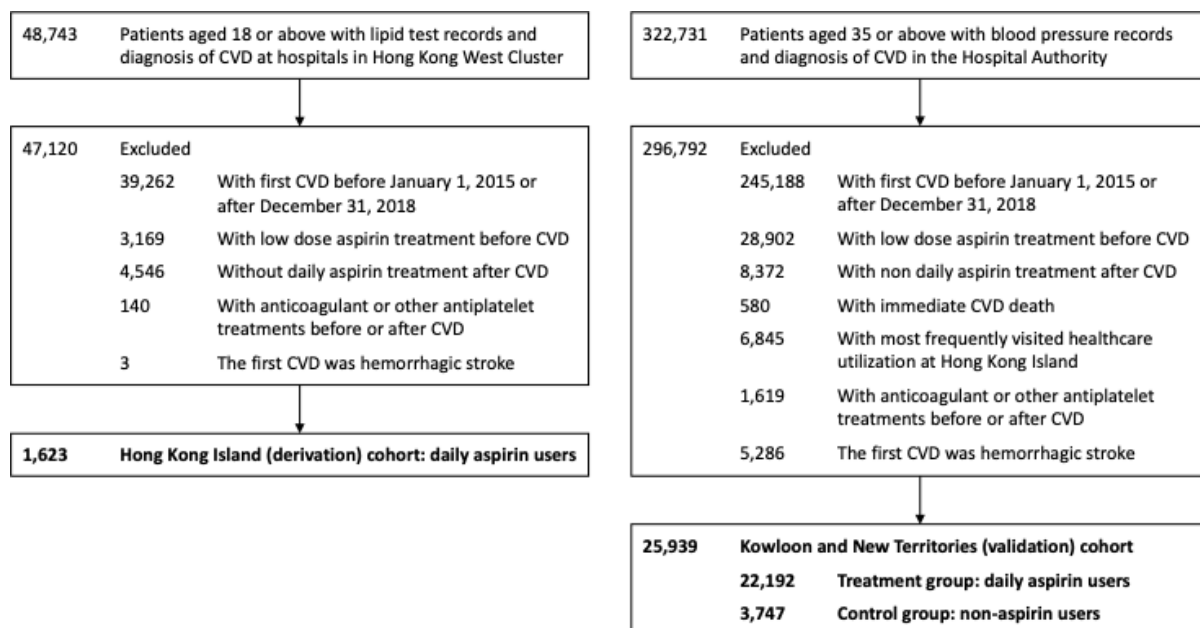
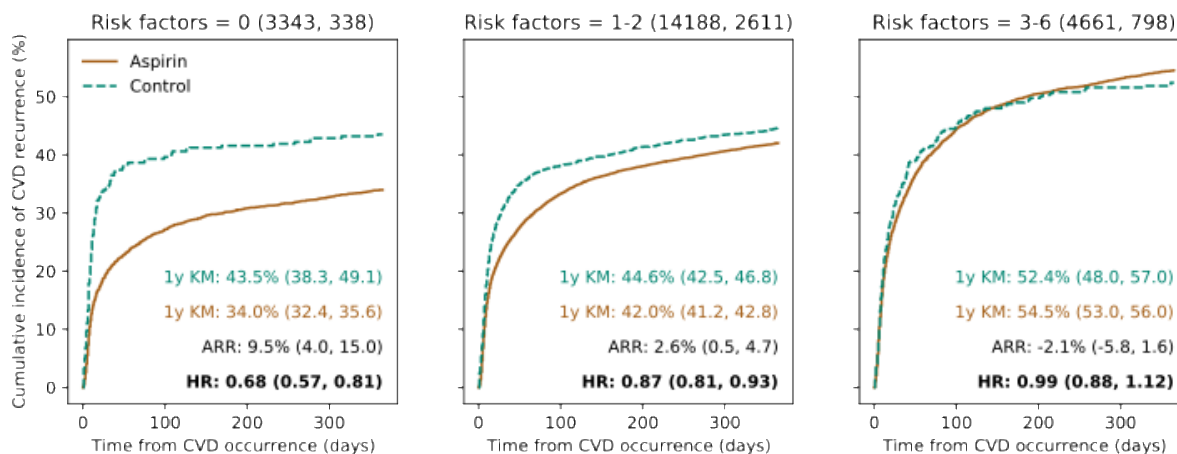


Figure 1. Selection of patients. Hong Kong West Cluster is a part of Hong Kong Island. CVD = cardiovascular diseases.

623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657



658
659
660
661
662
663

Figure 2. Outcomes by risk categories and treatment in the validation cohort. Kaplan-Meier curves stratified by different risk categories and treatment groups (aspirin versus control). Number of patients in aspirin group and control group in each risk category were shown after each subtitle (aspirin, control). One year Kaplan-Meier CVD recurrence probability (1y KM) in the two groups, the corresponding absolute risk reduction (ARR), and the hazard ratio (HR) with 95% confidence interval for aspirin versus control groups are shown.