

# BMJ Open Ten-year risk prediction models of complications and mortality of Chinese patients with diabetes mellitus in primary care in Hong Kong: a study protocol

Eric Yuk Fai Wan,<sup>1</sup> Esther Yee Tak Yu,<sup>1</sup> Weng Yee Chin,<sup>1</sup> Colman Siu Cheung Fung,<sup>1</sup> Ruby Lai Ping Kwok,<sup>2</sup> David Vai Kiong Chao,<sup>3</sup> King Hong Chan,<sup>4</sup> Eric Ming-Tung Hui,<sup>5</sup> Wendy Wing Sze Tsui,<sup>6</sup> Kathryn Choon Beng Tan,<sup>7</sup> Daniel Yee Tak Fong,<sup>8</sup> Cindy Lo Kuen Lam<sup>1</sup>

**To cite:** Wan EYF, Yu EYT, Chin WY, *et al.* Ten-year risk prediction models of complications and mortality of Chinese patients with diabetes mellitus in primary care in Hong Kong: a study protocol. *BMJ Open* 2018;**8**:e023070. doi:10.1136/bmjopen-2018-023070

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-023070>).

Received 19 March 2018  
Revised 19 June 2018  
Accepted 13 September 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Cindy Lo Kuen Lam;  
[cklam@hku.hk](mailto:cklam@hku.hk)

## ABSTRACT

**Introduction** Diabetes mellitus (DM) is a major disease burden worldwide because it is associated with disabling and lethal complications. DM complication risk assessment and stratification is key to cost-effective management and tertiary prevention for patients with diabetes in primary care. Existing risk prediction functions were found to be inaccurate in Chinese patients with diabetes in primary care. This study aims to develop 10-year risk prediction models for total cardiovascular diseases (CVD) and all-cause mortality among Chinese patients with DM in primary care.

**Methods and analysis** A 10-year cohort study on a population-based primary care cohort of Chinese patients with diabetes, who were receiving care in the Hospital Authority General Outpatient Clinic on or before 1 January 2008, were identified from the clinical management system database of the Hospital Authority. All patients with complete baseline risk factors will be included and followed from 1 January 2008 to 31 December 2017 for the development and validation of prediction models. The analyses will be carried out separately for men and women. Two-thirds of subjects will be randomly selected as the training sample for model development. Cox regressions will be used to develop 10-year risk prediction models of total CVD and all-cause mortality. The validity of models will be tested on the remaining one-third of subjects by Harrell's C-statistics and calibration plot. Risk prediction models for diabetic complications specific to Chinese patients in primary care will enable accurate risk stratification, prioritisation of resources and more cost-effective interventions for patients with DM in primary care.

**Ethics and dissemination** The study was approved by the Institutional Review Board of the University of Hong Kong—the Hospital Authority Hong Kong West Cluster (reference number: UW 15–258).

**Trial registration number** NCT03299010; Pre-results.

## INTRODUCTION

Diabetes mellitus (DM) is a well-recognised public health issue, affecting 415 million people and costing HK\$5.2 trillion in global

## Strengths and limitations of this study

- This is a 10-year retrospective population-based cohort study cohort of Chinese patients with diabetes mellitus in primary care which can represent the situation in Hong Kong.
- Two-thirds of samples in the cohort will be randomly selected for developing risk prediction models while remaining one-third would be used for validation to ensure the performance of models.
- Risk prediction nomograms and charts will be established based on the risk prediction models for a convenient use in clinical setting.
- Multiple imputation will be used to handle missing data to minimise the bias in developing risk prediction models.
- Misclassification bias may exist by using diagnosis coding such as International Classification of Primary Care, Second edition and International Classification of Diseases, Ninth Revision, Clinical Modification to identify the outcome events of patients.

health expenditures worldwide.<sup>1</sup> DM can lead to many complications resulting in morbidity and mortality. According to the International Diabetes Federation (IDF), in 2015, diabetes led to 5.0 million (14.5% of all deaths) deaths worldwide which translated to one death in every 6 s and approximately 70% of DM-related deaths were attributed to cardiovascular diseases (CVD).<sup>1</sup> The development of diabetes-related complications significantly increases medical costs.<sup>1</sup>

To prevent DM complications, the American Heart Association (AHA) guidelines recommend primary care providers to provide regular assessment and management of risk factors for patients especially those who are at

high risk of developing DM complications. Although the National Cholesterol Education Programme in the USA has suggested that all patients with diabetes be treated as if they had coronary heart disease (CHD), however, the observed rates of CVD vary vastly among different patients with diabetes.<sup>2</sup> The American Diabetes Association (ADA) and the Canadian Diabetes Association guidelines both include 10-year overall CVD risk stratification into account to identify high-risk patients for more intensive medical and psychosocial interventions.<sup>3</sup> The guidance of statin prescription from the American College of Cardiology and the AHA, which is consistent with the ADA, also takes predicted 10-year overall CVD risk into account.<sup>4</sup> The ADA recommends aspirin treatment for patients with diabetes with a 10-year predicted over CVD risk higher than 10%.<sup>3</sup> Studies in the USA, the UK, Australia, New Zealand and Hong Kong showed that systematic risk assessment and risk-stratified management initiatives in primary care settings could improve clinical outcomes such as haemoglobin A1c (HbA1c), blood pressure (BP) and low-density lipoprotein cholesterol (LDL-C), as well as reduce utilisation of health services including accident and emergency attendance and hospital admissions.<sup>5</sup> In 2009, the Hong Kong Hospital Authority (HA) made an initiation to enhance the quality of DM care in all HA primary care clinics by the introduction of the multidisciplinary Risk Assessment and Management Programme-DM (RAMP-DM) to systematically assess the CVD risk of patients with DM and then managed according to risk-stratified protocols.<sup>6</sup>

A key to cost-effective management of DM is an accurate risk assessment and stratification system that identifies high-risk patients for more intensive medical and psychosocial interventions. At the same time, an accurate estimation of risk distribution can inform policy-makers to allocate appropriate resources and plan services that can maximise population health benefit for patients with DM. Most of risk-stratified interventions in the guidance were based on common prediction functions for 10-year risk including the Framingham,<sup>7,8</sup> QRisk<sup>9,10</sup> and the European Systematic Coronary Risk Evaluation.<sup>11</sup> However, most of existing prediction models were established and validated for Western populations. Our previous studies developed a series of models for 5 years DM-related complications including CVD, end-stage renal disease (ESRD) and all-cause mortality,<sup>12–14</sup> and showed that other 5-year risk prediction model based on non-Chinese populations such as Framingham, the Action in Diabetes and Vascular Disease: Preterax and Diamicron-MR Controlled Evaluation, Swedish and New Zealand CVD risk scores either underestimated or overestimated the risk for Chinese patients.<sup>12–14</sup> Indeed, the prevalence of CVD in Chinese populations was only half of that in Caucasian populations.<sup>15,16</sup> A recent observational study also illustrated that CVD risk, even in Asian populations, varied widely among the Malay, Asian Indian and Chinese populations.<sup>17</sup> In addition, several multiethnic studies showed substantial difference in the incidence rate among different racial groups with a generally higher incidence rate of renal disease in Chinese than in non-Chinese patients with DM.<sup>18–20</sup> The

IDF report revealed that 1.3 million Chinese died because of DM in 2015, which represented the highest prevalence of DM-related mortality across various ethnic groups and was twofold higher than those found in Europeans and Australians.<sup>1</sup> These discrepancies in CVD, ESRD and all-cause mortality risk may be related to the differences in the disease profile and other determinants such as genetics, healthcare policy and culture.<sup>15,16,21–23</sup> Therefore, Chinese population specific risk prediction models are necessary.

Inaccurate risk stratification may lead to inappropriate risk-stratified interventions. There is a need for new robust risk prediction models, and thus the aim of this study protocol to develop the models to predict 10-year CVD risk and mortality for primary care Chinese patients in order to enable accurate risk stratification of patients with DM in the HA ongoing RAMP-DM or other primary care systematic risk-stratified multidisciplinary management programmes. Furthermore, robust risk prediction models for the overall prediction of first CVD and all-cause mortality can inform policy-makers in service planning and resource allocation.

### Aims and objectives

This study protocol aims to develop 10-year risk prediction models for total CVD and all-cause mortality among Chinese patients with diabetes in primary care. Risk prediction models for individual DM complications including CHD, heart failure, stroke and ESRD will also be developed.

The objectives are to:

1. Calculate the 10-year incidence of total CVD, all-cause mortality and each major DM complication in Chinese patients with DM in primary care.
2. Determine the risk factors that significantly predict total CVD, all-cause mortality and each major DM complication for Chinese patients with DM in primary care.
3. Develop and validate risk prediction models for total CVD, all-cause mortality and each major DM complication for Chinese patients with DM in primary care.
4. Develop a risk prediction nomogram and chart for the risk of total CVD, all-cause mortality for Chinese patients with DM in primary care.

The following hypotheses will be tested:

1. Patient sociodemographic, clinical parameters, disease characteristics and treatment modalities (these independent variables are described in the Methods/Design section) are predictive of 10-year risk of total CVD, all-cause mortality and individual DM complication as a dependent variable.
2. The risk prediction models for total CVD, all-cause mortality and individual DM complication developed in this study can have over 70% of discriminating power.

### METHODS AND ANALYSIS

#### Study design

A 10-year retrospective study on a population-based cohort of Chinese patients with DM in primary care.

## Subjects

The cohort will include all patients with a documented clinical diagnosis of DM and were receiving care in the HA primary care General Outpatient Clinics (GOPC) on or before 1 January 2008 identified from the HA clinical management system (CMS) database.

The inclusion criteria are patients aged 18 years old or older, had at least one GOPC/Family Medicine Clinics attendance on or within 1 year before 1 January 2008 and had a CMS record of the coding of International Classification of Primary Care, Second edition (ICPC-2) of T89 (Diabetes insulin dependent) or T90 (Diabetes non-insulin dependent) on or before 1 January 2008.

The exclusion criteria are patients who had a diagnosis of any DM complications defined by the relevant ICPC-2 or The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) (shown in the section below) on or before 1 January 2008 and patients exclusively managed by Specialist Outpatient Clinic (SOPC) on or before 1 January 2008.

## Sample size calculation

The required sample size is based on the requirements for the development and validation of the least common DM complication of ESRD. Specifically, based on our previous study, the 5-year incidence of ESRD was 1.9%,<sup>6</sup> which can be extrapolated to a 10-year incidence of ESRD was 3.8% by assuming a constant incidence rate over time. To develop the risk prediction model for ESRD by multivariable Cox proportional hazard regressions with forward stepwise selection on 16 potential risks factors, we need 21 053 subjects using the 1 in 50 rule that 1 candidate predictor can be studied for every 50 events.<sup>24</sup> The split samples on a 2:1 basis will be applied, and thus, 10 527 subjects are needed to validate the risk prediction models. Overall, a total 31 580 male and 31 580 female are needed for the development (training dataset) and validation of risk prediction models stratified by gender.

## Definition of indicator DM complications

The incidence of four major DM complications (CHD, stroke, heart failure and ESRD), total CVD and all-cause mortality will be calculated. The incidence is counted from the earliest date of documented diagnosis defined by the relevant ICPC-2 and/or ICD-9-CM coding recorded in the HA CMS database from 1 January 2008 to 31 December 2017. The relevant ICPC-2 and ICD-9-CM codes of each DM complication and mortality are determined by the academic and HA clinician coinvestigators as listed below:

1. CHD (ischaemic heart disease, myocardial infarction (MI), coronary death or sudden death) is defined by any of ICPC-2 K74 to K76 and ICD-9-CM 410.x, 411.x to 414.x, 798.x.
2. Stroke (fatal and non-fatal stroke) is defined by any of ICPC-2 K89 to K91 or ICD-9-CM 430.x to 438.x.
3. Heart failure is defined by any of ICPC-2 K77 or ICD-9-CM 428.x.

4. CVD is defined as the presence of any of CHD, heart failure and stroke ICPC-2 or ICD-9-CM codes listed in 1, 2 and 3 above.
5. ESRD is defined by any of ICD-9-CM 250.3x, 585.x, 586.x, or an estimated glomerular filtration rate (eGFR) <15 mL/min/1.73 m<sup>2</sup>.
6. Mortality is identified from the Hong Kong Death Registry.

## Risk factors to be included in the risk prediction models

Risk factors (independent variables) previously found to be associated with DM complications from the literature<sup>2</sup> and those that are routinely available in primary care are selected to strike a balance between comprehensiveness and feasibility. The potential risk factors that will be explored include those related to patient's sociodemographics, clinical parameters, disease characteristics and treatment modalities. Patient sociodemographics include sex, age and smoking status. Clinical parameters include body mass index (BMI), HbA1c, systolic BP (SBP) and diastolic BP, lipid profile (total cholesterol, high-density lipoprotein cholesterol, LDL-C, triglyceride), eGFR and albuminuria. Disease characteristics include the duration of DM and comorbidity. Treatment modalities include the use of specific antihypertensive drugs, insulin, specific oral antidiabetic drugs and lipid-lowering agents. The operational definitions of the risk factors are shown in online supplementary file 1. These factors, except sex, age, duration of DM and comorbidity, are modifiable, which have implications for practice.

## Data collection

In the middle of 2018, anonymous data from 1 January 2008 to 31 December 2017 of all patients with DM who satisfy the inclusion criteria and without any exclusion criteria will be extracted by the HA statistics team from the HA CMS database. We have successful experience in working with the HA in the extraction of similar data from 2009 to 2013 for our extended evaluation on quality of care and effectiveness of RAMP-DM study,<sup>5</sup> and we have obtained preliminary agreement from the HA for the data extraction in the present study.

## Outcome measures

1. The incidence of total CVD, all-cause mortality and each of four major DM complications over 10 years.
2. Factors predictive of total CVD, all-cause mortality and each of four major DM complications over 10 years.
3. Ten-year risk prediction models for total CVD, all-cause mortality and each of four major DM complications.
4. Factors that have sufficient power to classify Chinese patients with DM in primary care into risk group in terms of total CVD and all-cause mortality.

## Data processing and analysis

The cohort will be stratified by gender. Descriptive statistics will be used to calculate the incidence of total CVD, all-cause mortality and each of four major DM complications will be analysed annually and cumulatively over



10years with a 95% CI. The distribution of risk factors will be cross-tabulated by complication or mortality events. The 10-year cumulative incidence of various DM complications and mortalities will be further analysed by Kaplan-Meier method. The Kaplan-Meier survival curve will be used to describe the survivorship of total CVD, all-cause mortality and each of four major DM complications in the study cohort over 10years. Unadjusted associations between the risk factors and odds of events will be assessed by independent t-test for continuous variables or  $\chi^2$  test for categorical variables.

The cohort will be randomly split on a 2:1 basis, with two-thirds sample used for developing the risk prediction models, and the other one-third sample used for validation of the risk prediction models. The analyses will be carried out separately for men and women.

### Development of risk prediction models

Cox proportional hazard regressions with forward stepwise method will be used to develop the risk prediction models for total CVD, all-cause mortality and each of four major DM complications. If the main term of a clinical parameter is selected in the models, the quadratic term of such clinical parameter will be evaluated. Afterwards, the interaction terms between selected predictors and age will be also examined in the risk prediction models. Cox regression is the method most commonly used in risk prediction models in the Framingham Heart Study<sup>78</sup> and United Kingdom Prospective Diabetes Study.<sup>25 26</sup> It allows us to estimate the risk of disease or death for an individual, given their prognostic variables. A positive HR means a higher likelihood of event associated with that specific variable. Conversely, a negative HR means a lower likelihood of the event associated with that specific variable. The key proportional hazards assumption will be assessed by examining plots of the scaled Schoenfeld residuals against time for the covariates. Any non-random pattern indicates a violation of the proportional hazards assumptions in which case transformation of covariates may be necessary. For example, all continuous variables were naturally logarithmically transformed to minimise the influence of extreme values and to improve discrimination and calibration of the models. A parametric approach such as exponential or Weibull distribution for the hazard function can also be carried out. A total of six risk prediction models will be established for total CVD, all-cause mortality and each of four major DM complications. The log of the HR of each selected risk factor in the final model will be used as coefficient weights in the prediction model of each relevant outcome. The risk equations for 10years follow-up will be established by combining these weights with the survivor function.<sup>7</sup>

### Validation of risk prediction models

To validate the risk prediction models for total CVD, all-cause mortality and each of four major DM complications, the remaining one-third validation sample will be used to estimate the risk level of the subjects. The

performance of discrimination of the model will be assessed by Harrell's C-statistic, which is a measure similar to the area under the curve after consideration of the censoring pattern of the patients. A Harrell's C-statistic less than 0.7 indicates limited discriminating power, 0.7–0.9 is acceptable and higher than 0.9 suggests strong discrimination of the predictive models.<sup>27</sup> The D-statistic,  $R^2$  statistic and Brier score will be also calculated for evaluating the predictive power of the model. The D-statistic is a measure of discrimination where higher value indicates better discrimination. The  $R^2$  statistic is a measure of explained variation with a higher value indicating better performance. The Brier score is a measure of goodness of fit in which a lower value means higher accuracy.

Calibration will be used to measure how closely predicted outcomes agree with actual outcomes. Calibration of the model's ability to correctly estimate the absolute risks will be examined by modified Hosmer-Lemeshow test and calibration plots. The modified Hosmer-Lemeshow test for time-to-event data measures how well the predicted probability of the expected event rate agrees with the observed event rate, where a  $p > 0.05$  indicates good model calibration. In a calibration plot of the observed incidence of events against the predicted risk shows the scatter along the 45° line of perfect fit between predicted risk and observed incidence of event throughout the entire risk spectrum.

### Development of a risk prediction nomogram and chart

In order to enable the 10-year risk prediction models for total CVD and all-cause mortality to be applied in busy clinical setting, risk prediction nomograms and charts will be developed for men and women. For the nomogram, the patient's score for each predictor is plotted on the appropriate scale and vertical lines are drawn to the line of points to obtain the corresponding scores. The score of each predictor will be transformed based on the estimated standardised beta coefficient of each predictor from the risk prediction model. For continuous predictor such as age, the line with interval depends on its units from minimum to maximum values among studies subjects (eg, 20, 40, 60, 80years old) will be displayed on the nomogram plot and the corresponding scores will be obtained based on the estimated standardised beta coefficient of the predictor (eg, assign age of 20 to 0 point, age of 40 to 2 points, age of 60 to 4 points and age of 80 to 6 points). For categorical predictor such as gender, each level of the predictor will be ranked a corresponding score based on the estimated standardised beta coefficient of the predictor (eg, assign female to 0 point and male to 3 points). All scores are summed to obtain a total score. The total score is plotted on the total line with corresponding predicted risk of CVD. Moreover, we will develop risk prediction charts similar to those developed by the Joint British Society. The most significant predictors, up to a maximum of five, found in the full Cox regression models will be selected to classify subjects into 10-year CVD risk groups of <10% (low risk), 10%–20%

(medium risk) and >20% (high risk). The Kaplan-Meier survival curves of each risk group will be developed and compared by log-rank tests to confirm the HRs are significantly different among all risk groups.

STATA software V.13 will be used for data analyses. A significance level of 5% is used in all statistical tests.

## DISCUSSION

Primary care is the entry point of the entire medical system. Therefore, primary care doctors need to act as gatekeepers for medical resources. Given the large number and substantial heterogeneity of patients with DM, the aim of study was to develop several risk prediction models to facilitate primary care providers in identifying Chinese patients at higher risk of developing diabetic complications. This will allow interventions to be implemented to reduce the individual, societal and economic burden. The nomogram and chart can help inform clinicians regarding interventions based on overall risk of diabetic complications instead of only a single risk factor; in addition, these tools can also be used to educate, motivate and empower patients to prevent future diabetic complications. In terms of policy implications, misclassification may likely lead to excessive medical treatment, low cost-effectiveness in primary prevention and potentially unnecessary exposure to the risk of adverse drug effects. The accurate risk prediction model is particularly useful for screening programmes to inform decisions concerning service provision for DM primary care to achieve the maximum population health benefit.

We have also identified some potential improvements in the performance of existing risk prediction models. First, gender is a factor that is of concern in the analyses of risk factors and CVD/mortality because males are typically associated with a higher risk of CVD/mortality,<sup>28</sup> but statistical adjustment for gender is often insufficient to control for varying risk factor profiles and CVD/mortality incidence.<sup>29</sup> Second, there are possible interaction effects between age and risk factors on the CVD/mortality as the magnitude of the effect of specific risk factors such as LDL-C on the CVD/mortality may decrease with age.<sup>30</sup> Thus, the interaction term between age and risk factors should be considered when developing the risk prediction models. Third, many studies including our previous studies illustrated that there were curvilinear association (J or U shape) between HbA1c/SBP/BMI and the risk of CVD/mortality,<sup>31–33</sup> and thus, the quadratic term of such clinical parameters should be evaluated when developing the risk prediction models. Fourth, recent research suggested additional clinical parameters such as severity of renal impairment measured by eGFR and albumin/creatinine ratio, and variability of risk factors including HbA1c and SBP should be considered to enhance the performance of prediction models.<sup>34–37</sup> Finally, the feasibility of measuring predictors in a routine primary care setting should be considered when establishing risk prediction models. For example, some predictors such as

serum bicarbonate and phosphate included in an ESRD risk prediction model may not be available in routine clinical practice, especially in the primary care setting.<sup>38</sup> Accordingly, the rationale for inclusion of such non-standard parameters for risk prediction is often unclear, despite the inclusion of such predictors improved the model prediction accuracy. However, to develop a model with wide applicability in the real world, predictors should be selected to strike a balance between comprehensiveness and feasibility.

From our previous experience, there were several potential difficulties we may encounter during the development of risk prediction model. First, as anonymous data are extracted from the HA CMS database, there may be some variations of the operation definitions throughout this study. To avoid confusion and unnecessary misunderstanding, operational definitions (online supplementary file 1) clearly list the definitions of the clinical parameters, disease characteristics and treatment modalities according to clinical and professional advices. Moreover, it is important to ensure the quality of the data from data collection and correct use of data. Regular communications between our team and HA will be held to cross-validate the method of collecting raw data and the use of the data. Second, the outcome events in this study will rely on the diagnosis coding such as ICPC-2 and ICD-9-CM, which may cause misclassification bias. Due to the privacy policy, all extracted data will be anonymous, and thus, the cross-checking the records in the dataset for auditing purposes will not be feasible. However, in routine clinical practice in Hong Kong, clinicians in the clinical and hospital settings provide ICPC-2 and ICD-9-CM codes, respectively, for each episode of attendance.<sup>39–40</sup> Moreover, due to the heavily subsidised healthcare system in Hong Kong, patients with chronic diseases and serious complication, for example, MI, mostly treated in the HA public healthcare system. Therefore, our data from HA database should have captured nearly all CVD and ESRD outcomes of patients with DM who are managed in the HA primary care setting. Lastly, the missing data possibly exist in the dataset extracted from the database. To preserve the completeness and power of the data, multiple imputation will be used to deal with the missing data. Comparing with complete case analysis and single imputation, multiple imputation takes incomplete data into account to avoid unnecessary biases and producing more reliable and applicable risk prediction models.<sup>41</sup> Thus, multiple imputation has become one of the most widely used methods for handling missing data in medical research.<sup>42–43</sup> This approach was also used in the development of the CVD prediction model in the QRisk study.<sup>9–10</sup>

## CONCLUSION

There is a need for the development of 10-year risk prediction models based on the up-to-date population-based cohort for diabetic complications, particularly first total CVD. Chinese patients in primary care

will enable accurate risk stratification, better prioritisation of resources and more cost-effective interventions for patients with diabetes in primary care. They can also better inform and empower patients to prevent potential DM complications. At the health policy level, the results can inform decisions on service provision in the care of DM in primary care to achieve maximum population health benefit. The prediction models can also be used as an outcome measure on the potential benefits of complication prevention in clinical trials on DM interventions in primary care.

### Patient involvement

No patients were involved in setting the research question or the outcomes measures, designing the investigation or interpreting the data. There are no plans to involve patients in dissemination of the results.

### Author affiliations

<sup>1</sup>Department of Family Medicine and Primary Care, The University of Hong Kong, Hong Kong

<sup>2</sup>Department of Primary and Community Services, Hospital Authority Head Office, Hospital Authority, Hong Kong

<sup>3</sup>Department of Family Medicine and Primary Healthcare, Kowloon East Cluster, Hospital Authority, Hong Kong

<sup>4</sup>Department of Family Medicine & Primary Healthcare, Kowloon Central Cluster, Hospital Authority, Hong Kong

<sup>5</sup>Department of Family Medicine, New Territories East Cluster, Hospital Authority, Hong Kong

<sup>6</sup>Department of Family Medicine and Primary Healthcare, Hong Kong West Cluster, Hospital Authority, Hong Kong

<sup>7</sup>Department of Medicine, University of Hong Kong, Hong Kong

<sup>8</sup>School of Nursing, University of Hong Kong, Hong Kong

**Acknowledgements** The authors are most grateful to the Food and Health Bureau, HKSAR and the Hong Kong Hospital Authority, in particular chief of service in primary care in each cluster and Statistics and Workforce Planning Department at the Hong Kong Hospital Authority.

**Contributors** CLKL is the principal investigator of the study. CLKL and EYFW initially conceived the study. EYFW, EYTY, WYC, CSCF, DYT and CLKL helped with the design and implementation of the programmes, coordination of the study, drafted and revised the manuscript. EYFW, EYTY, WYC, CSCF, RLPK, DVKC, KHC, EM-TH, WWST, KCBT, DYT and CLKL revised the manuscript. All authors approved the final version.

**Funding** This study has been funded by the Health and Medical Research Fund, Food and Health Bureau, HKSAR (Project no: 14151181).

**Disclaimer** No funding organisation had any role in the design and conduct of the study; collection, management, analysis and interpretation of the data; and preparation of the manuscript. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

**Competing interests** None declared.

**Patient consent** Not required.

**Ethics approval** Ethics approval of this study was granted by the Institutional Review Board of the University of Hong Kong/ Hospital Authority Hong Kong West (UW 15-258).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

### REFERENCES

- International Diabetes Federation. *IDF diabetes atlas*. 7 edn. Brussels: International Diabetes Federation, 2015.
- Chamnan P, Simmons RK, Sharp SJ, *et al*. Cardiovascular risk assessment scores for people with diabetes: a systematic review. *Diabetologia* 2009;52:2001–14.
- American Diabetes Association. Standards of medical care in diabetes--2010. *Diabetes Care* 2010;33(Suppl 1):S11–61.
- Stone NJ, Robinson JG, Lichtenstein AH, *et al*. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* 2014;63(25 Pt B):2889–934.
- Jiao F, Fung CS, Wan YF, *et al*. Long-term effects of the multidisciplinary risk assessment and management program for patients with diabetes mellitus (RAMP-DM): a population-based cohort study. *Cardiovasc Diabetol* 2015;14:105.
- Fung CS, Chin WY, Dai DS, *et al*. Evaluation of the quality of care of a multi-disciplinary risk factor assessment and management programme (RAMP) for diabetic patients. *BMC Fam Pract* 2012;13:116.
- Wilson PW, D'Agostino RB, Levy D, *et al*. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97:1837–47.
- D'Agostino RB, Wolf PA, Belanger AJ, *et al*. Stroke risk profile: adjustment for antihypertensive medication. The Framingham Study. *Stroke* 1994;25:40–3.
- Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017;357:j2099.
- Hippisley-Cox J, Coupland C, Vinogradova Y, *et al*. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008;336:1475–82.
- Mancia G, Fagard R, Narkiewicz K, *et al*. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J Hypertens* 2013;31:1281–357.
- Wan EYF, Fong DYT, Fung CSC, *et al*. Development of a cardiovascular diseases risk prediction model and tools for Chinese patients with type 2 diabetes mellitus: a population-based retrospective cohort study. *Diabetes Obes Metab* 2018;20.
- Wan EYF, Fong DYT, Fung CSC, *et al*. Prediction of new onset of end stage renal disease in Chinese patients with type 2 diabetes mellitus - a population-based retrospective cohort study. *BMC Nephrol* 2017;18:257.
- Wan EYF, Fong DYT, Fung CSC, *et al*. Prediction of five-year all-cause mortality in Chinese patients with type 2 diabetes mellitus - A population-based retrospective cohort study. *J Diabetes Complications* 2017;31:939–44.
- Diabetes Drafting Group. Prevalence of small vessel and large vessel disease in diabetic patients from 14 centres. The World Health Organisation Multinational Study of Vascular Disease in Diabetics. Diabetes Drafting Group. *Diabetologia* 1985;28:615–40.
- Chi ZS, Lee ET, Lu M, *et al*. Vascular disease prevalence in diabetic patients in China: standardised comparison with the 14 centres in the WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia* 2001;44:S82–6.
- Liu JJ, Lim SC, Yeoh LY, *et al*. Ethnic disparities in risk of cardiovascular disease, end-stage renal disease and all-cause mortality: a prospective study among Asian people with Type 2 diabetes. *Diabet Med* 2016;33:332–9.
- Burrows NR, Li Y, Williams DE. Racial and ethnic differences in trends of end-stage renal disease: United States, 1995 to 2005. *Adv Chronic Kidney Dis* 2008;15:147–52.
- McClellan WM, Warnock DG, Judd S, *et al*. Albuminuria and racial disparities in the risk for ESRD. *J Am Soc Nephrol* 2011;22:1721–8.
- Kong AP, Xu G, Brown N, *et al*. Diabetes and its comorbidities--where East meets West. *Nat Rev Endocrinol* 2013;9:537–47.
- Forouhi NG, Sattar N. CVD risk factors and ethnicity--a homogeneous relationship? *Atheroscler Suppl* 2006;7:11–19.
- Byrne CD, Wild SH. *The metabolic syndrome*: John Wiley & Sons, 2011.
- Zhao W, Chen J. Implications from and for food cultures for cardiovascular disease: diet, nutrition and cardiovascular diseases in China. *Asia Pac J Clin Nutr* 2001;10:146–52.
- Steyerberg EW, Eijkemans MJ, Van Houwelingen JC, *et al*. Prognostic models based on literature and individual patient data in logistic regression analysis. *Stat Med* 2000;19:141–60.



25. Stevens RJ, Kothari V, Adler AI, *et al.* The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56). *Clin Sci* 2001;101:671–9.
26. Kothari V, Stevens RJ, Adler AI, *et al.* UKPDS 60: risk of stroke in type 2 diabetes estimated by the UK Prospective Diabetes Study risk engine. *Stroke* 2002;33:1776–81.
27. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285–93.
28. Mozaffarian D, Benjamin EJ, Go AS, *et al.* Heart disease and stroke statistics--2015 update: a report from the American Heart Association. *Circulation* 2015;131:e29–322.
29. Kelly BB, Fuster V. *Promoting cardiovascular health in the developing world: a critical challenge to achieve global health*: National Academies Press, 2010.
30. Lewington S, Whitlock G, Clarke R, *et al.* Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. *Lancet* 2007;370:1829–39.
31. Wan EYF, Fung CSC, Wong CKH, *et al.* Association of Hemoglobin A1c Levels With Cardiovascular Disease and Mortality in Chinese Patients With Diabetes. *J Am Coll Cardiol* 2016;67:456–8.
32. Wan EY, Fung CS, Fong DY, *et al.* A curvilinear association of body mass index with cardiovascular diseases in Chinese patients with type 2 diabetes mellitus - a population-based retrospective cohort study. *J Diabetes Complications* 2016;30:1261–8.
33. Wan EYF, Yu EYT, Fung CSC, *et al.* Do we need a patient-centered target for systolic blood pressure in hypertensive patients with type 2 diabetes mellitus? *Hypertension* 2017;70:1273–82.
34. Fox CS, Matsushita K, Woodward M, *et al.* Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without diabetes: a meta-analysis. *The Lancet* 2012;380:1662–73.
35. Matsushita K, Coresh J, Sang Y, *et al.* Estimated glomerular filtration rate and albuminuria for prediction of cardiovascular outcomes: a collaborative meta-analysis of individual participant data. *Lancet Diabetes Endocrinol* 2015;3:514–25.
36. Wan EY, Fung CS, Yu EY, *et al.* Association of visit-to-visit variability of systolic blood pressure with cardiovascular disease and mortality in primary care chinese patients with type 2 diabetes-a retrospective population-based cohort study. *Diabetes Care* 2017;40:dc161617.
37. Wan EY, Fung CS, Fong DY, *et al.* Association of variability in hemoglobin A1c with cardiovascular diseases and mortality in Chinese patients with type 2 diabetes mellitus - A retrospective population-based cohort study. *J Diabetes Complications* 2016;30:1240–7.
38. Tangri N, Stevens LA, Griffith J, *et al.* A predictive model for progression of chronic kidney disease to kidney failure. *JAMA* 2011;305:1553–9.
39. Wong MC, Jiang JY, Tang JL, *et al.* Health services research in the public healthcare system in Hong Kong: an analysis of over 1 million antihypertensive prescriptions between 2004-2007 as an example of the potential and pitfalls of using routinely collected electronic patient data. *BMC Health Serv Res* 2008;8:138.
40. Fung V, Cheung N, Szeto K, *et al.* *Hospital authority clinical vocabulary table: the past, the present, and the future. hospital authority clinical vocabulary table: the past, the present, and the future*: AHIMA, American Health Information Management Association, 2004.
41. Royston P. Multiple imputation of missing values. *Stata Journal* 2004;4:227–41.
42. Wan EYF, Fung CSC, Yu EYT, *et al.* Effect of multifactorial treatment targets and relative importance of hemoglobin A1c, blood pressure, and low-density lipoprotein-cholesterol on cardiovascular diseases in chinese primary care patients with type 2 diabetes mellitus: a population-based retrospective cohort study. *J Am Heart Assoc* 2017;6:e006400.
43. Wan EY, Fong DY, Fung CS, *et al.* Incidence and predictors for cardiovascular disease in Chinese patients with type 2 diabetes mellitus - a population-based retrospective cohort study. *J Diabetes Complications* 2016;30:444–50.