APPRAISALS IN META-JOURNAL HOUR 14

By Nurfaizah, BH Chew and SQ Yew

The paper:

DEVELOPMENT OF A CLINICAL RISK SCORE PREDICTION TOOL FOR 5-, 9-, AND 13-YEAR RISK OF DEMENTIA

Why was this study conducted?

Dementia is the seventh leading cause of death and no effective treatment yet existed. An early intervention on modifiable risk factors of dementia could help prevent or delay its progression. The risk score model is a simple and convenient method for the general population to assess the probability of diseases using preclinical risk factors. This longitudinal prospective cohort study was conducted to develop a discriminative risk score model for the general population and predicts the 5-, 9-, and 13-year individual dementia risk for men and women. This study used a large sample size, risk factors of dementia that are commonly used in general practice, and risk stratification according to gender. These advantages highlighted the importance of developing this new prediction model of dementia as compared to the existing dementia risk score models.

YouTube

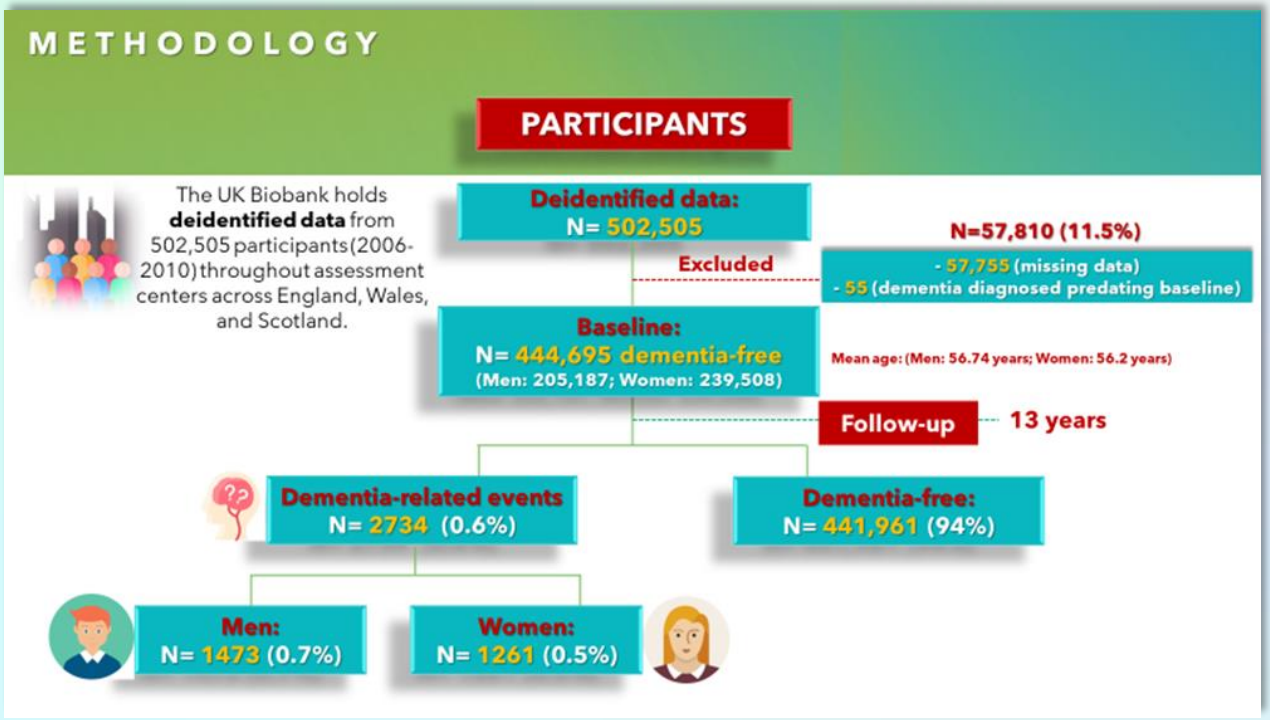
Watch the video recording on:

Click [HERE] and don't forget to subscribe to our channel!

How was it done?

Participants:

This study used a large UK population was conducted between March 13, 2006, and October 1, 2010. Data analysis was performed from June 7 to September 15, 2021. The UK Biobank holds de-identified data from 502,505 participants throughout assessment centers across England, Wales, and Scotland. After 57,755 participants were excluded due to missing data and dementia diagnosis predating the baseline, a total of 444,695 dementia-free participants (205 187 men; mean [SD] age, 56.74 [8.18] years; and 239 508 women; mean [SD] age, 56.20 [8.01] years) at baseline were included. About 0.6% (N=2734) individuals displayed dementia-related events at follow-up and were categorized into dementia group which included 0.7% (N= 1473) men and 0.5% (N=1261) women. A total of 441961 participants did not have a diagnosis of dementia or dementia-related events at follow-up (dementia-free group).



Primary outcome:

Individual analyses of time end points were concluded on the first dementia diagnosis during the follow-up period. Dementia diagnoses were established according to the International Statistical Classification of Diseases, Version 10 (ICD-10) terms from UK Biobank data field 41 270 (ICD-10 codes F01-F04 and G30), which included Alzheimer disease, vascular dementia, unspecified dementia, organic amnesic syndrome, and dementia in other diseases classified elsewhere.

Exposure factors:

EXPOSURE FACTORS

Demographic
- Age
- Ethnicity
- Educational level
- BMI
- Employment

Lifestyle
- Smoking
- Frequency of alcohol consumption
- Walking pace

Sleep Phenotype
- Sleep duration
- Early awakening
- Napping
- Sleeplessness or insomnia
- Snore

Comorbidities
- Respiratory disease
- Cancer
- Cerebrovascular disease
- Diabetes
- Cardiovascular disease
- Hypertension

All measurements were taken at baseline. Participants were asked the touchscreen questions.

How the study developed a discriminative risk score model to predict the 5-, 9-, and 13-year individual dementia risk for men and women?

The point risk score prediction model was developed using optimal exposure factors of dementia that are practical and readily available to healthcare professionals. The flow process of developing the point risk score prediction model as follows:

The data were divided into **training** and **testing data sets** to **establish** and **validate** a prediction model separately.

The Least Absolute Shrinkage and Selection Operator (LASSO) regression and forward and backward stepwise multivariate Cox proportional hazards regression – **To identify potential optimal predictors that are readily available to healthcare professionals and develop an optimal risk prediction model**.

The **relative contribution of each risk predictor to the dementia** population was calculated using a population-attributable fraction (PAF).

A **point risk score model** that stratifies individuals for **5-, 9-, and 13-year risk of dementia** was developed.

Feature Selection

1st

Mann-Whitney test

-**To exclude the features without significant differences** between dementia and dementia-free groups ($p<0.05$).

2nd

LASSO regression

-The analysis was conducted with **10-fold cross-validation to select the most important predictors** from the **training data set.**

3rd

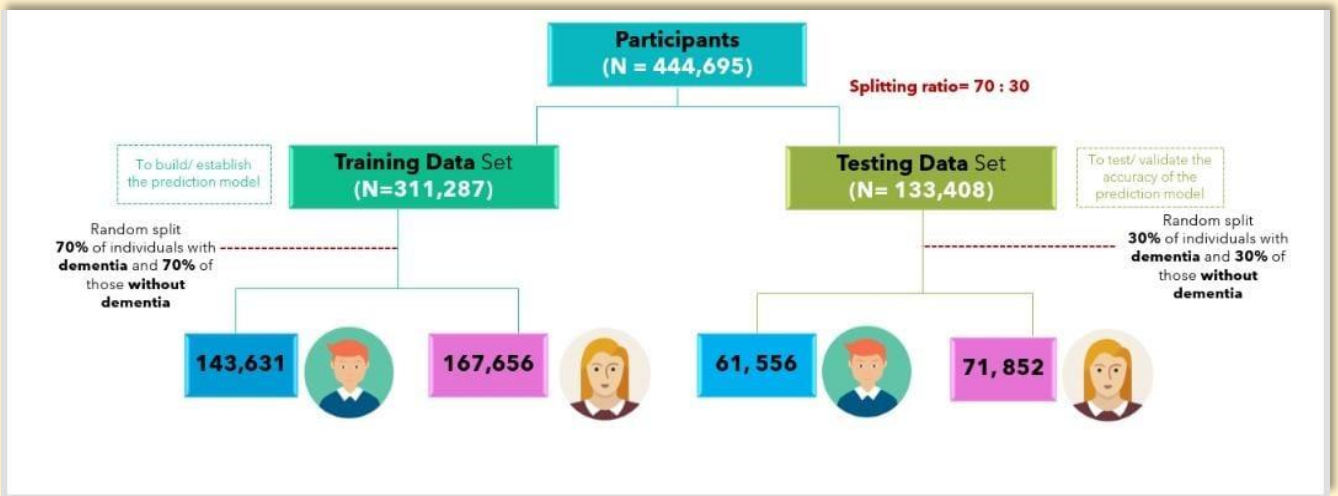
Multivariate Cox proportional hazards regression model

-The remaining features were evaluated for **associations between exposure factors and dementia risks.**

-Both forward and backward stepwise analyses were further used to identify the **optimal predictors.**

Point Risk Score Prediction Model

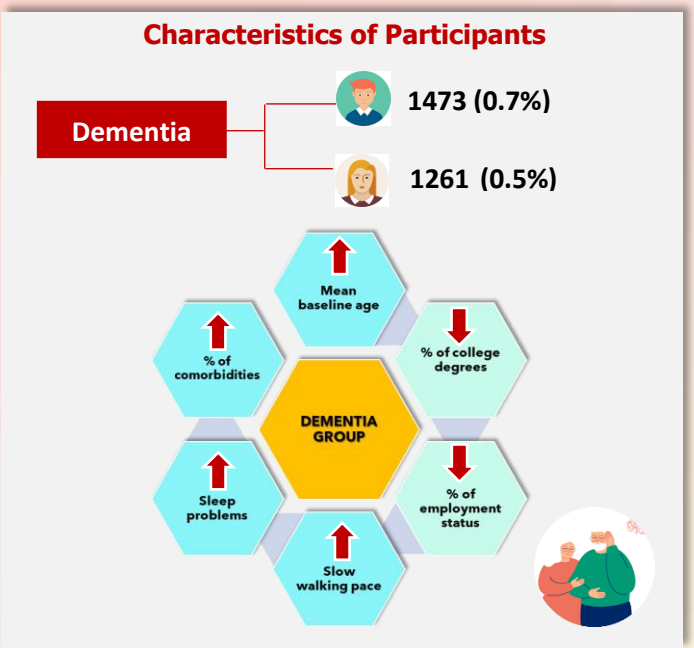
Validation of Risk Prediction Model



What was the finding?

Table 1. Summary of the outcomes

OUTCOMES	MEN	WOMEN
Occurrence of dementia (13 years followed-up)	0.7%	0.5%
C-statistic (Final multivariate Cox proportional hazards regression model)	Training:0.86	Training: 0.85
	Testing:0.85	Testing: 0.87
Shared modifiable independent risk factors	✓ Age ✓ No paid employment status ✓ Respiratory disease ✓ Cerebrovascular disease ✓ Diabetes ✓ Hypertension	
Weighted PAF for all independent risk factors	31.7 %	53.35%
Total point score of the risk score model	−18 to 30	−17 to 30
Prediction accuracy	9- year: 97.6% 13-year: 100%	9- year: 99.6% 13-year: 100%



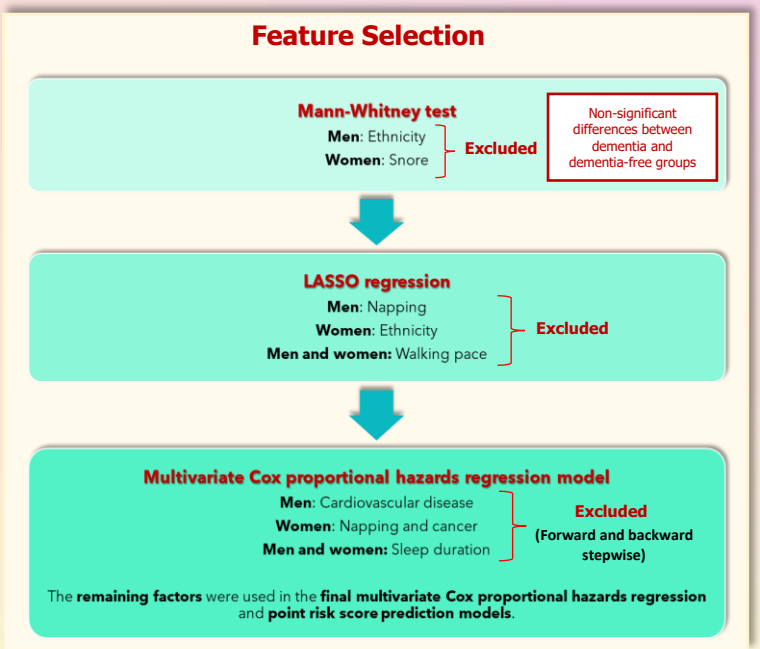


Table 2. Area under the Curve (AUC) of Training and Testing Data Sets in Men and Women

OUTCOMES	MEN		WOMEN	
AREA UNDER THE CURVE (AUC)	TRAINING DATA SET	TESTING DATA SET	TRAINING DATA SET	TESTING DATA SET
5-year	0.86	0.85	0.87	0.91
9-year	0.86	0.86	0.87	0.87
13-year	0.86	0.83	0.84	0.87

Table 3. Independent risk factors of dementia shared by men and women and exclusive risk predictor for men and women

EXPOSURE FACTORS	TESTING DATA SET	
	Men	Women
Age	HR, 1.14; 95% CI, 1.12-1.17	HR, 1.17; 95% CI, 1.14-1.20
No paid employment status	HR, 1.93; 95% CI, 1.47-2.53	HR, 1.97; 95% CI, 1.41-2.76
Respiratory disease	HR, 1.87; 95% CI, 1.53-2.30	HR, 1.82; 95% CI, 1.46-2.27
Cerebrovascular disease	HR, 3.40; 95% CI, 2.66-4.35	HR, 4.04; 95% CI, 3.01-5.43
Diabetes	HR, 1.61; 95% CI, 1.26-2.06	HR, 2.06; 95% CI, 1.53-2.78
Hypertension	HR, 1.39; 95% CI, 1.13-1.72	HR, 1.36; 95% CI, 1.07-1.72
	Men	Women
Exclusive risk predictor	**Sometimes sleepiness** (HR, 1.29; 95%CI, 1.05-1.58) had a 29% higher risk of dementia than those reporting sleepiness never or rarely.	**Low educational level** (HR, 1.43; 95%CI, 1.08-1.90) **Sleepiness often or all of the time** (HR, 1.86; 95%CI, 1.19-2.90) increased the risk of dementia.

Point Risk Score Prediction Model and Validation

5 Steps to Calculate the Point Score:

Step 1: Reference Value (Middle value of each category)
Continuous variable: Age in men (Reference: middle category- 50 to 55 years)
Categorical variable: Reference group (e.g. degree) = reference value.

Step 2: Regression Coefficient
The regression coefficient of each exposure factor in the Cox proportional hazards regression.

Step 3: Distance
The regression coefficient and reference value are used to calculate the **difference** between the reference and non-reference categories for each exposure factor.

E.g. Age (Men):

- Value of reference group =53
- Regression coefficient = 0.15
- Ref. value for men aged 36 – 40 years = 38
- **Difference: (38-53) x 0.15 = -2.25**

Step 4: Distance Constant
A distance constant corresponding to the change of 1 score for each exposure factor was set.

Age (men):
Regression coefficient x 1.5
=0.15 x 1.5
= **0.23**

Step 5: Calculation of Point Score
Point score = Difference/ distance constant

Age (men):
-2.25/ 0.23 = - 9.78 ~ **-10**

Table 4. The prevalence, commonality, and weighted Population-attributable Fraction (PAF) for all risk factors of dementia

EXPOSURE FACTORS	DEMENTIA (%)	
All risk factors	**Men**	**Women**
	31.7	**53.35**
Socioeconomic adversity ▪Non degree ▪Not paid	13.38	27.35
Comorbidities ▪Respiratory disease ▪Cerebrovascular disease ▪Diabetes ▪Hypertension ▪Cardiovascular disease (*women only*)	15.22	18.77
Others ▪Sleepiness ▪Underweight ▪Low frequency of alcohol consumption	2.84 0.07 0.20	2.33 0.15 0.48

How much can we take out from this research/paper?

This paper reported a meaningful endeavour to improvise on existing dementia prediction models by selecting clinically more available variables as the predictors. Having a large number of people diagnosed with dementia in the UK Biobank database is another feasible factor for the study. However, it is puzzling that the investigators were all from China and none was from or affiliated with an institution in the UK. This may affect contextualisation of the findings, and the process and justification for conducting the study was not explained. Under the acknowledgement, it appears that UK Biobank Resource is accessible by a certain application, but no further details are provided.

Beside the mismatch between the background of the investigator-researchers and the study settings, there are some mis-labelling of study designs and inadequate descriptions of the study. The paper describes a retrospective case-control study stratified by the gender with categorised outcomes at specified time points. The data was analysed using multivariate Cox proportional hazards regression instead of logistic regression as expected without explanation. The authors did not explain the justification of calculating risk of dementia specifically at 5-, 9-, and 13 years. These 5-, 9- and 13-year time points are not defined whether the number of dementia diagnosed at the earlier time points were accumulative from the preceding years, and/or excluded from the latter time points. This causes the actual number of people with dementia at 5- and 9-year to be unknown, and inability to judge whether the 5-year is equally credible as the 13-year and if so why the need for the 13-year prediction.

Another mis-labelling is the diagnostic study by the authors of this study which is rather a prognostic study or possibly an aetiologic study, or even a mix of these two. It is less of a diagnostic study because the diagnosis of dementia is generally a clinical diagnosis (cross-sectional in study design) where the age in year would have included the effect of the duration of the time points. The major lacking in this study is the lack of description and support for the diagnosis of dementia whether it was done and recorded in a 'gold' standard manner, and the accuracy of many self-reported predictors retrieved from the database. The self-reported comorbidities rather than being ascertained using patient's registries may lead to respondent bias, especially among those participants with lower educational levels.

It is not explained or provided proper justification on the rigorous selection of predictors through processes of Mann-Whitney test, LASSO regression and then multivariate Cox proportional hazards regression modelling. This statistical strategy is usually done in aetiology studies (explanatory model where confounders to the causal factor/s are excluded) than in diagnostic or prognostic studies (predictive model where all important determinants/predictors are included, and no confounders are excluded). Without any clarification on the statistical strategy used, it could happen that over-rigorous selection of prediction results in the over-performing prediction model. Another inadequate almost absence is the description of penalization of the prediction models. The MUST-NOT forget of all these is the model/s has yet to be externally validated. It ended in an internal validation using the same source database that has been split into the testing dataset and reported >95% accuracy. However, the weighted PAF of all modifiable risk factors for dementia accounted for 31.7% in men and 53.4% in women. Again, this estimation was not explained, neither the weighting procedure for the PAF.

With the above study designs in mind, the interpretation of the results is properly guided and cautioned. Of the many predictors, it is logical to observed strong predictors being the age and having a cerebrovascular disease, and others include engaging in employment or occupation that is more cognition-demanding, staying physically healthy, better sleep quality, and stay away from smoking and excessive alcohol. It was reported that the total point score ranged from −18 to 30 in men and from −17 to 30 in women, whereby higher scores correspond to higher risks (in percentage) of developing dementia. However, the authors did not discuss the clinical significance/implication of such scoring (i.e, what is the recommended cut-off score above which individuals in the general population require rigorous intervention to prevent dementia). In the scarcity of "effort, money, and time", surely clinicians would like to know who should be prioritised for interventions in clinical practice. Nevertheless, until this prediction model is externally validated in own society and setting, it may be a good piece of scientific evidence for the UK Biobank's population (people at hospitalisation in England, Wales, and Scotland) and not any where else.

Reference

1. Ren L, Liang J, Wan F, Wang Y, Dai X-J. Development of a clinical risk score prediction tool for 5-, 9-, and 13-year risk of dementia. JAMA Network Open.2022;5(11):e2242596. doi:10.1001/jamanetworkopen.2022.42596.