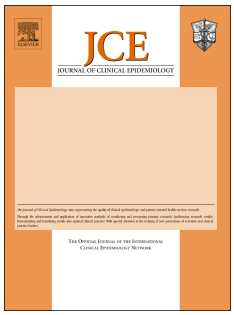

Elicited Clinician Knowledge Did Not Improve Dementia Risk Prediction in Individuals with Mild Cognitive Impairment

Meng Wang, Thierry Chekouo, Zahinoor Ismail, Nils D. Forkert, David B. Hogan, Aravind Ganesh, Richard Camicioli, Dallas Seitz, Michael J. Borrie, Ging-Yuek Robin Hsiung, Mario Masellis, Paige Moorhouse, Carmela Tartaglia, Eric E. Smith, Tolulope Sajobi, for the Alzheimer's Disease Neuroimaging Initiative

PII: S0895-4356(23)00052-5

DOI: https://doi.org/10.1016/j.jclinepi.2023.03.009

Reference: JCE 11038

To appear in: *Journal of Clinical Epidemiology*

Received Date: 10 January 2023

Revised Date: 10 March 2023

Accepted Date: 13 March 2023

Please cite this article as: Wang M, Chekouo T, Ismail Z, Forkert ND, Hogan DB, Ganesh A, Camicioli R, Seitz D, Borrie MJ, Hsiung G-YR, Masellis M, Moorhouse P, Tartaglia C, Smith EE, Sajobi T, for the Alzheimer's Disease Neuroimaging Initiative, Elicited Clinician Knowledge Did Not Improve Dementia Risk Prediction in Individuals with Mild Cognitive Impairment, *Journal of Clinical Epidemiology* (2023), doi: https://doi.org/10.1016/j.jclinepi.2023.03.009.

Elicited Clinician Knowledge Did Not Improve Dementia Risk Prediction in Individuals with Mild Cognitive Impairment

Authors: Meng Wang[1,2], Thierry Chekouo[3], Zahinoor Ismail[1,2,4], Nils D. Forkert[2,5], David B. Hogan[1,2], Aravind Ganesh[2], Richard Camicioli[6], Dallas Seitz[4], Michael J. Borrie[7], Ging-Yuek Robin Hsiung[8], Mario Masellis[9], Paige Moorhouse[10], Carmela Tartaglia[11], Eric E. Smith[2], and Tolulope Sajobi[1,2] for the Alzheimer's Disease Neuroimaging Initiative*

[1]Department of Community Health Sciences & O'Brien Institute of Public Health, University of Calgary, Canada
[2]Department of Clinical Neurosciences & Hotchkiss Brain Institute, University of Calgary, Canada
[3]Division of Biostatistics, School of Public Health, University of Minnesota, US
[4]Department of Psychiatry, University of Calgary, Canada
[5]Department of Radiology, University of Calgary, Canada
[6]Department of Medicine, Division of Neurology, University of Alberta, Canada
[7]Department of Medicine, Division of Geriatric Medicine, Western University, Canada
[8]Division of Neurology, Department of Medicine, the University of British Columbia, Canada
[9]Sunnybrook Health Sciences Centre, Toronto, Canada
[10]Division of Geriatric Medicine, Department of Medicine, Dalhousie University, Canada
[11]Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Canada

*Patient data source used in preparation of this manuscript were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this manuscript. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Corresponding Author:
Dr. Tolulope Sajobi
Department of Community Health Sciences
Cumming School of Medicine
University of Calgary
3280 Hospital Drive NW
Calgary T2N 4Z6 AB
Tel: +1(403)-210-8586
Email: ttsajobi@ucalgary.ca

Table:3 Figures: 2
Declaration of interest: none

ABSTRACT

Objective: This study aims to develop and validate a Bayesian risk prediction model that combines research cohort data with elicited expert knowledge to predict dementia progression in people with mild cognitive impairment (MCI).

Study Design and Setting: This is a prognostic risk prediction modeling study based on cohort data (Alzheimer's Disease Neuroimaging Initiative [ADNI]; n=365) of research participants with MCI and elicited expert data. Bayesian Cox models were used to combine expert knowledge and ADNI data to predict dementia progression in people with MCI. Posterior distributions were obtained based on Gibbs sampler and the predictive performance was evaluated using ten-fold cross-validation via c-index, integrated calibration index (ICI), and integrated brier score (IBS).

Results: 365 people with MCI were included, mean age was 73 years (SD=7.5) and 39% developed dementia within 3 years. When expert knowledge was incorporated, the c-index, ICI, and IBS values were 0.74 (95% CI 0.70-0.79), 0.06 (95% CI 0.05-0.08), and 0.17 (95% CI 0.14-0.19), respectively. These were similar to the model without expert knowledge data.

Conclusion: The addition of expert knowledge did not improve model accuracy in this ADNI sample to predict dementia progression in individuals with MCI.

Key words: Bayesian; prior; elicitation; dementia; MCI; prediction

Running Title: A Bayesian Dementia Risk Prediction after Expert Elicitation

INTRODUCTION

Dementia (major neurocognitive disorder) is typically preceded by mild cognitive impairment (MCI, also called mild neurocognitive disorder), a syndrome associated with objective impairment in cognition but without major functional disability or loss of independence. While the etiology and outcome of MCI are variable, a substantial number of patients will progress to dementia.[1]

Due to the heterogeneity of the MCI population, it is difficult in clinical practice to provide an individualized estimated risk of progression to dementia for a person with MCI. Clinical decision support tools, like risk scores, can help clinicians predict the probability of MCI progressing to dementia by synthesizing the effects of multiple predictors using underlying data and models.[2] However, there are few dementia risk scores for individuals with MCI that are suitable for application in routine clinical practice. This may be due to challenges with feasibility, lack of validation, or unavailability of the included predictors. When counseling regarding the risk for progression in MCI, clinicians require shared guidelines, experience, validated approaches, and heuristics to determine an individual's prognosis. Structured expert elicitation can help experts to express their knowledge in a quantitative form and reduce biases in the process.[3–5] Although expert knowledge can feed directly into decision making itself, if there is data available, combining the expert knowledge with modeling of the data is usually preferred.[3,4,6] Informative prior elicitation has a fundamental role in Bayesian inference that enables one to make use of expert knowledge and historical data.[7] In contrast, non-informative priors, also called vague, flat, or diffuse priors, do not make use of real prior information and plays minimal role in the final inference.[8]

This study aims to develop and validate a Bayesian risk prediction model that integrates patients' data with elicited expert knowledge to predict dementia in people with MCI.

METHODS

Data source

We used data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu). ADNI is a longitudinal multicenter study that started in 2004, with a primary goal of testing different biomarkers for the progression of MCI and early AD.[9] ADNI is a research cohort, participants included in ADNI were between 55-90 years of age and English or Spanish speakers in the US and Canada. There were 598 individuals with MCI in ADNI, defined as having a Mini-Mental State Examination (MMSE)[10] score between 24-30, reported subjective complaints, objective memory deficits, and a Clinical Dementia Rating[11] score of 0.5. Only patients with complete data (n=365) of relevant predictors at ADNI baseline or screening visits were considered in this study (matched with the listed variables used in expert elicitation shown in Supplemental Table S1). Supplemental Figure S1 describes the identification of the study cohort.

Outcome and Predictor measures

Our study outcome was the time to all-cause dementia over a three-year period for participants with MCI at ADNI enrollment. In ADNI, AD was diagnosed using the National Institute of Neurological, Communicative Disorders, and Stroke and Alzheimer's Disease and Related Disorders Association criteria for possible or probably AD[12] for all participants. A total of 34 predictors were considered in our analysis, including variables identified in our structured expert elicitation[13,14] and available in ADNI baseline and screening visit. All predictors were

based on data files downloaded from ADNI/LONI. Full operational definitions for the predictors measured in ADNI were provided in Supplemental Appendix A and Table S1. All datasets were downloaded on or before July 4th, 2022.

Expert elicitation

The structured expert elicitation methodology was used to elicit dementia progression risk for each possible predictor, from a total of 11 clinician experts. Ten experts (six neurologists, two geriatricians, and two psychiatrists) fully participated in the introductory meetings, elicitation surveys, and discussion meetings. Details on a protocol for this study and the expert elicitation process and results have been published elsewhere.[13,14] In our expert elicitation survey (Supplemental Appendix B), experts were first asked to rank the potential predictors for MCI to dementia progression and rate the importance of each predictor on a seven-point Likert scale. Next, for each predictor ranked as at least somewhat important by one or more experts, the experts were asked to imagine that the predictor was present in a person at low risk and then to give: 1) the lowest plausible annual progression rate to dementia if that predictor was present, 2) the highest annual progression rate if that predictor was present, 3) the best guess as to the actual annual progression rate, and 4) the confidence, expressed a percentage, that the true annual progression rate lay within the lowest and high rates previously given.

Prior distribution summaries were obtained based on the above elicited information from ten experts. Three-year dementia risk was estimated based on the annual progression rate estimated by experts, by assuming a constant hazard for each year. Standardized 90% confidence intervals were obtained using linear extrapolation based on the lowest, highest, and best guess estimates from experts, as well as how confident they were.[15,16] The normal priors for the regression coefficients were assumed in this study, which is the most common choice for

informative priors.[7] Four approaches were explored for obtaining the mean of the normal distribution for each regression coefficient, including using the mean, the median, the minimum, and the maximum of experts' best guesses. Two different methods were explored for obtaining the variance of the normal distribution for each regression coefficient, including using the variance of the ten experts' best guesses and the average of the variance from each expert's estimates based on the calculated standardized 90% confidence intervals. The detailed process of prior derivations is included in Supplemental Appendix C.

Analysis

Two types of Bayesian Cox regression models with piecewise constant baseline hazards (details in Supplemental Appendix D) were fitted: one with expert knowledge data (informative priors) and the other one without expert knowledge data (non-informative priors). Normal priors for the regression coefficients and gamma prior for the baseline hazards were used, respectively. A Bayesian Cox regression model with informative priors was used for combining expert knowledge with patient data. To build informative priors for the regression coefficients, we used elicited data to approximate the mean and variance of normal prior distributions. The priors for the hazards and regression coefficients were assumed to be independent. Because of multiple experts, a linear pooling method with equal weights was used for aggregation.[3,17] For Bayesian Cox regression with non-informative priors, we imposed a normal prior distribution with mean 0 and variance 10^6 for each regression coefficient. For each constant baseline hazard, both the shape and inverse-scale parameters on the gamma priors were set at 10^{-4}, which provides reasonably non-informative priors.[18] The Gibbs sampler using the adaptive rejection Metropolis sampling algorithm[19] was used to sample from the full conditionals (the number of burn-in iterations = 10000; the number of iterations after burn-in ranged from 100,000 to 400,000;

thinning ranged from 20 to 80; starting values of the Markov chains were based on maximum likelihood estimates and prior information). The convergence of Markov chain Monte Carlo algorithm was evaluated based on the posterior autocorrelations and effective sample sizes. The predictive performance was evaluated based on ten-fold cross-validation *via* Harrell's concordance index (c-index), integrated brier score (IBS), and integrated calibration index (ICI) for all models. The ICI is a calibration metric for survival outcomes,[20] with smaller ICI value indicates a better calibrated model. The IBS uses a squared loss function,[21] and a smaller IBS value indicates a better combination of discrimination and calibration.

Sensitivity analyses were conducted to assess the robustness of the analysis results, by varying the number of predictors used in the models and approaches for aggregating elicited data from experts. In the first sensitivity analysis (condition 1) three experts with inconsistent estimates were excluded. Inconsistence was defined as when one of the following is true: the highest estimate was lower than a best guess, a best guess was lower than the lowest estimate, or an estimated progression risk was lower than baseline risk for risk factors that the expert agreed on. In addition, variables were excluded if any expert considered the variable to be less than somewhat important (19 predictors remained). In the second sensitivity analysis (condition 2), a Bayesian Cox regression was first fitted to the whole sample and then significant predictors were selected based on posterior summaries. In the third sensitivity analysis (condition 3), the top ten important predictors were selected based on ten experts' ranking of the predictors using a seven-point Likert scale. For each condition, Bayesian models were refitted, and model performance was obtained using ten-fold cross-validation. All analyses were conducted in SAS (9.4)[22] and R (4.2).[23] The conduct and reporting of this study followed the transparent reporting of a

multivariable prediction model for individual prognosis or diagnosis[5,6] (supplemental Appendix F).

RESULTS

Table 1 and supplemental Table S3 show that out of 365 research subjects with MCI, mean age was 73 years (SD=7.5), and 142 (39%) developed dementia within 3 years. Supplemental Table S2 shows the comparisons between individuals included and excluded because of missing data. Figure 1 shows that people with positive cerebrospinal fluid (CSF) or fluorodeoxyglucose-positron emission tomography (FDG-PET) findings had significantly higher risk of developing dementia, based on crude comparisons.

Figure 2 shows the prior distribution for the two most important variables ranked by the ten experts along with the average distribution. The impact of age on the progression of dementia differed between experts with most experts indicating that they were very confident on how much age contributes to dementia progression (very narrow distribution for almost every expert). The best guesses on the regression coefficient for age (per year of change) ranged from 0.01 (95% CI 0.004-0.02) to 0.08 (95%CI 0.06-0.11), which implies that experts think that the estimated hazard ratio ranged from 1.05 (95%CI 1.02-1.09) to 1.49 (95%CI 1.35-1.72) per 5 years of aging for people with MCI, holding other predictors constant. The best guesses on the regression coefficient for CSF or FDG-PET positivity ranged from 0.66 (95%CI 0.44-0.89) to 3.0 (95%CI 2.48-3.51), which suggests that experts anticipate that the estimated hazard ratio ranged from 1.94 (95%CI 1.55-2.43) to 20.0 (95%CI 11.9-33.6) for people with positive CSF or FDG-PET findings, compared with people without positive CSF or FDG-PET findings holding other predictors constant.

Table 1 Sample characteristics

	N=365
Age, mean (SD)	73.17 (7.50)
Female, n (%)	140 (38.4)
Education in Years, median [Q1-Q3]	16 [14, 18]
High education, n (%)	246 (67.4)
Married or common-law, n (%)	274 (75.1)
Hypertension, n (%)	171 (46.8)
Parkinsonism, n (%)	1 (0.3)
Stroke, n (%)	6 (1.6)
Diabetes, n (%)	28 (7.7)
CVD, n (%)	57 (15.6)
OSA, n (%)	35 (9.6)
VB12 deficiency, n (%)	6 (1.6)
TBI, n (%)	14 (3.8)
Dyslipidemia, n (%)	146 (40.0)
Hypothyroidism, n (%)	55 (15.1)
Gait, n (%)	33 (9.0)
Kidney disease, n (%)	167 (45.8)
Liver disease, n (%)	11 (3.0)
Impaired hearing, n (%)	35 (9.6)
Apathy, n (%)	59 (16.2)
Psychosis, n (%)	9 (2.5)
impulsivity disinhibition, or agitation, n (%)	81 (22.2)
Depression, n (%)	111 (30.4)
Sleep problem, n (%)	52 (14.2)
CSF or FDG-PET positive, n (%)	256 (70.1)
Total WMH volume, ml, mean (SD)	2.93 (6.54)
CWMH, n (%)	46 (12.6)
Whole brain volume, L, mean (SD)	1.02 (0.11)
Global Atrophy, n (%)	225 (61.6)
Hippocampal volume, ml, mean (SD)	6.51 (1.12)
Hippocampal Atrophy, n (%)	252 (69.0)
Informant report cognitive symptoms, n (%)	361 (98.9)
MMSE, mean (SD)	27.26 (1.80)
3-year dementia, n (%)	142 (39)
APOE E4, n (%)	207 (56.7)
Family history, n (%)	188 (51.5)
Alcohol abuse, n (%)	15 (4.1)
Smoked, n (%)	136 (37.3)
BMI, mean (SD)	26.66 (4.49)

Obesity, n (%)	71 (19.5)

Note: SD: standard deviation; Q1: first quartile; Q3: third quartile; CSF: cerebrospinal fluid; FDG-PET: A fluorodeoxyglucose (FDG)-positron emission tomography (PET); MMSE: Mini-Mental State Exam; gait: signs of impaired gait including slowed gait; WMH: total white matter hyperintensities; CWMH: confluent white matter changes; *APOE*: apolipoprotein E; TBI: traumatic brain injury; CVD: cardiovascular disease; OSA: obstructive sleep apnea; VB12: vitamin B12 deficiency; BMI: body mass index; High education defined as education years 16 + years university degree or above. Obesity defined as BMI >30; smoked defined as current smoker or former smoker. *APOE* E4 is the presence of at least one E4 allele.

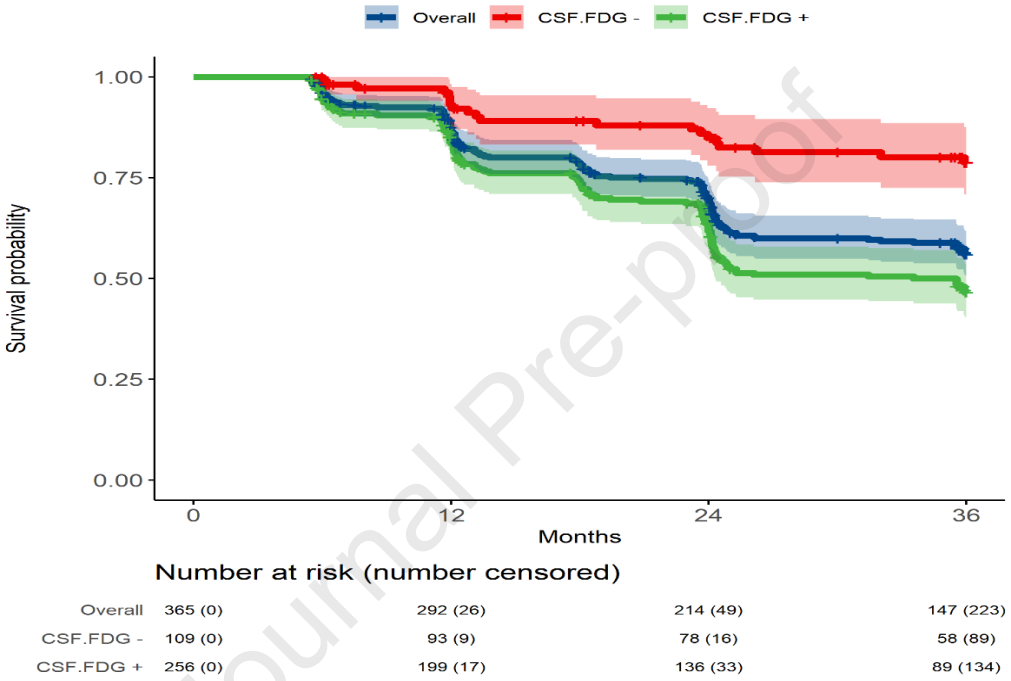

Figure 1 Kaplan-Meier curves of overall sample, and by the CSF and FDG-PET findings.

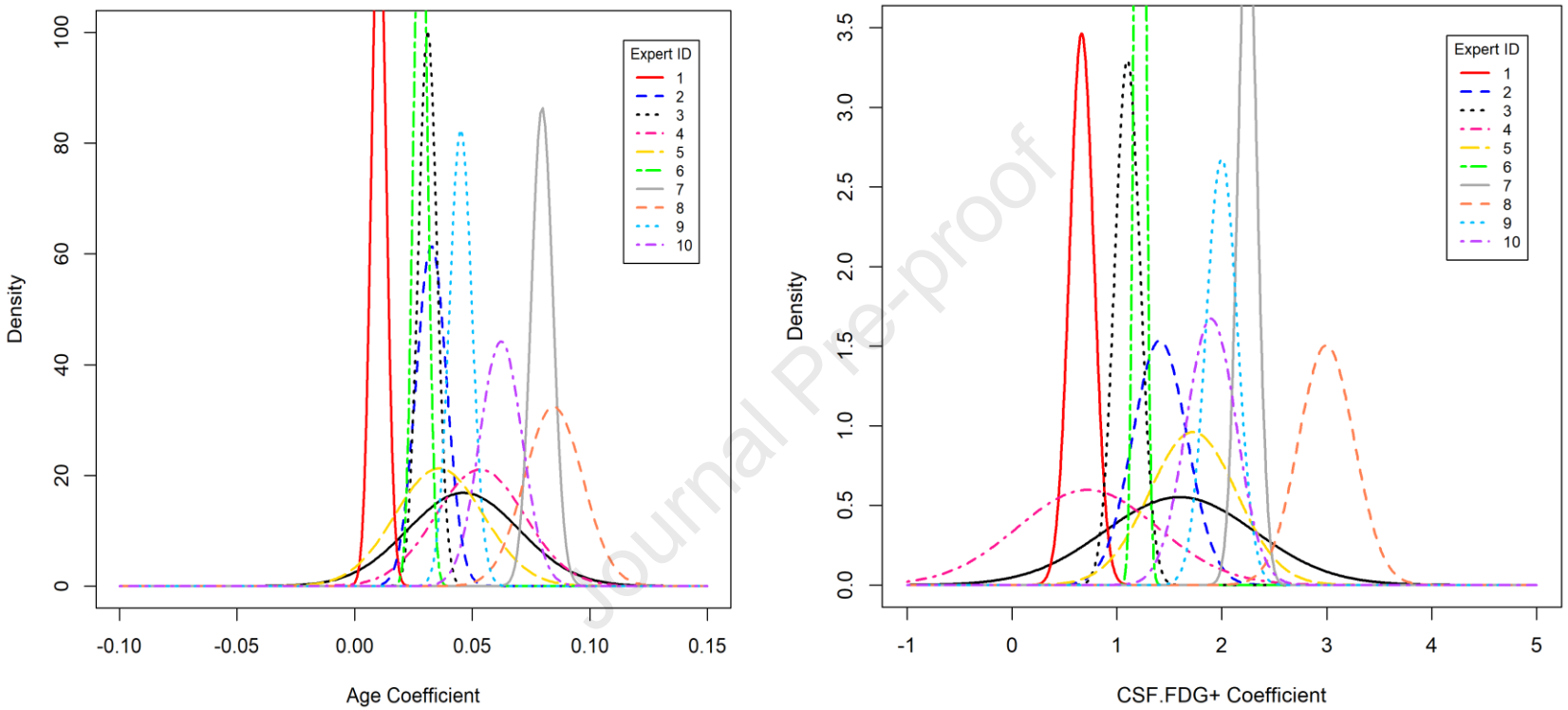

Figure 2 Prior distribution from experts for the two most important variables (the black solid line is the average distribution based on the mean of the best guesses and variance of the best guesses of the ten experts)

Table 2 Bayesian Cox and Cox regression model using ten-fold cross validation for ADNI data

	ADNI (n=365)		
	c-index [95% CI]	**ICI [95% CI]**	**IBS [95% CI]**
Informative (SEE)			
$\boldsymbol{\beta} \sim N(mean, \sum_{best})$	0.74 [0.70-0.79]	0.06 [0.05-0.08]	0.17 [0.14-0.19]
$\boldsymbol{\beta} \sim N(mean, \sum_{CI})$	0.71 [0.66-0.77]	0.06 [0.05-0.08]	0.16 [0.15-0.18]
$\boldsymbol{\beta} \sim N(median, \sum_{best})$	0.74 [0.69-0.79]	0.07 [0.05-0.08]	0.17 [0.15-0.19]
$\boldsymbol{\beta} \sim N(median, \sum_{CI})$	0.71 [0.65-0.75]	0.07 [0.05-0.10]	0.16 [0.14-0.19]
$\boldsymbol{\beta} \sim N(min, \sum_{best})$	0.75 [0.71-0.78]	0.05 [0.04-0.07]	0.16 [0.15-0.18]
$\boldsymbol{\beta} \sim N(max, \sum_{best})$	0.75 [0.69-0.80]	0.07 [0.06-0.08]	0.17 [0.15-0.19]
Noninformative			
$\boldsymbol{\beta} \sim N(0, \mathbf{10e6})$	0.75 [0.70-0.79]	0.06 [0.05-0.08]	0.16 [0.14-0.18]
Model without any priors (Frequentist approach)	0.72 [0.68-0.76]	0.06 [0.05-0.08]	0.16 [0.14-0.17]

NB: for each model, we specified piecewise constant baseline hazards (each follow-up year); c-index= Harrell's c-index; ICI = Integrate Calibration Index; IBS=brier score; CI=confidence interval, based on the 10-fold cross validation. $\boldsymbol{\beta} \sim N(mean, \sum_{best})$: we aggregated experts' data by using the mean of expert's best guess as the mean, using the variance of the experts' best guesses as the variance for each regression coefficient; $\boldsymbol{\beta} \sim N(mean, \sum_{CI})$*:* we aggregated experts' data by using the mean of expert's best guesses as the mean, and calculated variance from the min and max possible values from the experts' data for each regression coefficient; $\boldsymbol{\beta} \sim N(median, \sum_{best})$: we aggregated experts' data by using the median of expert's best guess as the mean, using the variance of the experts' best guesses as the variance for each regression coefficient; $\boldsymbol{\beta} \sim N(median, \sum_{CI})$*:* we aggregated experts' data by using the median of expert's best guesses as the mean, and calculated variance from the min and max possible values from the experts' data for each regression coefficient; $\boldsymbol{\beta} \sim N(min, \sum_{best})$: we aggregated experts' data by using the minimum of expert's best guess as the mean, using the variance of the experts' best guesses as the variance for each regression coefficient; $\boldsymbol{\beta} \sim N(max, \sum_{best})$: we aggregated experts' data by using the maximum of expert's best guess as the mean, using the variance of the experts' best guesses as the variance for each regression coefficient. Informative prior data came from structured expert elicitation.

Table 3 Bayesian Cox and Cox regression model using ten-fold cross validation for ADNI data-Sensitivity Analysis

Condition 1: excluded 3 experts with inconsistent estimates and excluded variables if any expert think it is less than somewhat important			
Informative (expert elicitation)	**c-index [95% CI]**	**ICI [95% CI]**	**IBS [95% CI]**
$\beta \sim N(mean, \sum_{best})$	0.76 [0.72-0.80]	0.08 [0.06-0.10]	0.16 [0.14-0.18]
Noninformative			
$\beta \sim N(0, 10e6)$	0.77 [0.73-0.80]	0.07 [0.06-0.08]	0.16 [0.14-0.18]
Model without any priors			
(Frequentist approach)	0.76 [0.74-0.79]	0.07 [0.06-0.10]	0.16 [0.14-0.18]
Condition 2: Included only Significant Predictors based on Posterior summaries			
Informative (expert elicitation)	**c-index [95% CI]**	**ICI [95% CI]**	**IBS [95% CI]**
$\beta \sim N(mean, \sum_{best})$	0.76 [0.70-0.83]	0.06 [0.05-0.07]	0.16 [0.14-0.19]
Noninformative			
$\beta \sim N(0, 10e6)$	0.75 [0.70-0.81]	0.07 [0.05-0.08]	0.16 [0.14-0.18]
Model without any priors			
(Frequentist approach)	0.76 [0.70-0.81]	0.06 [0.05-0.07]	0.16 [0.14-0.18]
Condition 3: Top 10 predictors from expert elicitation			
Informative (expert elicitation)	**c-index [95% CI]**	**ICI [95% CI]**	**IBS [95% CI]**
$\beta \sim N(mean, \sum_{best})$	0.74 [0.68-0.80]	0.06 [0.05-0.08]	0.16 [0.14-0.18]
Noninformative			
$\beta \sim N(0, 10e6)$	0.76 [0.70-0.81]	0.06 [0.05-0.07]	0.16 [0.14-0.18]
Model without any priors			
(Frequentist approach)	0.76 [0.70-0.81]	0.06 [0.05-0.07]	0.16 [0.14-0.18]

NB: for each model, we specified piecewise constant baseline hazards (each follow-up year). c-index= Harrell's c-index; ICI = Integrate Calibration Index; IBS=brier score; CI=confidence interval, based on bootstrapping. $\boldsymbol{\beta} \sim N(mean, \sum_{\boldsymbol{best}})$: we aggregated experts' data by using the mean of expert's best guess as the mean, using the variance of the experts' best guesses as the variance for each regression coefficient. In condition 1, inconsistence was defined as when one of the following is true: the highest estimate was lower than a best guess, a best guess was lower than the lowest estimate, or an estimated progression risk was lower than baseline risk for risk factors that the expert agreed on. In addition, variables were excluded if any expert considered the variable to be less than somewhat important (19 predictors remained). In condition 2, a Bayesian Cox regression was first fitted to the whole sample and then significant predictors were selected based on posterior summaries. In condition 3, the top ten important predictors were selected based on ten experts' ranking of the predictors using a seven-point Likert scale.

Table 2 shows Bayesian Cox and Cox regression model performances using ten-fold cross-validation for the ADNI data. When expert knowledge was incorporated, for example, using the mean of the best guesses among experts as the center of the normal distribution for each predictor, and using the variance of best guesses among experts as the variance of the normal distribution for each predictor, the c-index, ICI, and IBS values were 0.74 (95% CI 0.70-0.79), 0.06 (95% CI 0.05-0.08), and 0.17 (95% CI 0.14-0.19), respectively. The other aggregation methods for the informative priors of the regression coefficients demonstrated similar performances (Table 2 and Supplemental Table S5). When non-informative priors were used, the performance was similar to informative priors. In addition, for the Cox regression without any priors (Frequentist approach), the c-index, ICI, and IBS values were 0.72 (95% CI 0.68-0.76), 0.06 (95% CI 0.05-0.08), and 0.16 (95% CI 0.14-0.17), respectively. The effects of predictors between informative and noninformative models differed (>15% difference) in predictors such as apathy, MMSE, and *APOE* E4 (the presence of at least one E4 allele), but not in age, CSF or FDG-PET finding, and psychosis for example. Some comparisons between posterior and prior distributions were included in Supplemental Appendix E.

Table 3 details the results of the sensitivity analysis. The statistically significant predictors based on the informative model were age, CSF or FDG-PET finding, hippocampal atrophy, psychosis, MMSE score, depression, *APOE* E4, Vitamin B12 deficiency, and marital status (Supplemental Appendix E). In the non-informative model age and *APOE* E4 were not significant based on posterior summaries. We found that the model performance was slightly improved compared with Table 2 across the conditions, though informative priors still resulted in similar model performance as non-informative priors (Table 3).

DISCUSSION

This study proposed a Bayesian risk prediction model to combine expert knowledge with clinical data for dementia prognosis in people with MCI. Instead of solely depending on cohort data, this study explicitly incorporated expert knowledge from clinicians. We did not observe improvement of the predictive model performance when incorporating expert knowledge, compared with models without expert knowledge. There is mixed evidence in literature on whether expert knowledge is useful for improving clinical risk prediction models. For example, a previous study in cancer research incorporated expert knowledge in Bayesian variable selection and found that the predictive performance did not change with the introduction of informative priors, regardless of the parameters considered (such as the amount of weight assigned to the prior).[6] On the other hand, another study found that the experienced cardiologist's assessment had higher accuracy, compared with risk prediction models, in predicting the presence of obstructive coronary artery disease in patients with chest pain.[24] Another study combined data mining and case-based reasoning for facilitating more informed evidential decision making for pathology test ordering by general practitioners.[25] They concluded that the approach has advantages for decision support criteria (such as evidence base, situational relevance, and flexibility), without comparing model accuracy.[25] Therefore, it may be too soon to conclude that expert knowledge is not be useful for improving the prediction of dementia in persons with MCI. Some of the variables ranked as very important by the experts, such as psychosis, signs of parkinsonism, history of stroke or TIA, and informant recognition of cognitive symptoms, were not collected in sufficient detail in ADNI to include in the models. With more comprehensive data, it may be possible to that expert knowledge provides unique information.

In practice, when a patient is evaluated, a clinician must match patient-specific information with individual judgement based on prior knowledge with limited decision support.[25,26] Clinical knowledge is mostly based on substantive knowledge of the field, knowledge about biological mechanisms underlying brain aging, and practical experience. The limitation of this approach is that it may reinforce individual expert bias. Variations in physician practices when seeing patients with MCI are likely not captured in guidelines.[26] A strength of this study is that we addressed this limitation by including 11 experts from a broad range of backgrounds. On the other hand, a risk score often uses data-driven approaches, which can synthesize the effects of multiple predictors that could be too complex for a medical expert to process.[2] However, data-driven approaches are sensitive to sample size and depend on the quality of data being used for modelling. Data-driven approaches may have difficulty handling rare but very predictive variables (such as psychosis in MCI) that clinicians are readily aware of. Moreover, a clinical risk prediction model, like any other decision support tool, should support rather than replace individual clinician judgement, and this supportive role may be better achieved by including clinical expert knowledge using a standardized approach, such as we have applied in this study.

This study has some limitations. First, these models are developed based on data from the ADNI registry, which is a volunteer-based research cohort focused on AD that excluded patients with clinically diagnosed vascular cognitive impairment and other causes of dementia. The ADNI data is not a representative sample of people with MCI, since it is over-represented with males, white people, and people with high education. This population may be more homogeneous compared to the heterogeneous populations that clinicians see in their clinics. A challenging aspect of the expert elicitation is the choice of which measure to elicit. First, it

needs to be appropriate for experts. It is advised that elicited variables are preferred to be observable. For example, regression coefficients may be difficult to elicit directly from experts.[4] Second, the elicited variables must form the basis for estimating regression coefficients. Therefore, we asked experts to think of changing one variable at a time holding the others constant. However, this is not a common way for clinicians to think about risk, which may have contributed to variance across experts in their responses. In deriving a Bayesian Cox model with piecewise constant hazards, this study assumed linearity and additivity of the relationship between log hazards scale and the predictors, at each interval (per year over three years). Further studies need to explore more complex relationships, though it may be difficult to elicit such data from clinicians. Moreover, the conventional Cox regression is underpowered based on the ADNI sample, which might influence our study conclusions on whether clinician knowledge improves accuracy of the risk prediction model. This study used complete case analysis to deal with missing data, further study will explore the impact of different imputation methods on our study conclusions. Another limitation is that some expert knowledge of dementia risk may be used in practices that have not been subject to standardization and classification, such as the way a person speaks or follows directions. These aspects of the clinical encounter may inform the clinician's intuition but are not captured in cohort studies like ADNI. Future research may explore the use of vignettes or videotaped assessments that can be viewed by multiple experts to assist with elicitation of the most important risk factors.

In conclusion, we developed and validated a Bayesian risk prediction model that combines information from patient data and elicited expert knowledge to predict dementia progression in individuals with MCI. While we failed to show that incorporating clinical expert knowledge enhanced prediction in this study, we suggest that future studies with more

comprehensive risk factor data could explore the incorporation of clinical expertise into multivariable risk prediction models to improve model performance and facilitate implementation of more clinically relevant models.

Ethics: This study was approved by the University of Calgary Conjoint Health Research Ethics Board (REB19-046)

Acknowledgements/Conflicts/Funding Sources: This PhD study was supported by Alberta Innovates Graduate Student Scholarship, University of Calgary Cumming School of Medicine Graduate Scholarship, and Harley N. Hotchkiss Doctoral Scholarship in Neuroscience. Dr. Sajobi is supported by a Natural Sciences and Engineering Research Council Discovery grant. ADNI acknowledgement: Data collection and sharing for this project was funded by the ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the NIA, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing

funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCE

1. Chen Y, Qian X, Zhang Y, et al. Prediction Models for Conversion From Mild Cognitive Impairment to Alzheimer's Disease: A Systematic Review and Meta-Analysis. *Front Aging Neurosci*. 2022;14:840386. doi:10.3389/fnagi.2022.840386
2. Koopman RJ, Mainous AG. Evaluating multivariate risk scores for clinical decision making. *Fam Med*. 2008;40(6):412-416.
3. Bojke L, Soares M, Claxton K, et al. Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. *Health Technol Assess*. 2021;25(37):1-124. doi:10.3310/hta25370
4. O'Hagan A, ed. *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley & Sons; 2006.
5. Hemming V, Burgman MA, Hanea AM, McBride MF, Wintle BC. A practical guide to structured expert elicitation using the IDEA protocol. *Methods Ecol Evol*. 2018;9(1):169-180. doi:10.1111/2041-210X.12857
6. Boulet S, Ursino M, Thall P, et al. Integration of elicited expert information via a power prior in Bayesian variable selection: Application to colon cancer data. *Stat Methods Med Res*. 2020;29(2):541-567. doi:10.1177/0962280219841082
7. Ibrahim JG, Chen MH, Sinha D. Bayesian Survival Analysis. *Wiley StatsRef Stat Ref Online*. Published online 2014.
8. Hoff PD. *A First Course in Bayesian Statistical Methods*. Vol 580. Springer; 2009.
9. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI). *Neurology*. 2010;74(3):201-209. doi:10.1212/WNL.0b013e3181cb3e25
10. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-198. doi:10.1016/0022-3956(75)90026-6
11. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *Int Psychogeriatr*. 1997;9 Suppl 1:173-176; discussion 177-178. doi:10.1017/s1041610297004870

12. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984;34(7):939-944. doi:10.1212/wnl.34.7.939
13. Wang M, Smith EE, Forkert ND, et al. Integrating expert knowledge for dementia risk prediction in individuals with mild cognitive impairment (MCI): a study protocol. *BMJ Open*. 2021;11(11):e051185. doi:10.1136/bmjopen-2021-051185
14. Wang M, Sajobi T, Hogan D, et al. Expert Elicitation of Risk Factors for Progression to Dementia in Individuals with Mild Cognitive Impairment. *Alzheimers Dement Rev*. Under Revision 2022:1-16.
15. Adams-Hosking C, McBride MF, Baxter G, et al. Use of expert knowledge to elicit population trends for the koala (Phascolarctos cinereus). *Divers Distrib*. 2016;22(3):249-262. doi:10.1111/ddi.12400
16. Bedford T, Cooke R. *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge University Press; 2001. doi:10.1017/CBO9780511813597
17. Stone, M. The Opinion Pool. *Ann Math Statist*. Published online 1961:1339--1342. doi:10.1214/aoms/1177704873
18. SAS Institute Inc. The PHREG Procedure. In: *SAS/STAT® 15.1 User's Guide*. SAS Institute Inc.; 2018.
19. Gilks WR, Best NG, Tan KKC. Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *J R Stat Soc Ser C Appl Stat*. 1995;44(4):455-472. doi:10.2307/2986138
20. Austin PC, Harrell FE, van Klaveren D. Graphical calibration curves and the integrated calibration index (ICI) for survival models. *Stat Med*. 2020;39(21):2714-2742. doi:10.1002/sim.8570
21. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. 1999;18(17-18):2529-2545. doi:10.1002/(sici)1097-0258(19990915/30)18:17/18<2529::aid-sim274>3.0.co;2-5
22. SAS Institute Inc. SAS/STAT® 15.1 User's Guide (The PHREG Procedure).
23. RStudio Team (2021). RStudio: Integrated Development Environment for R.
24. Havistin R, Ivanov A, Patel P, et al. Analysis of clinical risk models vs. clinician's assessment for prediction of coronary artery disease among predominantly female population. *Coron Artery Dis*. 2022;33(3):182-188. doi:10.1097/MCA.0000000000001090
25. Zhuang ZY, Churilov L, Burstein F, Sikaris K. Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners. *Eur J Oper Res*. 2009;195(3):662-675. doi:10.1016/j.ejor.2007.11.003
26. Smellie WSA. Methodology for constructing guidance. *J Clin Pathol*. 2005;58(3):249-253. doi:10.1136/jcp.2004.018374

Author contributions: Tolulope Sajobi, Eric E Smith, and Meng Wang are responsible for the conceptualization of this study. Eric E Smith and Tolulope Sajobi supervised Meng Wang's doctoral work including clinician expert recruitment, elicitation process and data analysis. Meng Wang designed the questionnaire, analyzed data, summarized findings, and drafted the manuscript. Eric E Smith facilitated expert elicitation process. Thierry Chekouo provided Bayesian analysis expertise and reviewed all results with Meng Wang. All authors provided feedback on the interpretation of the analysis and results, critically reviewed, provided suggestions, and approved this manuscript.

What is new:

- This study mathematically modelled the decision-making process by which clinicians often arrive at dementia prognostic decisions.
- This study did not observe improvement of model predictive performance when incorporating expert knowledge.
- This study potentially enhances the use of expert knowledge.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: