Risk Prediction Model for Late Life Depression: Development and Validation on Three Large European Datasets

Risk Prediction Model for Late Life Depression: Development and Validation on Three Large European Datasets

Luca Cattelani, Martino Belvederi Murri, Federico Chesani, Lorenzo Chiari, Stefania Bandinelli, and Pierpaolo Palumbo

Abstract—Assessing the risk to develop a specific disease is the first step towards prevention, both at individual and population level. The development and validation of Risk Prediction Models (RPMs) is the norm within different fields of medicine but still underused in psychiatry, despite the global impact of mental disorders. In particular, there is a lack of RPMs to assess the risk of developing depression, the first worldwide cause of disability and harbinger of functional decline in old age. We present DRAT-up, the first prospective RPM to identify late life depression among community-dwelling subjects aged 60 to 75. The development of DRAT-up was based on appraisal of relevant literature, extraction of robust risk estimates and integration into model parameters. A unique feature is the ability to estimate risk even in the presence of missing values. To assess the properties of DRAT-up a validation study was conducted on three European cohorts, namely ELSA, InCHIANTI and TILDA, with 20206, 1359, and 3124 eligible samples, respectively. The model yielded accurate risk estimation in the three datasets from a small number of predictors. The Brier scores were 0.054, 0.133, and 0.041, while the values of Area Under the Curve (AUC) were 0.761, 0.736, and 0.768, respectively. Sensitivity analyses suggest robustness to missing values: setting any individual feature to unknown caused Brier scores to increase of 0.004, and AUCs to decrease of 0.045 in the worst cases. DRAT-up can be readily used for clinical purposes and to aid policy making in the field of mental health.

Index Terms—late life depression, prevention, risk assessment, risk factors, risk prediction model.

I. INTRODUCTION

Risk prediction models (RPMs) are widely used in the medical field to estimate the probability of individuals or populations to develop a specific condition or outcome (usually a disease), based on the knowledge on multiple risk factors [1]. RPMs can be useful both for clinical purposes and to guide policy making. Since they are able to provide information on specific risks of individuals, RPMs can indeed aid physicians in decision making and clinical management. For instance, intensive clinical monitoring or specific clinical interventions may be reserved only to those meeting criteria for high clinical risk. Moreover, RPMs can be implemented by policy-makers to plan large-scale interventions aimed at specific subpopulations characterized by a higher risk to develop the disease. Although to date few RPMs have been evaluated in their economic impact, they may be also used to improve resource allocation, thus improving the appropriateness and cost-effectiveness of interventions [2].

So far, RPMs have been used to predict various conditions including cardiovascular diseases [3], [4], cancer [5], surgery outcomes [6] or hospital readmissions [7]. However, they are still underutilized in the mental health field [8] despite the growing relevance of mental health disorders worldwide. In particular, depression is one of the leading causes of disability and a major contributor to the global burden of disease[1]. Furthermore, the impact of late life depression is expected to grow significantly over the next years considering the aging of the population [9].

Late life depression is particularly disabling: it causes personal and familial suffering, heightens suicide risk, worsens the outcomes of associated physical conditions and increases healthcare costs [10], [11]. Treatments for old age depression include antidepressant drugs [12], [13] or promising psychosocial interventions, such as psychotherapy, problem solving therapy or physical exercise [14]–[17]. Despite this, clinical outcomes remain suboptimal worldwide due to lack of resources and frequent under-diagnosis [9], [18]. In recent years, the importance of prevention and early identification of depression has been increasingly recognized [8]. In fact, various preventive strategies are becoming available, including rehabilitation and psychotherapy as well as medications and organizational strategies [19]–[21]. Moreover, the prognosis of depression improves if treatment is delivered earlier [22]. Thus, RPMs would be particularly useful to focus preventive strategies to individuals at high-risk for late life depression.

At present only two RPMs for late life depression have been developed. The first was based on a path analysis of 299 older subjects and included gender, income, education, activities of daily living and somatic symptoms as predictors [23]. This study was however limited by a cross-sectional design and a low predictive power (the model explained only 28% of the variance of the outcome). The second RPM, by Okamoto

L. Cattelani and F. Chesani are with the Department of of Computer Science and Engineering, University of Bologna, Bologna, Italy.

M. Belvederi Murri is with the Section of Psychiatry, Department of Neuroscience, Ophthalmology, Genetics and Infant-Maternal Science, University of Genoa, Genoa, Italy.

L. Chiari and P. Palumbo are with the Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", University of Bologna, Bologna, Italy.

L. Chiari is also with the Health Sciences & Technologies Interdepartmental Center for Industrial Research, University of Bologna, Bologna, Italy.

S. Bandinelli is with the Geriatric Unit, Local Health Unit Tuscany Centre, Florence, Italy.

Manuscript received , 2018; revised , 2018.

[1] http://www.who.int/mediacentre/factsheets/fs369/en/

and Harasawa, was based on a stepwise linear discrimination analysis of a sample of 754 subjects aged 65 and older. Predictors were hearing problems, poor appetite, financial leeway, emotional support and perceived usefulness [24]. This RPM correctly identified 78% of subjects as depressed, but was again based on cross sectional data, thus making it impossible to establish its real prospective value. The reliability of existing RPMs for late life depression is also limited by other important features:

1) choice of predictors based on a single study featuring a small population, often from a single geographical site;
2) exclusion of subjects with missing data, which is common in social and medical science [25];
3) lack of external validation, which severely limits the generalizability of findings.

Given these premises, the aim of the present study, was to develop a methodologically robust RPM for late life depression. For this purpose, we implemented the following approach:

1) a-priori appraisal of scientific literature to identify risk factors and their distribution in the target population (evidence-based approach) [26];
2) development of the RPM using a recently developed methodology, that is able to work in the presence of missing data. A similar method was implemented in a recent RPM to predict the incidence of falls among older subjects [27], [28];
3) validation of the model on multiple, large, longitudinal datasets from different geographical areas.

II. METHODS

A. Development

The Depression Risk Assessment Tool DRAT-up has been developed following the same approach employed for FRAT-up, a Web-based fall risk assessment tool for community-dwelling older people [27]. DRAT-up is based on probabilistic logic. Unlike artificial neural networks and other artificial intelligence approaches, this method does not require direct training on data but rather on estimates derived from the existing literature. The systematic review and meta-analysis on risk factors for depression in community older people, by Cole and Dendukuri [26], was used to identify the risk factors acknowledged as statistically significant.

We call $\delta = (\delta_1, \delta_2, ..., \delta_5)$ the vector of the five dichotomous risk factors (Table I), $\delta_i \in \{0, 1, NA\}$. Namely, $\delta_i = 0$, 1, or NA if, respectively, the i-th risk factor is not present, is present, or if this information is not available. We call Y the random variable representing future depression, $Y \in \{0, 1\}$.

The risk score $f(\delta)$ given by DRAT-up is the probability of future depression after 2 years, conditional on the available information about risk factors exposure.

As better explained in Appendix, subsection VI-A, it is calculated as

$$f(\delta) = 1 - (1 - C_0) \prod_{i \in K} (1 - C_i)^{\delta_i} \prod_{j \in U} (1 - p_j C_j) \quad (1)$$

TABLE I
RISK FACTORS ODDS RATIOS (OR), OR 95% CREDIBLE INTERVALS, PREVALENCE, AND PROBABILITY CONTRIBUTIONS

	OR (credible interval) [26]	Prevalence	C_i
Gender (female)	1.4 (1.2, 1.8)	0.530 [30]	0.024
Disability	2.5 (1.6, 4.8)	0.178 [31]	0.084
Sleep disorder	2.6 (1.9, 3.7)	0.285 [32]	0.089
Bereavement	3.3 (1.7, 4.9)	0.041 [33]	0.124
Current depression	2.3 (1.1, 7.1)	0.123 [29]	0.074

where C_0 is the probability of future depression when no risk factor is present; C_i is the probability of future depression due to risk factor i; K (U) is the set of risk factors on which we have (respectively, we do not have) information about their exposure; and p_j is the prevalence of risk factor j.

The idea underlying this modelling formula is that future depression may result from any of its risk factors, or may occur idiopathically (i.e., due to an unexplained mechanism, see the term C_0). Risk factors for which we have information on their exposure (risk factors i, $i \in K$) give a contribution to the risk of depression in the product which is either 1 or $1-C_i$, depending on whether δ_i is 0 or 1. This term can equivalently be rewritten as $1 - \delta_i C_i$. Risk factors for which we do not have information (risk factors j, $j \in U$) give a contribution to the risk in the product which is $1 - p_j C_j$, that is obtained replacing the risk factor indicator δ_j with the prevalence p_j.

Following the approach described in [27], the probability contributions C_i were derived from the odds ratios reported in [26], $C_0 = 0.061$ was derived from [29], and the risk factors prevalence rates from other sources [29]–[33], as detailed in Table I.

B. Validation

DRAT-up was successively validated on three datasets of European cohorts of older people, namely the English Longitudinal Study of Ageing (ELSA) [34], [35], the Invecchiare nel Chianti (InCHIANTI) study [36], and The Irish Longitudinal Study on Ageing (TILDA) [37]. ELSA is designed to be used for the investigation of a multidisciplinary set of topics relevant to understanding the ageing process. Both objective and subjective data are collected including health trajectories, disability, cognition, physical and mental health, and household and family structure. InCHIANTI is a population-based study of older persons living in the Chianti area in Tuscany (Italy). It was driven by the desire to improve the understanding in how multiple physiologic subsystems influence the ability to walk in the long run. It includes data on physiological and neurological impairments, disability, diseases and injuries, and marital status. TILDA is a study on ageing that collects information on all aspects of health, economic and social circumstances. It presents a breadth of physical, mental health, cognitive, social and economic measures.

We included only samples relative to subjects without missing information on depression outcome and aged between 60 and 75 years. This choice was made to aid the early identification of depression in a crucial age [38], [39]. Therefore, samples from the last wave and samples where the outcome

is a missing value for other reasons (e.g. subsequent death or refusal/impossibility to answer) were not included.

Risk factor exposure and outcome (i.e. future depression) were operationalized in each dataset as described in Table II. Their distributions are reported in Table III. In particular, the 'Current depression' risk factor is obtained by applying a cut-off at 16 to the CES-D score in InCHIANTI and TILDA, or at 4 to the short version CES-D in ELSA. The 'Depression' outcome is obtained by applying the same cut-off to the following wave.

We chose the AUC and the Brier score as summary performance metrics. We decomposed the Brier score in its three constituents: reliability (REL), resolution (RES), and uncertainty (UNC): $Brier = REL - RES + UNC$. Having conditioned on the risk score, REL is the bias of risk scores with respect to the conditional means of the observations, RES is the variance of conditional means, and UNC is the variance of the observations.

$$REL = \frac{1}{n}\sum_{k=1}^{m} n_k(f_k - \overline{Y_k})^2$$

$$RES = \frac{1}{n}\sum_{k=1}^{m} n_k(\overline{Y_k} - \overline{Y})^2$$

$$UNC = \overline{Y}(1 - \overline{Y})$$

Where n_k is the number of observations receiving risk score f_k; m is the number of possible different risk scores f_k; $n = \sum_{k=1}^{m} n_k$ is the total sample size; $\overline{Y_k}$ is the mean Y over all observations receiving risk score f_k; and $\overline{Y}$ is the mean Y over all observations [40]. We compared the Brier score of DRAT-up with the one of a dummy tool, namely a tool that assigns to each sample the outcome prevalence assumed from the literature. This is indeed the behaviour of DRAT-up when information on all risk factors is missing.

We compared DRAT-up with the model developed by Okamoto and Harasawa [24] on the ELSA and TILDA datasets.

In ELSA, subjective usefulness was derived from the question 'Have you recently felt you were playing a useful part in things?'. The four possible responses were grouped in two categories: yes ('better than usual' and 'same as usual') and no ('less than usual' and 'much less than usual'). Feeling of economic leeway was derived from the item 'Shortage of money stops me from doing things I want to do', dichotomizing the answers in two categories: low ('often' and 'sometimes') and high ('not often' and 'never'). Social support was derived from summing the scores obtained on 12 questions about social support from partner, children, other immediate family members, and friends. This total score was dichotomized based on the median. Hearing problems were assessed from the question 'Is your hearing (using a hearing aid as usual)', dichotomizing the answers in two categories: bad ('fair' and 'poor') and good ('excellent', 'very good', and 'good'). Appetite could not be assessed. Subjective usefulness could not be derived in waves 2, 4, and 5.

In TILDA, subjective usefulness was derived from the 11th item of the CASP-19 scale for quality of life in early old age [41] ('I feel that my life has meaning'). The four possible responses were grouped in two categories: yes ('often', 'sometimes') and no ('rarely', 'never'). Feeling of economic leeway was derived from the 9th item of CASP-19 ('Shortage of money stops me from doing the things that I want to do'), dichotomizing the answers in two categories: low ('often' and 'sometimes') and high ('not often' and 'never'). Social support was derived from dichotomizing the Berkman-Syme Social Network Index [42] on its mean score. Appetite was derived from dichotomizing the second item of the CES-D scale ('I did not feel like eating; my appetite was poor'): poor (3-7 days during the last week) and good (2 days or less during the last week). Hearing problems were derived from the respondents self-rated hearing: bad ('fair', 'poor') and good ('excellent', 'very good', 'good').

Missing values were replaced by prevalence rates reported in [24]. We judged that the InCHIANTI dataset could not be harmonized to calculate the Okamoto score.

C. Unknown values

DRAT-up manages missing input information (i.e. unknown risk factor exposure) the same way as FRAT-up [27]. More in particular, the risk score is calculated averaging over the distribution of missing values (formula 1).

In order to test its robustness under these circumstances, we calculated how its performance varies for different percentages of missing values, from the original percentage in the dataset up to 100%, with subjects taken to be missing randomly sampled.

More in particular, we call $a_{i,o}$ the fraction of missing values on the i-th risk factor in the dataset at hand. We define the mixture distribution

$$P_i(\gamma) = (1 - \gamma)P_o + \gamma P_{i,NA}$$

where $\gamma \in [0,1]$, P_o is the data distribution in the dataset and $P_{i,NA}$ is the distribution of data after imposing $\delta_i = NA$ on all the samples. For γ that ranges from 0 to 1, the fraction of missing values on the i-th risk factor in the $P_i(\gamma)$ mixture ranges from $a_{i,o}$ to 1.

For each risk factor i, we have calculated the Brier score and the AUC as functions of the fraction of missing values on i, using the results for mixture distributions reported in Appendix, subsection VI-B. As a benchmark, we have further calculated them numerically, averaging over 1000 random mixtures for each γ.

III. RESULTS

DRAT-up was validated on 24689 samples, relative to 11704 individuals, from the 3 datasets, as detailed in Table IV.

A. Performance

The AUC, the Brier score, and its three components attained by DRAT-up on the 3 datasets are reported in Table V. 95% confidence intervals (CI) are computed with 1000 iterations of the bootstrap method. The AUC ranged from 0.736 in InCHIANTI to 0.768 in TILDA. Receiver operating characteristic

TABLE II
OPERATIONAL DEFINITIONS OF RISK FACTORS AND OUTCOME IN EACH DATASET. ADL, IADL, AND CES-D STAND FOR ACTIVITIES OF DAILY LIVING, INSTRUMENTAL ADL, AND CENTER FOR EPIDEMIOLOGIC STUDIES DEPRESSION SCALE.

	ELSA	InCHIANTI	TILDA
Gender	Gender item	Gender item	Gender item
Disability	ADL or IADL >0	ADL or IADL >0	ADL or IADL >0
Sleep disorder	CES-D item 11 >0	"How would you rate your overall sleep quality during the last month?" score 3-4 (poor or very poor)	Rating "Most of the time" at any of the 3 items on: daytime sleepiness, difficulty falling asleep, early waking
Bereavement	Not used	Widow for 0-12 months	Not used
Current depression	Short version CES-D ≥ 4	CES-D ≥ 16	CES-D ≥ 16
Depression	Short version CES-D ≥ 4 after 2 years	CES-D ≥ 16 after 3 years	CES-D ≥ 16 after 2 years

TABLE III
PREVALENCE RATES OF RISK FACTORS AND OUTCOME IN EACH DATASET. SHARES OF RESPECTIVELY TRUE, FALSE, AND UNKNOWN. 95% CONFIDENCE INTERVALS IN BRACKETS.

	ELSA	InCHIANTI	TILDA
Female	0.540 (0.533, 0.546) / 0.460 (0.454, 0.467) / 0.000	0.528 (0.503, 0.553) / 0.472 (0.447, 0.497) / 0.000	0.530 (0.512, 0.547) / 0.470 (0.453, 0.488) / 0.000
Disability	0.197 (0.192,0.203) / 0.802 (0.796, 0.808) / 0.001 (0.001, 0.001)	0.102 (0.0854, 0.119) / 0.898 (0.881, 0.915) / 0.000	0.107 (0.097, 0.118) / 0.893 (0.882, 0.903) / 0.000
Sleep disorder	0.379 (0.373, 0.385) / 0.611 (0.605, 0.617) / 0.010 (0.008, 0.011)	0.241 (0.219, 0.265) / 0.752 (0.728, 0.776) / 0.007 (0.003, 0.011)	0.322 (0.305, 0.338) / 0.678 (0.661, 0.695) / 0.000 (0.000, 0.001)
Bereavement	0.000 / 0.000 / 1.000	0.012 (0.006, 0.018) / 0.988 (0.982, 0.994) / 0.000	0.000 / 0.000 / 1.000
Current depression	0.053 (0.050, 0.056) / 0.937 (0.933, 0.940) / 0.010 (0.009, 0.012)	0.149 (0.132, 0.169) / 0.846 (0.827, 0.864) / 0.005 (0.001, 0.009)	0.042 (0.035, 0.049) / 0.956 (0.949, 0.964) / 0.002 (0.001, 0.004)
Depression	0.055 (0.052, 0.058) / 0.945 (0.942, 0.948) / 0.000	0.171 (0.151, 0.191) / 0.829 (0.809, 0.849) / 0.000	0.039 (0.032, 0.046) / 0.961 (0.954, 0.968) / 0.000

TABLE IV
NUMBER OF INCLUDED WAVES, PERSONS AND SAMPLES BY DATASET

	ELSA	InCHIANTI	TILDA
Sample years by wave	2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2012	1998-2000, 2002-2003, 2004-2006	2010-2012
Last wave (outcome only)	2012-2013	2007-2009	2012-2013
Persons	7920	660	3124
Samples	20206	1359	3124

TABLE V
PERFORMANCE. 95% CONFIDENCE INTERVALS IN BRACKETS.

	ELSA	InCHIANTI	TILDA
AUC	0.761 (0.746, 0.775)	0.736 (0.703, 0.769)	0.768 (0.717, 0.815)
Brier score	0.054 (0.052, 0.056)	0.133 (0.118, 0.148)	0.041 (0.036, 0.045)
Reliability	0.008 (0.008, 0.009)	0.015 (0.011, 0.023)	0.010 (0.009, 0.012)
Resolution	0.006 (0.005, 0.007)	0.023 (0.018, 0.032)	0.006 (0.004, 0.010)
Uncertainty	0.052 (0.049, 0.054)	0.142 (0.127, 0.154)	0.037 (0.031, 0.044)

Fig. 1. ROC curve of DRAT-up on the three validation datasets.

B. Robustness to missing/unknown values

Results from the analysis on the robustness to missing values are shown in Figures 2 and 3. The AUC decreased at most by 0.045. The Brier score increases at most of 0.004, thus reaching the skill of the dummy tool in the worst case.

IV. DISCUSSION

The present study reports on DRAT-up, an evidence-based RPM to identify depression among community-dwelling older adults. By employing a novel approach to risk calculation [27] and using only five predictors, DRAT-up yielded a fairly accurate estimate of incident depression while overcoming the most

(ROC) curves are displayed in Fig. 1. The Brier score is dominated by the uncertainty term, which only depends on the prevalence of depression, and mainly explains the differences of the Brier score in the three cohorts.

The AUCs for the Okamoto score on the ELSA and TILDA datasets are respectively 0.672 (95% CI 0.657-0.690) and 0.683 (95% CI 0.628-0.735).
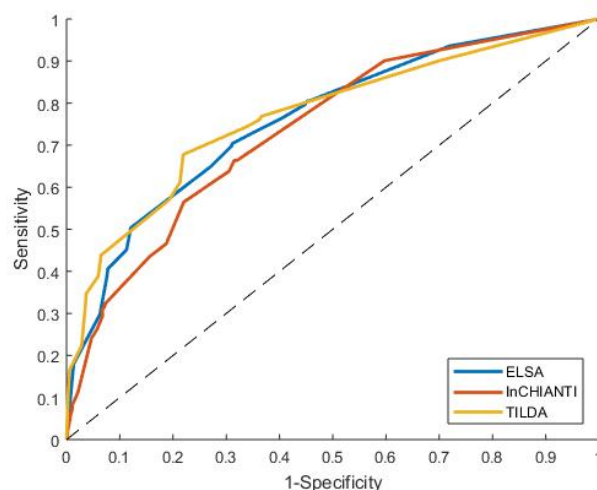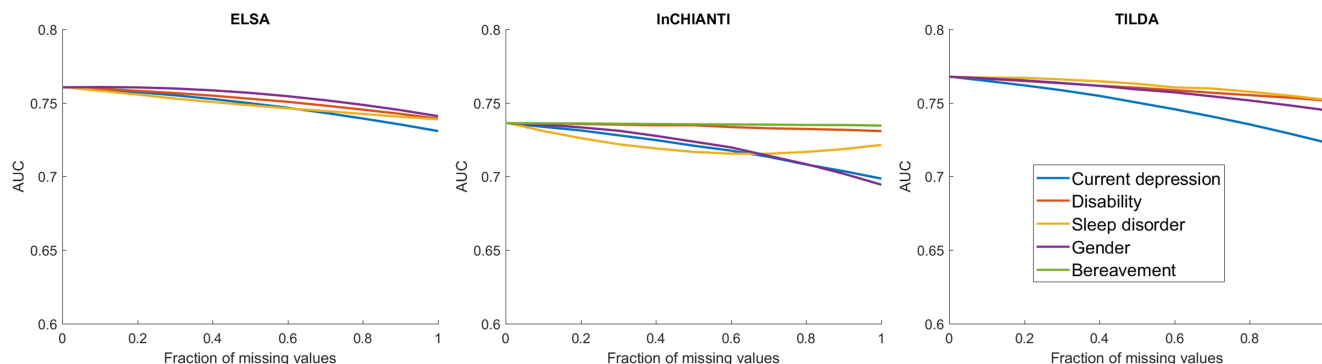
Fig. 2. Robustness of the AUC of DRAT-up to missing values. AUC is displayed as a function of the fraction of missing values on the single risk factors for the three validation datasets.
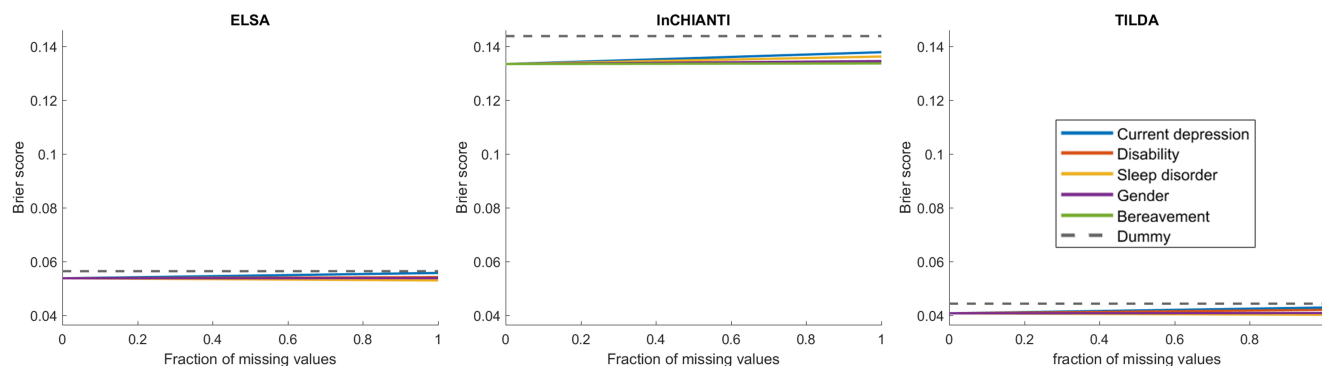

Fig. 3. Robustness of the Brier score of DRAT-up to missing values. The Brier score is displayed as a function of the fraction of missing values on the single risk factors for the three validation datasets. The dashed line (labeled as 'Dummy') indicates the performance of DRAT-up when no information on risk factor exposure is available on any sample. In this case DRAT-up assigns to each sample the depression prevalence rate assumed from the literature.

important methodological shortcomings of previous models [8]. Upon direct comparison with a previously-proposed RPM [24], DRAT-up showed a higher discriminative ability.

DRAT-up was developed using robust evidence on risk factors that are specific for depression among older individuals, obtained from prospective studies over a medium term period [26]. The review allowed to identify a-priori robust, reliable predictors while excluding a number of factors that are likely to be spuriously associated with depression. The model was validated on three large epidemiological datasets from different parts of Europe to establish its reliability in different populations, without the need for a localized training procedure.

As presently elaborated, DRAT-up was characterized by a fair performance in all three datasets, with similar ROC curves. Taking together, these features constitute an important step forward from previous RPMs within the mental health field, underlined in a recent literature review [8]. Another important strength of DRAT-up is the ability to estimate risk even in the presence of missing values, which is critical issue in risk prediction. Very few RPMs, to our knowledge, take into account the presence of missing data [25], [43]. In our study, missing data was replaced with evidence from epidemiological studies, i.e. the prevalence of each factor in the target population. This method proved to be effective: in most cases, it was possible to remove data for each predictor at a time without

strongly affecting the model performance in terms of AUC values and Brier scores. However, the importance of each factor may be different, and it may change between different datasets. For instance, despite the absence of information on bereavement in ELSA and TILDA, the model performance was not largely affected. Similarly, the removal of bereavement from the InCHIANTI dataset had a very small impact on the performance of the model. Bereavement is associated with the greatest risk for incident depression (OR=3.3), but its impact may be limited likely because of an overall low prevalence at population-level. On the contrary, the removal of the risk factor 'Current depression' caused the Brier score to approach the performance of the dummy tool in ELSA and TILDA.

To wholly interpret the results of this study, some further issues need to be discussed.

First, the performance of the model presented nontrivial differences between the three datasets, and this may depend on different reasons. The cohorts are expected to be intrinsically different, and this improves the evaluation of the model generalizability. For instance, we encountered an almost double prevalence of disability in the ELSA population than among InCHIANTI or TILDA; depression baseline prevalence was much higher in InCHIANTI than other datasets despite similar definitions. These differences may partly reflect true population heterogeneity. An additional degree of heterogeneity may depend on methodological factors, such as the operational

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2884079, IEEE Journal of Biomedical and Health Informatics

JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. , NO. , 2018   6

definitions of the predictors or the outcome. For instance, the incidence of newly-diagnosed depression was greater in InCHIANTI than in the other datasets. This may partially depend on the longer timeframe between predictor and outcome assessment (after 3 years for InCHIANTI as opposed to 2 years for the other datasets). Future studies should validate RPMs on cross-country data employing homogeneous methodologies.

Second, the predictors employed in the model were identified from a methodologically rigorous, albeit dated, meta-analysis [26]. Despite being the most methodologically rigorous review available, an update of the evidence could further improve the model and is underway. The model will be further extended to include other relevant predictors, such as social and geographical factors [44]–[47], marital status [48] and presence of chronic diseases [49]–[51].

Third, in all datasets used for validation of the model, depression was defined based on the CES-D questionnaire. The CES-D is generally considered an accurate and reliable indicator of the diagnosis of Major Depression in younger and older individuals, but there has been debate on which cut-off score represents the best choice. Recently, a meta-analysis suggested that a cut-off of 20 instead of 16 could slightly improve specificity, although at the price of a small decrease of sensitivity [52]. Several reports based on the InChianti study have used 20 as a cut-off score [53], while those from TILDA more often used 16 [54]. In the present study we have chosen to adopt the more inclusive cut-off score of 16, in consideration of the clinical importance of mild depression in older populations [55]. Nonetheless, future studies should validate these findings using the cut-off of 20 or other reliable indicators of depression.

Lastly, DRAT-up is based on data about community-dwelling individuals, and its validity among institutionalized subjects yet remains to be determined.

V. CONCLUSION

In conclusion, DRAT-up is a new risk prediction model to identify late life depression among community-dwelling individuals, two to three years in advance, using only five simple predictors. The development of DRAT-up overcomes several important methodological shortcomings that hampered previous studies, namely cross-sectional design, model development from a single dataset, inability to cope with missing data and lack of external validation [8]. Whereas, DRAT-up yielded a fair accuracy on large datasets from different European countries, and was able to cope with missing data.

Further improvements of DRAT-up could derive from the inclusion of other predictors derived from robust scientific evidence, calibration based on population-specific estimates (e.g. at catchment area level) and periodic updates [56]. Moreover, further studies are required to evaluate the clinical consequences of this model [57]. Although depression is a significant predictor of incident disability in most studies [58], [59], accurate prediction does not necessarily translate to increased clinical attention or better outcomes. Novel treatments for late life depression and their implementation into the community are urgently needed [60]. Clinicians and policy

makers may use DRAT-up to aid the early identification of subjects at-risk for depression [22]. Moreover, DRAT-up may be used to increase the specificity of preventive interventions against depression in order to improve the clinical outcomes while reducing disability and societal costs [61], [62]. Clearly, further studies are necessary to establish the optimal correspondence between varying degrees of depression risk and appropriate interventions. Meanwhile, this study confirms the feasibility of RPMs in the mental health field.

Clinicians and policy makers may take advantage from this model in service planning or clinical activity.

REFERENCES

[1] J. A. A. G. Damen, L. Hooft, E. Schuit, T. P. A. Debray, G. S. Collins, I. Tzoulaki, C. M. Lassale, G. C. M. Siontis, V. Chiocchia, C. Roberts, M. M. Schlüssel, S. Gerry, J. A. Black, P. Heus, Y. T. van der Schouw, L. M. Peelen, and K. G. M. Moons, "Prediction models for cardiovascular disease risk in the general population: systematic review," *BMJ*, vol. 353, 2016. [Online]. Available: http://www.bmj.com/content/353/bmj.i2416

[2] A. van Giessen, J. Peters, B. Wilcher, C. Hyde, C. Moons, A. de Wit, and E. Koffijberg, "Systematic Review of Health Economic Impact Evaluations of Risk Prediction Models: Stop Developing, Start Evaluating," *Value in Health*, vol. 20, no. 4, pp. 718–726, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1098301516300328

[3] F. Fowkes, G. Murray, I. Butcher, C. Heald, R. Lee, L. Chambless, A. Folsom, A. Hirsch, M. Dramaix, G. DeBacker, J. Wautrecht, M. Kornitzer, A. Newman, M. Cushman, K. Sutton-Tyrrell, F. Fowkes, A. Lee, J. Price, R. D'Agostino, J. Murabito, P. Norman, K. Jamrozik, J. Curb, K. Masaki, B. Rodriguez, J. Dekker, L. Bouter, R. Heine, G. Nijpels, C. Stehouwer, L. Ferrucci, M. McDermott, H. Stoffers, J. Hooi, J. Knottnerus, M. Ogren, B. Hedblad, J. Witteman, M. Breteler, M. Hunink, A. Hofman, M. Criqui, R. Langer, A. Fronek, W. Hiatt, R. Hamman, H. Resnick, J. Guralnik, and M. McDermott, "Ankle brachial index combined with framingham risk score to predict cardiovascular events and mortality: A meta-analysis," *JAMA*, vol. 300, no. 2, pp. 197–208, jul 2008. [Online]. Available: http://dx.doi.org/10.1001/jama.300.2.197

[4] G. C. M. Siontis, I. Tzoulaki, K. C. Siontis, and J. P. A. Ioannidis, "Comparisons of established risk prediction models for cardiovascular disease: systematic review," *BMJ*, vol. 344, no. may24 1, pp. e3318–e3318, 2012. [Online]. Available: http://www.bmj.com/cgi/doi/10.1136/bmj.e3318

[5] J. A. Usher-Smith, F. M. Walter, J. D. Emery, A. K. Win, and S. J. Griffin, "Risk Prediction Models for Colorectal Cancer: A Systematic Review," *Cancer Prev. Res.*, vol. 9, no. 1, pp. 13–26, jan 2016. [Online]. Available: http://cancerpreventionresearch.aacrjournals.org/content/9/1/13

[6] R. Zhang, O. Borisenko, I. Telegina, J. Hargreaves, A. R. Ahmed, R. Sanchez Santos, C. Pring, P. Funch-Jensen, B. Dillemans, and J. L. Hedenbro, "Systematic review of risk prediction models for diabetes after bariatric surgery," *Br. J. Surg.*, vol. 103, no. 11, pp. 1420–1427, oct 2016. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/27557164 http://doi.wiley.com/10.1002/bjs.10255

[7] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, and S. Kripalani, "Risk prediction models for hospital readmission: A systematic review," *JAMA*, vol. 306, no. 15, pp. 1688–1698, oct 2011. [Online]. Available: http://dx.doi.org/10.1001/jama.2011.1515

[8] F. Bernardini, L. Attademo, S. D. Cleary, C. Luther, R. S. Shim, R. Quartesan, and M. T. Compton, "Risk Prediction Models in Psychiatry: Toward a new frontier for the prevention of mental illnesses," *J. Clin. Psychiatry*, vol. 78, no. 5, pp. 572–583, may 2017. [Online]. Available: https://doi.org/10.4088/JCP.15r10003

[9] W. V. McCall and K. W. Kintziger, "Late life depression: a global problem with few resources." *The Psychiatric clinics of North America*, vol. 36, no. 4, pp. 475–81, dec 2013. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0193953X13000828

[10] G. S. Alexopoulos, "Depression in the elderly," *Lancet*, vol. 365, no. 9475, pp. 1961–1970, 2005.

[11] M. Luppa, C. Sikorski, T. Motzek, A. Konnopka, H.-H. König, and S. G. Riedel-Heller, "Health Service Utilization and Costs of Depressive Symptoms in Late Life - A Systematic Review," *Curr. Pharm. Des.*, vol. 18, no. 36, pp. 5936–5957, oct 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22681171

[12] R. M. Kok and C. F. Reynolds, "Management of Depression in Older Adults: A Review." *JAMA*, vol. 317, no. 20, pp. 2114–2122, may 2017. [Online]. Available: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2017.5706

[13] E. Tedeschini, Y. Levkovitz, N. Iovieno, V. E. Ameral, J. C. Nelson, and G. I. Papakostas, "Efficacy of antidepressants for late-life depression: a meta-analysis and meta-regression of placebo-controlled randomized trials." *The Journal of clinical psychiatry*, vol. 72, no. 12, pp. 1660–8, dec 2011. [Online]. Available: http://article.psychiatrist.com/?ContentType=START&ID=10007683

[14] M. Belvederi Murri, M. Amore, M. Menchetti, G. Toni, F. Neviani, M. Cerri, M. B. L. Rocchi, D. Zocchi, L. Bagnoli, E. Tam, A. Buffa, S. Ferrara, M. Neri, G. S. Alexopoulos, S. Zanetidou, and Safety and Efficacy of Exercise for Depression in Seniors (SEEDS) Study Group, "Physical exercise for late-life major depression." *The British journal of psychiatry : the journal of mental science*, vol. 207, no. 3, pp. 235–42, sep 2015. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/26206864

[15] F. Holvast, B. Massoudi, R. C. Oude Voshaar, and P. F. M. Verhaak, "Non-pharmacological treatment for depressed older patients in primary care: A systematic review and meta-analysis," *Plos One*, vol. 12, no. 9, p. e0184666, 2017. [Online]. Available: http://dx.plos.org/10.1371/journal.pone.0184666

[16] U. Jonsson, G. Bertilsson, P. Allard, H. Gyllensvärd, A. Söderlund, A. Tham, and G. Andersson, "Psychological Treatment of Depression in People Aged 65 Years and Over: A Systematic Review of Efficacy, Safety, and Cost-Effectiveness," *PLOS ONE*, vol. 11, no. 8, p. e0160859, aug 2016. [Online]. Available: http://dx.plos.org/10.1371/journal.pone.0160859

[17] C. Okolie, M. Dennis, E. Simon Thomas, and A. John, "A systematic review of interventions to prevent suicidal behaviors and reduce suicidal ideation in older people." *International psychogeriatrics*, vol. 29, no. 11, pp. 1801–1824, 2017. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/28766474

[18] M. Belvederi Murri, M. Amore, M. Respino, and G. S. Alexopoulos, "The symptom network structure of depressive symptoms in late-life: Results from a European population study," *Molecular Psychiatry*, aug 2018. [Online]. Available: http://www.nature.com/articles/s41380-018-0232-0

[19] C. Mihalopoulos and T. Vos, "Costeffectiveness of preventive interventions for depressive disorders: an overview," *Expert Review of Pharmacoeconomics & Outcomes Research*, vol. 13, no. 2, pp. 237–242, apr 2013. [Online]. Available: http://www.tandfonline.com/doi/full/10.1586/erp.13.5

[20] R. C. Baldwin, "Preventing late-life depression: A clinical update," *International Psychogeriatrics*, vol. 22, no. 8, pp. 1216–1224, 2010.

[21] E. M. Whyte and B. Rovner, "Depression in late-life: Shifting the paradigm from treatment to prevention," *International Journal of Geriatric Psychiatry*, vol. 21, no. 8, pp. 746–751, 2006.

[22] L. Ghio, S. Gotelli, A. Cervetti, M. Respino, W. Natta, M. Marcenaro, G. Serafini, M. Vaggi, M. Amore, and M. Belvederi Murri, "Duration of untreated depression influences clinical outcomes and disability," *Journal of Affective Disorders*, vol. 175, pp. 224–228, 2015.

[23] K.-R. Shin, D. Jung, I. Jo, and Y. Kang, "Depression Among Community-Dwelling Older Adults in Korea: A Prediction Model of Depression," *Arch. Psychiatr. Nurs.*, vol. 23, no. 1, pp. 50–57, feb 2009. [Online]. Available: http://dx.doi.org/10.1016/j.apnu.2008.03.001

[24] K. Okamoto and Y. Harasawa, "Prediction of symptomatic depression by discriminant analysis in Japanese community-dwelling elderly," *Archives of Gerontology and Geriatrics*, vol. 52, no. 2, pp. 177–180, 2011.

[25] J. D. Dziura, L. A. Post, Q. Zhao, Z. Fu, and P. Peduzzi, "Strategies for Dealing with Missing Data in Clinical Trials: From Design to Analysis," *Yale J. Biol. Med.*, vol. 86, no. 3, pp. 343–358, sep 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24058309 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3767219/

[26] M. G. Cole and N. Dendukuri, "Risk Factors for Depression Among Elderly Community Subjects: A Systematic Review and Meta-Analysis," *Am. J. Psychiatry*, vol. 160, no. 6, pp. 1147–1156, jun 2003.

[27] L. Cattelani, P. Palumbo, L. Palmerini, S. Bandinelli, C. Becker, F. Chesani, and L. Chiari, "FRAT-up, a web-based fall-risk assessment tool for elderly people living in the community," *Journal of Medical Internet Research*, vol. 17, no. 2, jan 2015. [Online]. Available: http://doi.org/10.2196/jmir.4064

[28] P. Palumbo, J. Klenk, L. Cattelani, S. Bandinelli, L. Ferrucci, K. Rapp, L. Chiari, and D. Rothenbacher, "Predictive Performance of a Fall Risk Assessment Tool (FRAT-up) for Community-Dwelling Older People in 4 European Cohorts," *J. Am. Med. Dir. Assoc.*, vol. 17, no. 12, pp. 1106–1113, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1525861016302936

[29] J. R. M. Copeland, A. T. F. Beekman, A. W. Braam, M. E. Dewey, P. Delespaul, R. Fuhrer, C. Hooijer, B. A. Lawlor, S.-L. Kivela, A. Lobo, H. Magnusson, A. H. Mann, I. Meller, M. J. Prince, F. Reischies, M. Roelands, I. Skoog, C. Turrina, M. W. DeVries, and K. C. M. Wilson, "Depression among older people in Europe: the EURODEP studies." *World psychiatry : official journal of the World Psychiatric Association (WPA)*, vol. 3, no. 1, pp. 45–9, feb 2004.

[Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16633454 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1414664/

[30] "Database - Eurostat." [Online]. Available: http://ec.europa.eu/eurostat/web/population-demography-migration-projections/population-data/database#

[31] A. Börsch-Supan, M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck, and S. Zuber, "Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)," *International Journal of Epidemiology*, vol. 42, no. 4, pp. 992–1001, aug 2013. [Online]. Available: https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dyt088

[32] R. E. Roberts, S. J. Shema, G. A. Kaplan, and W. J. Strawbridge, "Sleep Complaints and Depression in an Aging Cohort: A Prospective Perspective," *American Journal of Psychiatry*, vol. 157, no. 1, pp. 81–88, jan 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10618017 http://psychiatryonline.org/doi/abs/10.1176/ajp.157.1.81

[33] C. L. Turvey, C. Carney, S. Arndt, R. B. Wallace, and R. Herzog, "Conjugal Loss and Syndromal Depression in a Sample of Elders Aged 70 Years or Older," *American Journal of Psychiatry*, vol. 156, no. 10, pp. 1596–1601, oct 1999. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10518172 http://psychiatryonline.org/doi/abs/10.1176/ajp.156.10.1596

[34] A. Steptoe, E. Breeze, J. Banks, and J. Nazroo, "Cohort Profile: The English Longitudinal Study of Ageing," *Int. J. Epidemiol.*, vol. 42, no. 6, pp. 1640–1648, dec 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23143611

[35] M. Marmot, "English Longitudinal Study of Ageing: Waves 0-5, 1998-2011 [computer file]," 2013.

[36] L. Ferrucci, S. Bandinelli, E. Benvenuti, A. Di Iorio, C. Macchi, T. B. Harris, and J. M. Guralnik, "Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study." *Journal of the American Geriatrics Society*, vol. 48, no. 12, pp. 1618–25, dec 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11129752

[37] P. M. Kearney, H. Cronin, C. O'Regan, Y. Kamiya, G. M. Savva, B. Whelan, and R. Kenny, "Cohort Profile: The Irish Longitudinal Study on Ageing," *Int. J. Epidemiol.*, vol. 40, no. 4, pp. 877–884, aug 2011. [Online]. Available: https://doi.org/10.1093/ije/dyr116

[38] S. P. Passon, S. Ostling, and I. Skoog, "The incidence of first-onset depression in a population followed from the age of 70 to 85." *Psychological medicine*, vol. 31, no. 7, pp. 1159–68, oct 2001. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11681542

[39] Y. Yang, "Is Old Age Depressing? Growth Trajectories and Cohort Variations in Late-Life Depression," *Journal of Health and Social Behavior*, vol. 48, no. 1, pp. 16–32, mar 2007. [Online]. Available: http://journals.sagepub.com/doi/10.1177/002214650704800102

[40] A. H. Murphy, "A New Vector Partition of the Probability Score," *Journal of Applied Meteorology*, vol. 12, no. 4, pp. 595–600, jun 1973. [Online]. Available: https://doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2

[41] M. Hyde, R. D. Wiggins, P. Higgs, and D. B. Blane, "A measure of quality of life in early old age: The theory, development and properties of a needs satisfaction model (CASP-19)," *Aging & Mental Health*, vol. 7, no. 3, pp. 186–194, may 2003. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/12775399 http://www.tandfonline.com/doi/abs/10.1080/1360786031000101157

[42] L. F. Berkman and S. L. Syme, "Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents." *American journal of epidemiology*, vol. 109, no. 2, pp. 186–204, feb 1979. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/425958

[43] G. Stiglic, P. Kocbek, N. Fijacko, A. Sheikh, and M. Pajnkihar, "Challenges associated with missing data in electronic health records: A case study of a risk prediction model for diabetes using data from Slovenian primary care," *Health Informatics Journal*, p. 146045821773328, oct 2017. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1460458217733288

[44] E. A. Ambugo, "Cross-country variation in the sociodemographic factors associated with major depressive episode in Norway, the United Kingdom, Ghana, and Kenya," *Social Science and Medicine*, vol. 113, pp. 154–160, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.socscimed.2014.05.022

[45] J. R. Copeland, A. T. Beekman, M. E. Dewey, C. Hooijer, A. Jordan, B. A. Lawlor, A. Lobo, H. Magnusson, A. H. Mann, I. Meller, M. J. Prince, F. Reischies, C. Turrina, M. W. Devries, and K. C. Wilson, "Depression in Europe. Geographical distribution among older people,"

British Journal of Psychiatry, vol. 174, no. APR., pp. 312–321, apr 1999. [Online]. Available: https://doi.org/10.1192/bjp.174.4.312

[46] P. Henríquez-Sánchez, J. Doreste-Alonso, M. A. Martínez-González, M. Bes-Rastrollo, A. Gea, and A. Sánchez-Villegas, "Geographical and climatic factors and depression risk in the SUN project," *European Journal of Public Health*, vol. 24, no. 4, pp. 626–631, 2014.

[47] G. Juhasz, N. Eszlari, D. Pap, and X. Gonda, "Cultural differences in the development and characteristics of depression." *Neuropsychopharmacologia Hungarica*, vol. 14, no. 4, pp. 259–65, dec 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23269213

[48] X.-Y. Yan, S.-M. Huang, C.-Q. Huang, W.-H. Wu, and Y. Qin, "Marital Status and Risk for Late Life Depression: A Meta-Analysis of the Published Literature," *Journal of International Medical Research*, vol. 39, no. 4, pp. 1142–1154, aug 2011. [Online]. Available: http://imr.sagepub.com/cgi/content/long/39/4/1142 http://journals.sagepub.com/doi/10.1177/147323001103900402

[49] C.-Q. Huang and D. B. Rong, "Chronic diseases and risk for depression in old age: a meta-analysis of published literature." *Ageing research reviews*, vol. 9, no. 2, pp. 131–41, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.arr.2009.05.005

[50] M. Sakagami, T. Kitahara, T. Okayasu, A. Yamashita, A. Hasukawa, I. Ota, and T. Yamanaka, "Negative prognostic factors for psychological conditions in patients with audiovestibular diseases," *Auris Nasus Larynx*, vol. 43, no. 6, pp. 632–636, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.anl.2016.02.006

[51] V. Valkanova and K. P. Ebmeier, "Vascular risk factors and depression in later life: A systematic review and meta-analysis," *Biological Psychiatry*, vol. 73, no. 5, pp. 406–413, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.biopsych.2012.10.028

[52] G. Vilagut, C. G. Forero, G. Barbaglia, and J. Alonso, "Screening for depression in the general population with the center for epidemiologic studies depression (ces-d): A systematic review with meta-analysis," *PLoS One*, vol. 11, no. 5, pp. 1–17, 2016.

[53] E. Vermeulen, I. A. Brouwer, K. Stronks, S. Bandinelli, L. Ferrucci, M. Visser, and M. Nicolaou, "Inflammatory dietary patterns and depressive symptoms in Italian older adults," *Brain, Behavior, and Immunity*, vol. 67, pp. 290–298, 2018. [Online]. Available: https://doi.org/10.1016/j.bbi.2017.09.005

[54] R. Briggs, S. P. Kennelly, and R. A. Kenny, "Does baseline depression increase the risk of unexplained and accidental falls in a cohort of community-dwelling older people? Data from The Irish Longitudinal Study on Ageing (TILDA)," *International Journal of Geriatric Psychiatry*, vol. 33, no. 2, pp. e205–e211, feb 2018. [Online]. Available: http://doi.wiley.com/10.1002/gps.4770

[55] T. W. Meeks, I. V. Vahia, H. Lavretsky, G. Kulkarni, and D. V. Jeste, "A tune in "a minor" can "b major": A review of epidemiology, illness course, and public health implications of subthreshold depression in older adults," *Journal of Affective Disorders*, vol. 129, no. 1-3, pp. 126–142, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.jad.2010.09.015

[56] K. G. M. Moons, A. P. Kengne, D. E. Grobbee, P. Royston, Y. Vergouwe, D. G. Altman, and M. Woodward, "Risk prediction models: II. External validation, model updating, and impact assessment," *Heart*, vol. 98, no. 9, pp. 691–698, 2012.

[57] A. J. Vickers and E. B. Elkin, "Decision Curve Analysis: A Novel Method for Evaluating Prediction Models," *Medical Decision Making*, vol. 26, no. 6, pp. 565–574, nov 2006. [Online]. Available: https://doi.org/10.1177/0272989X06295361

[58] D. Connolly, J. Garvey, and G. McKee, "Factors associated with ADL/IADL disability in community dwelling older adults in the Irish longitudinal study on ageing (TILDA)," *Disability and Rehabilitation*, vol. 39, no. 8, pp. 809–816, apr 2017. [Online]. Available: https://www.tandfonline.com/doi/full/10.3109/09638288.2016.1161848

[59] E. D'Orsi, A. J. Xavier, A. Steptoe, C. de Oliveira, L. R. Ramos, M. Orrell, P. Demakakos, and M. G. Marmot, "Socioeconomic and lifestyle factors related to instrumental activity of daily living dynamics: results from the English Longitudinal Study of Ageing." *Journal of the American Geriatrics Society*, vol. 62, no. 9, pp. 1630–9, sep 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/25243677

[60] G. S. Alexopoulos and M. L. Bruce, "A model for intervention research in late-life depression," *International Journal of Geriatric Psychiatry*, vol. 24, no. 12, pp. 1325–1334, dec 2009. [Online]. Available: http://doi.wiley.com/10.1002/gps.2287

[61] M. Luppa, H.-H. König, D. Heider, T. Motzek, G. Schomerus, and S. G. Riedel-Heller, "Direct costs associated with depressive symptoms in late life: a 4.5-year prospective study," *International Psy-*

chogeriatrics, vol. 25, no. 02, pp. 292–302, 2013. [Online]. Available: http://www.journals.cambridge.org/abstract_S1041610212001688

[62] O. I. Okereke, C. F. Reynolds, D. Mischoulon, G. Chang, N. R. Cook, T. Copeland, G. Friedenberg, J. E. Buring, and J. A. E. Manson, "The VITamin D and OmegA-3 TriaL-Depression Endpoint Prevention (VITAL-DEP): Rationale and design of a large-scale ancillary study evaluating vitamin D and marine omega-3 fatty acid supplements for prevention of late-life depression," *Contemporary Clinical Trials*, vol. 68, no. February, pp. 133–145, 2018. [Online]. Available: https://doi.org/10.1016/j.cct.2018.02.017

[63] G. Parisi, "Balance research funds across Europe," *Nature*, vol. 530, p. 33, feb 2016. [Online]. Available: http://dx.doi.org/10.1038/530033d

VI. Appendix

A. Justification of Equation 1

Here we show that Equation 1 gives the output risk score $f(\delta)$ as the probability of future depression conditional on the available information about risk factors exposure.

As our model assumption, when exposure on all risk factor is known (i.e. $K = \{1, 2, ..., 5\}$), the output risk score $f(\delta)$ is given by the 'noisy-or' formula

$$f(\delta) = P(Y = 1|\delta_i, i \in K) = P(Y = 1|\delta) =$$
$$1 - (1 - C_0) \prod_{i=1}^{5} (1 - C_i)^{\delta_i}.$$

We now partition the set of risk factor indices $\{1, 2, ..., 5\}$ in K, U, and a singleton $\{s\}$, and proceed by induction on the set U. In particular, we calculate the output risk score $f(\delta)$ for the case of δ_s unknown, starting from the case of δ_s known.

When δ_s is known, we assume that the output risk score $f(\delta)$ is given by

$$f(\delta) = P(Y = 1|\delta_i, i \in K \cup \{s\}) =$$
$$1 - (1 - C_0) \prod_{i \in K \cup \{s\}} (1 - C_i)^{\delta_i} \prod_{j \in U} (1 - p_j C_j).$$

From the conditional probability law we have

$$P(Y|\delta_i, i \in K) =$$
$$P(Y|\delta_s = 1, \delta_i, i \in K)P(\delta_s = 1|\delta_i, i \in K)+$$
$$P(Y|\delta_s = 0, \delta_i, i \in K)P(\delta_s = 0|\delta_i, i \in K).$$

Considering δ_s independent from any δ_i with $i \neq s$, and $P(\delta_s = 1) = p_s$ as it is assumed by our model, it is easy to derive that the output risk score $f(\delta)$ when δ_s is unknown is given by

$$f(\delta) = P(Y = 1|\delta_i, i \in K) =$$
$$1 - (1 - C_0) \prod_{i \in K} (1 - C_i)^{\delta_i} \prod_{j \in U \cup \{s\}} (1 - p_j C_j)$$

which concludes our proof by induction. We note that under our modeling assumptions, the output risk score $f(\delta)$ is given by the product of a number of terms that does not depend on the number of unknown risk factors. Instead, more generally the output risk score $f(\delta)$, defined as a conditional probability on the risk factors with available information, would be calculated as the summation over a number of terms that grows exponentially with the number of unknown risk factors.

B. Brier score and AUC of mixture distributions

Given a mixture distribution $P_m = \gamma P_1 + (1-\gamma)P_2$ with $\gamma \in [0, 1]$, here we show that the Brier score is an affine function of the mixture parameter γ, while the AUC is quadratic. The proofs trivially follow from the definitions.

The Brier score is defined as the expected value of the squared difference between the outcome and the risk score:

$$Brier(f, P_{\delta, Y}) = E[(Y - f(\delta))^2]$$

where the expectation is taken over the distribution $P_{\delta, Y}$ of (δ, Y). Thus, it follows that the Brier score is linear in its second argument in the sense that,

$$Brier(f, P_m) = \gamma Brier(f, P_1) + (1 - \gamma) Brier(f, P_2).$$

which shows that the Brier score is affine in γ.

We proceed analogously for the AUC. AUC is defined as

$$AUC(f, P) = E_{P(\delta^h|Y^h=0)}[E_{P(\delta^d|Y^d=1)}[I(\delta^h, \delta^d)]]$$

where

$$I(\delta^h, \delta^d) = \begin{cases} 1 & \text{if } f(\delta^d) > f(\delta^h) \\ 0 & \text{if } f(\delta^d) < f(\delta^h) \\ 1/2 & \text{otherwise} \end{cases}$$

Here d and h are used as superscripts.

We further define $\widetilde{AUC}$, an extended version of AUC:

$$\widetilde{AUC}(f, P_1, P_2) = E_{P_1(\delta^h|Y^h=0)}[E_{P_2(\delta^d|Y^d=1)}[I(\delta^h, \delta^d)]]$$

where P_1 and P_2 are distinct probability distributions of (δ, Y). Clearly, $\widetilde{AUC}(f, P, P) = AUC(f, P)$.

As above, we consider the mixture distribution P_m. In our case, we forcefully introduce missing values on δ without altering Y, thus $P_1(Y) = P_2(Y)$ and the mixture holds also for the conditional distributions: $P_m(\delta|Y) = \gamma P_1(\delta|Y) + (1-\gamma)P_2(\delta|Y)$. Therefore

$$AUC(f, P_m) = \widetilde{AUC}(f, P_m, P_m) =$$
$$\gamma^2 AUC(f, P_1) + (1 - \gamma)^2 AUC(f, P_2)$$
$$\gamma(1-\gamma)[\widetilde{AUC}(f, P_1, P_2) + \widetilde{AUC}(f, P_2, P_1)]$$

which shows that AUC is quadratic in γ.