Risk prediction models of natural menopause onset: a systematic review

Hamidreza Raeisi-Dehkordi [1,2†], Stefanie Kummer [1†], Peter Francis Raguindin [1,2,3], Gordana Dejanovic [1], Petek Eylul Taneri [4,5], Isabel Cardona [1,6], Lum Kastrati [1,2], Beatrice Minder [7], Trudy Voortman [8,9], Pedro Marques-Vidal [10], Klodian Dhana [11], Marija Glisic [1,3], Taulant Muka [1*]

[1] *Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland.*

[2] *Graduate School for Health Sciences, University of Bern, Bern, Switzerland.*

[3] *Swiss Paraplegic Research, 6207, Nottwil, Switzerland.*

[4] *School of Nursing and Midwifery, National University of Ireland, Galway, Galway, Ireland.*

[5] *HRB-Trials Methodology Research Network, National University Of Ireland, Galway, Ireland.*

[6] *Department of Obstetrics and Gynaecology, McGill University Health Center, McGill University, Montreal, Canada.*

[7] *Public Health & Primary Care Library, University Library of Bern, University of Bern, Switzerland.*

[8] *Department of Epidemiology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands.*

[9] *Division of Human Nutrition and Health, Wageningen University & Research, Wageningen, the Netherlands.*

[10] *Department of Medicine, Internal Medicine, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland.*

[11] *Section on Nutrition and Nutritional Epidemiology, Department of Internal Medicine, Rush University, Chicago, Illinois, USA.*

---

† Authors contributed equally.

© The Author(s) 2022. Published by Oxford University Press on behalf of the Endocrine Society. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

source: https://doi.org/10.48350/171692 | downloaded: 12.10.2024

*Corresponding author: Taulant Muka, associate professor, Institute of Social and Preventive Medicine (ISPM), University of Bern, Mittelstrasse 43, Bern 3012, Switzerland.

ORCID: 0000-0003-3235-3073

Email: taulant.muka@ispm.unibe.ch

Disclosures: The authors have nothing to disclose.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Abstract

Context Predicting the onset of menopause is important for family planning and to ensure prompt intervention in women at risk of developing menopause-related diseases.

Objective To summarize risk prediction models of natural menopause onset and their performance.

Data Sources and Study Selection Five bibliographic databases were searched up to March 2022. We included prospective studies on perimenopausal women or women in menopausal transition, that reported either the univariable or multivariable model for risk prediction of natural menopause onset.

Data Extraction Two authors independently extracted data according to the CHARMS (critical appraisal and data extraction for systematic reviews of prediction modelling studies) checklist. Risk of bias was assessed using PROBAST (prediction model risk of bias assessment tool).

Data Synthesis Of 8'132 references identified, we included 14 articles based on 8 unique studies comprising 9'588 women (mainly Caucasian) and 3'289 natural menopause events. All the included studies used onset of natural menopause (ONM) as outcome, while four studies predicted early ONM as well. Overall, there were 180 risk prediction models investigated, with age, anti-Müllerian hormone (AMH) and follicle-stimulating hormone (FSH) being the most investigated predictors. Estimated C-statistic for the prediction models ranged from 0.62 to 0.95. Although all studies were rated at high risk of bias mainly due to the methodological concerns related to the statistical analysis, their applicability was satisfactory.

Conclusion Predictive performance and generalizability of current prediction models on ONM is limited given that these models were generated from studies at high risk of bias and from specific populations/ethnicities. Although in certain settings such models may be useful, efforts to improve their performance are needed as use becomes more widespread.

Keywords Risk prediction model, Onset of menopause, Perimenopause, Premenopausal women

Funding source None

PROSPERO registration code CRD42021248771.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

INTRODUCTION

Menopause as the complete cessation of the menstrual periods, marks the end of a woman's natural reproductive lifespan. This event occurs between the age range of 40-60 year, with a median age of approximately 49 to 51 years (1). The relatively wide age range for the onset of menopause would seem to indicate that females have a highly variable number of oocytes and/or lose them at a highly variable rate (2). With the aging of the population, especially in the last decades, it is estimated that 1.2 billion women will be menopausal (experiencing menopause) or postmenopausal by 2030 (3). Age at natural menopause is a crucial period for women of reproductive age since both an early or late menopause may have some health implications (4). Early menopause is associated with lower bone mass density, osteoporosis, depression and increased risk of type 2 diabetes, cardiovascular diseases, and all-cause mortality (5-8). On the other hand, late menopause has been associated with increased risk of sex-related cancers including breast, ovarian and endometrial cancers (9). Thus, predicting onset of natural menopause (ONM) may not only help clinicians regarding contraception (10), assisted reproduction (11) and management of menorrhagia (12), but also may alert women in reproductive ages of health-related complications due to either early or late menopause. With ongoing lifestyle changes and trends, many women postpone childbearing to the end of their reproductive lifespan. Therefore, an accurate estimation of time of menopause could improve the preventive management of unexpected early infertility (13).

While menopause is inevitable, research has indicated that ONM may vary depending on a wide range of factors including genetic factors, lifestyle and environmental exposures, sociodemographic factors and hormonal mechanisms (14,15). Thus, considering these known risk factors in univariable or multivariable models, may provide insight into the prediction of natural menopause precisely. The most studied predictor biomarker is the circulating anti-Mullerian hormone (AMH), which as a single biomarker or combined with age and other variables shows a C-statistics ranging from 0.70 to 0.90 (16,17). Estradiol and Follicle-Stimulating Hormone (FSH) have also been explored as predictor biomarkers of ONM, showing a C-statistic similar to AMH

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

(18,19). However, most articles have not performed validation (17,18,20), neither calibration (17,18,21,22). Calibration is an important step in prediction models because even a good-performing model based on its C-statistics, can over- or underestimate the outcomes risk. In addition, the majority of the current prediction models have not provided clear time prediction period to predict ONM. Uncertainty on generalizability of prediction models to different populations and ethnicities exists as well. Thus, there is a need to summarize the risk prediction models for ONM and understanding best predictors of natural menopause as well as current gaps in the literature. In the present study, we aimed to perform a comprehensive systematic review to identify prediction models to forecast the timing of ONM and age at menopause of premenopausal women.

METHODS

This systematic review was conducted according to a guideline on systematic reviews and meta-analysis (23), and reported according to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA)-guidelines (24). The study protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) database on May 15, 2021 (CRD42021248771).

Search strategy

A systematic literature search was conducted by a library information scientist (BM) in Embase (Ovid), Web of Science, Medline (Ovid), Cochrane CENTRAL and Google Scholar up to February 8, 2021 and was updated on March 25, 2022 using the following Medical Subject Headings (MeSH) and non-MeSH keywords: (i) risk prediction, predictive, prediction model, diagnostic, diagnostic model, predictive ability; (ii) menopause, premature menopause, perimenopause, post menopause, age of menopause, early menopause, late menopause, menopause onset; (iii) accuracy, sensitivity, specificity, positive predictive value PPV, negative predictive value NPV, likelihood ratio, area under the ROC curve (AUC), Youden's index, diagnostic odds ratio (DOR), C-statistics. Furthermore, we made a forward citation tracking and also hand-

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

searched the reference list of eligible studies. The selection of studies was not restricted based on date and language. The detailed search strategy is presented in **Supplementary Material 1** (25)**.**

Selection criteria

Published original studies fulfilling the following criteria were included: (i) prospective cohort studies, case-cohort, and nested case-control studies (i.e., studies with prospective data collection); (ii) conducted on premenopausal or perimenopausal women or women in menopausal transition; (iii) had followed women over time for the occurrence of menopause and reported ONM; (iv) reported either the univariable model or multivariable model for risk prediction of ONM; (v) had presented at least one of the following outcomes: accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), likelihood ratio, area under the curve (AUC), Youden's index, diagnostic odds ratio (DOR), and C-statistics.

Studies were excluded if they (i) were performed in animals; (ii) were cross-sectional or case-control studies without prospective data collection; (iii) were focused on menopause due to surgery, radiotherapy, or chemotherapy; (iv) provided data about model performance beyond our inclusion criteria like hazard ratios. Titles and abstracts of publications were manually screened by independent reviewers (HR-D, SK, PFR, GD, PET). Subsequently, the full-text screening was performed. Disagreements in judgment were resolved by discussion with a third reviewer (MG or TM).

Data extraction

For each included article, we extracted the following data: study and population characteristics (year of publication, country, study design, funding source, number of participants, follow up time, exclusion criteria, age, menopause status), variables considered for the prediction (hormones, imaging-like transvaginal ultrasound, menstrual cycle irregularities), number of natural menopause events, definition of menopause and early natural menopause, prediction time and performance of the model (C-statistics, sensitivity, specificity, AUC) using a standardized data extraction form based on the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Modelling Studies (CHARMS)-checklist (26). In the case of multiple publications, the most up-to-date or comprehensive paper was included unless the article investigated different predictors and study outcomes in different time periods within the same study (e.g., 2 years or 1 year before the onset of the last menstruation (22,27)) . Two reviewers (HR-D and SK) independently extracted all of the required data. Possible discrepancies were resolved by consulting a third reviewer (MG or TM).

Quality assessment

Risk of bias (ROB) and applicability to the intended population and setting was examined by using "A tool to assess risk of bias and applicability of prediction model studies PROBAST"(28). PROBAST contains main domains including (participants, predictors, outcome, and analysis) and assesses the ROB for the four domains and applicability, except for the analysis domain. Each domain is characterized by specific questions (20 questions overall). Questions may be answered as yes, probably yes, probably no, or no information, depending on the provided information of each included study. A domain is at high ROB when at least one of its items is judged as not appropriate (no or probably no). The overall risk of bias is considered as high ROB when one or more domains are considered to be at high risk.

RESULTS

Study selection

A flow chart of the study selection process is shown in **Figure 1**. A total of 8'132 citations were obtained from the initial database search. After removal of duplicates, 4'415 unique references remained for screening of the titles and abstracts, of which 95 articles underwent full-text reviewing. Fourteen articles based on eight unique studies (16-22,27,29-34) describing 180 prediction models of ONM satisfied the inclusion criteria and were selected for data extraction and synthesis **(Supplementary Table 2)** (25).

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Study characteristics

Population. Of the 14 articles included, six (16-21) used data from the Netherlands, four from USA (21,23,29,32), three (22,30,34) from Iran and one (32) article combined data from the Netherlands and Iran. Some cohorts were used multiple times for model development. The Doetinchem Cohort Study was used in De Kat et al. (16) and Dolleman et al. 2015 (21). The SWAN cohort (The Study of Women's Health Across the Nation) was used in Finkelstein et al. (22) and Greendale et al. (27) . The article by Depmann et al. 2016 is a follow-up study of Broer et al. (the Scheffer, van Rooij, de Vet (SRV) cohort study). The article of Dolleman et al. 2014 (20) consists of two cohorts, one of which is the SRV cohort. The TLGS cohort (Teheran Lipid and Glucose Study) was used in Gohari et al. (29), Tehrani et al. 2020 (31) and Tehrani et al. 2021(34). The study by Tehrani et al. 2014 (32) was performed using data from two cohorts, TLGS and the SRV cohort. Thus, 14 articles (16-22,27,29-34) based on 8 unique study populations (16,17,19,20,22,29,30,33) were included for final analysis.

All articles included premenopausal women with ages ranging from 18 to 63 years at baseline. Ten articles included Caucasians (16-21,29,31,32,34), while two articles included a black or white population (30,33). Two articles included cohorts of The Study of Women's Health Across the Nation (SWAN) (22,27), which included different ethnicities (Caucasian, African American, Hispanic, Chinese & Japanese). The proportion of hormone therapy users (either hormone replacement therapy or hormonal contraception) varied from 16 to 33% (16,21,30). The ONM was self-reported in all the included studies.

Women who underwent gynaecological surgery (oophorectomy, hysterectomy) or regularly took hormones (oral contraception, hormonal replacement therapy) were excluded from most studies. Other exclusion criteria of individual studies were pregnancy, infertility, unproved fertility, other endocrine disorders, or irregular/unpredictable menstrual cycles at the time of first assessment. The exclusion criteria for each study are summarized in **Supplementary Table 1** (25).

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Prediction models. Risk prediction models of included articles used biomarkers (AMH, FSH, Estradiol, Inhibin B), imaging (transvaginal ultrasonography to obtain antral follicle count), menstrual cycle irregularities/climacteric symptoms, or lifestyle factors (body mass index (BMI), smoking). All biomarkers were assessed at a single time point, which is usually at baseline. More than half of the articles (n = 8) developed and validated both the univariable and multivariable model (16-20,29,30). For each of these individual studies, at first, the accuracy of prediction was checked for the individual predictors. Next, the predictors with the best accuracy were combined to explore whether it improved the prediction of ONM. Five articles elaborated only multivariable models (22,27,29,32,34), and one article (33) provided univariable model. With the exception of one article (27), all articles had the prediction of ONM as an outcome. Menopause in the included articles was defined as 12 consecutive months without menstruation (3), and was based on self-reporting. Greendale et al. (27) validated risk prediction models that predicted the probability of a woman having crossed 2 years before, 1 year before, and the final menstrual period (FMP) itself. Four studies also provided the risk prediction of onset of early menopause (16,21,31,34). Nine articles reported C-Statistic (16-18,20,21,30-32,34), one article reported AUC (19) and another one only sensitivity/specificity (33), and three articles reported AUC/sensitivity/specificity (22,27,29). A 95% confidence interval was reported in eight articles (16,19,20,22,31-34).

Prediction horizon. The majority (8 of 14) of articles (16-18,20,21,31,32,34) did not report the prediction time of menopause onset. Finkelstein et al. (22) and Greendale et al. (27) used prediction times of 1, 2 and 3 years till women became menopausal, whereas Gohari et al. (29) provided forecast information in 3, 4 or 5 years. Kim et al. (30) and van Rooij et al. (19) predicted the time to menopause to be 5 and 4 years, respectively. Taylor et al. (33) used a prediction time that ranged within 2 or 4 years.

Study design. All articles were of prospective cohort design.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Sample size. The number of participants ranged from 81 to 2434. The mean follow-up period was 10.2 years. During follow-up, a mean of 34.3% of women reached menopause. The number of events (occurrence of menopause) were clearly reported in 10 articles (16-18,20,21,29-32,34), and ranged from 46 to 1'298. Four articles (19,22,27,33) did not report the number of women experiencing menopause during the follow-up.

Outcome assessment. One model (33) was developed using only self-reported data (menstrual cycle irregularities). However, most of the models (13 of 14) were developed using a combination of self-reported (interview or questionnaire) and biological measurements (e.g., hormones and antral follicle count).

Predictors. There were overall 180 models tested. In total, 20 different predictors were included. Most models included a set of similar predictors, consisting of age, (n = 121), Anti-Müllerian hormone (AMH) (n = 87), and follicle-stimulating hormone (FSH) (n = 44). Other commonly identified predictors are shown in **Supplementary Figure 1** (25).

Predictors of ONM in risk prediction models

Most common predictors

The three most common predictors were age, AMH and FSH, and findings on these predictors are summarized in **Table 1.**

Age. Age was the most common investigated predictor, used in 121 models (from a total of 180 models). Age as the only predictor was used in 7 studies (17-21,30,33); in 6 studies age as a sole predictor showed AUC or C-statistics ranging from 0.82 – 0.88 (17-21,30). Another study performed by Taylor et al. (33) showed poor sensitivity and good specificity of age in forecasting menopause within 2 or 4 years (sensitivity 42.4% (95% CI 26.8-60.0) and 40% (95% CI 28.3-52.3), specificity 87.4% (95% CI 82.9-91.1) and 94.8% (95% CI 90.2-97.4), respectively (33).

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

AMH. AMH was used in 87 predictive models, making it the second most common investigated predictor. AMH as a sole predictor of menopause onset provided a C-statistics ranging from 0.70 to 0.87 (16-19). Adding age into the models using AMH improved predictive ability across all articles (C-statistics/AUC 0.86 – 0.90) (17-19) **Supplementary Materials 2** (25).

FSH. FSH was used in 44 prediction models (of which 3 in univariable models) (17-19). The univariable models of Broer et al. (17) and Depmann et al. 2016 (18) using FSH as a predictor (18) showed C-statistics of 0.70 and 0.66, respectively. The AUC of van Rooij et al. (19) was 0.72 (95% CI 0.56-0.88). After adjusting for age, the model performance of all three mentioned articles improved (Depmann 0.85, Broer 0.88, and van Rooij 84) **Supplementary Materials 2** (25).

Final menopause prediction models

In the following section, we present the final models reported by the authors and /or the models with the best performance generated from the included studies; the findings and whether these final models were validated and/ or calibrated are summarized in **Table 2**. Overall C-statistics/AUC ranged from 0.710 – 0.987, while one article (33) showed a sensitivity and specificity of 93.9% (95% CI 80.6-98.9) and 90.8% (95% CI 86.7-94.0), respectively. Age and AMH were present in most included studies. Of the included studies, only six (16,19,21,22,30,31) and one (32) studies validated their models internally and externally, respectively, and 5 studies performed calibration (16,27,29,31,32). Of which, while all of the above-mentioned studies performed validation for their final models (16,19,21,22,30-32), only the study by Gohari et al. (29) performed calibration for their final model.

Best model performance of Broer et al. (17) and Tehrani et al. 2014 (32), consisting of age and AMH, showed a C-statistics of 0.90 and 0.88 (95% CI 0.83 – 0.94), respectively. De Kat et al. (16) and Gohari et al. (29) models both included AMH at a certain age (35 and 43 years, respectively). The C-statistics and AUC were 0.71 (95% CI 0.69 – 0.72) and 0.987, respectively. Addition of AMH decline rate, oral contraception use or smoking to De Kat's model did not have any impact on the model performance. Best model performance presented by Depmann et al. 2016 (18) included age, AMH and smoking (C-statistic 0.87). There was no change in performance between the different time-points

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

of age or smoking (baseline versus follow up). Dollemann's (2014) (20) final model including daughter's age and AMH and mother's age of natural menopause (ANM) presented a C-statistic of 0.92 (95% CI 0.88 – 0.96). The study showed that AMH was a more accurate menopause predictor than mother's age at menopause. Dolleman et al., 2015 (21) presented a multivariable model with good performance consisting of age, AMH, BMI, packyears of smoking and menstrual cycle status (C-statistic 0.91). A similar model by Finkelstein et al. study with age, AMH and BMI presented an AUC of 0.896 (95% CI 0.889 – 0.903) (22).

The final model by Greendale et al. (27) included FSH, Estradiol, age, menopause transition stage, race/ethnicity and time of venipuncture (in early follicular phase) either 2 years before final menstrual period (FMP), 1 year before FMP or at FMP. The models with different prediction times showed AUC of 0.902 – 0.945.

Kim et al. combined age with different hormones (AMH and FSH) and AFC and presented a C-statistic of 0.95.

Taylor (33) used six definitions of menstrual variability for predicting time to natural menopause within 2 or 4 years (i. more than 90 days since the most recent menstrual period; ii. 60 or more days of amenorrhea during the previous years; iii. cycle lengths that varied by 19 or more days; iv. cycle lengths too variable to report a usual length; v. cycles less regular than they had been at age 40; vi. change in the duration or heaviness of menstrual flow compared with age 40). The second definition of menstrual variability showed the highest sensitivity (93.9%, 95% CI 80.6 - 98.9%) and specificity (90.8%, 95% CI 86.7 – 94%) within 2 years.

Tehrani et al. 2020 (31) showed that prediction of age at menopause could be improved by multiple AMH measurements. Their final prediction model included age, AMH, annual AMH decline rate and sample storage time, and showed a C-statistic of 0.81 (95% CI 0.79 – 0.83). Van Rooij's (19) prediction model, which included age, AMH and Inhibin B, resulted in AUC of 0.92 (95% CI 0.86 – 0.99).

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

In a recent study done by Tehrani et al. 2021 (34), AMH measurement at different time points, baseline age, storage time of AMH and annual AMH decline rate were included in the final model, resulting in a C-statistic of 0.85 (95% CI 0.83-0.87).

Detailed explanation of the risk prediction models of the included studies are presented in **Supplementary Table/ Materials 2** (25).

ROB assessment

The detailed summary of ROB assessment based on each domain and its related sub-domains for the included studies is presented in **Supplementary Table 3** (25). All included studies were classified as high ROB (**Figure 2 and Figure 3**) (16-22,27,29-34). ROB assessment based on each domain showed that all of the included articles had a low ROB for "participants" domain, demonstrating that participants enrolled in the articles were representative for the models' targeted populations (**Figure 2**).

Thirteen articles (16-22,27,29-32,34) had a high ROB for "predictors" domain as predictors were not defined in a similar way for all participants or different assessment methods were used or since all of the predictors would not be available at the time the model intended to be used for prediction. The study performed by Taylor et al. (33) was the only study that all of its predictors would be available at the time the model is intended to be used for prediction (**Figure 2**). The ROB was unclear for 10 articles for "outcome" domain (**Figure 2**) (16-18,20,21,30-34); Eight out of 10 articles did not provide any information on the time interval between predictor assessment and outcome determination (16-19,22,29,31,34). The article done by Kim et al. did not provide any information on whether a prespecified or standard outcome definition (onset of menopause) was used or whether predictors were excluded from outcome definition (30). In addition, another article had an unclear ROB for this domain due to the lack of information on whether predictors were excluded from the outcome definition or whether outcome was determined without knowledge of predictors information (33). Twelve of the 14 articles were at high ROB for the "analysis" domain and the two remained articles were scored as unclear ROB for this domain (**Figure 2**). Number of participants with the

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

outcome (those who reached menopause) was not provided in 4 articles and subsequently the ROB was considered as unclear for this sub-domain (19,22,27,33), while the article performed by Kim et al. (30) was at high ROB due to the limited number of participants with the outcome relative to the number of candidate predictor parameters. Three articles did not provide information on whether all enrolled participants were included in the analysis and were scored as unclear for this sub-domain (19,27,29), nevertheless, the article done by Finkelstein et al. (22) was at high ROB for this sub-domain due to the inappropriate exclusion of 1108 enrolled women from the final analysis. Only three articles presented sufficient information on handling the participants with missing data appropriately and were scored as low ROB (16,21,34). Only one study was at high ROB due to the selection of predictors based on univariable analysis (33), while the other 13 articles were at low ROB for this sub-domain (16-22,27,29-32,34). Although four articles did not present information on complexities in the data analysis (21,29,31,33) and four articles used inappropriate analytical methods (19,22,27,30) and were scored as unclear or high ROB, respectively, six articles accounted for complexities in the data analysis appropriately (16-18,20,32,34). Regarding the complexities in the data analysis, three out of six articles (17,18,32) have used a Weibull survival model analysis which accounts for the length of follow-up and delayed entry of participants rather than logistic models in which predictions are not specific to a particular time period and the length of follow-up is being ignored. In addition, three other articles (16,20,34) have used a time-to-event analysis, such as Cox regression, which accounts for censored participants up to the end of their follow-up rather than logistic regression models which simply exclude participants with incomplete follow-up. Nine articles did not assess calibration and were at high ROB (17-22,30,33,34), indicating that the performance or the ability of their model to provide accurate individuals probabilities is unsatisfactory. Seven articles have used some form of validation and were scored as low ROB (16,19,21,22,30-32). Of which, 6 studies have performed internal validation methods such as bootstrapping (16,19,21,22,30,31) and only one study used cross-validation method (32). Only three articles reported that the predictors and regression coefficients in the final model correspond to reported results from multivariable analysis and subsequently were scored as low ROB (20,21,31).

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Concerns regarding applicability of the presented models are presented in **Supplementary Figures 2 and 3** (25). Assessment of applicability to "participants", "predictors", and "outcome" domains revealed that four, two, and eight articles were rated with high, unclear, and low overall applicability concern, respectively.

DISCUSSION

Summary of main findings

This study systematically reviewed the literature and identified 14 articles describing 180 either univariable or multivariable risk prediction models of menopause onset. These models varied significantly regarding participant inclusion, sample size, predictors, endpoints, and model performance. The different predictors included demographics, lifestyle, biomarkers, imaging, and clinical features of menopause transition. In total, 13 of the 14 articles used more than one predictor in their models. The most commonly used predictor was age, followed by AMH and FSH.

While there are some narrative reviews on this topic, they have major limitations. First, they have not performed a systematic appraisal of evidence, which could lead to bias. Our study provides the first comprehensive systematic review on the topic, using well-defined inclusion and exclusion criteria. Second, previous studies have focused on a limited number of predictors like AMH (35) or restricted their results to Western European countries (36). Of the 14 included articles, most were conducted on populations from the Netherlands (n = 7), the USA (n = 4) or Iran (n = 5). Risk prediction models may, however, act differently in populations of different racial or ethnic backgrounds. Systematic efforts to validate models in other populations are essential. Women who were using hormone therapy for medical reasons or as contraceptives or hormonal replacement therapy were in general excluded from the articles. However, De Kat et al., Dolleman et al. 2015 and Kim et al. included women who were taking estrogen (oral contraception, hormonal replacement therapy) (16,21,30). Dolleman et al. 2015 also included pregnant women (21).

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Pregnancy or exogenously supplied hormones affect the level of hormones, which can affect model`s performance.

In addition, more than half of the included studies did not report the time interval between the assessment of predictors and outcome. The interpretation of the results is strongly dependent on the prediction horizon, and among the articles reporting time of prediction, the focus was on short-term prediction of menopause onset. While long-term prediction models can be useful in terms of public health interventions to identify women at risk of developing early menopause, such models are lacking. In addition, predicting the onset of menopause becomes more difficult the longer this time lies in the future and is therefore more prone to error, which needs to be explored and addressed by future studies.

While overall, the presented prediction models were at high ROB, all the included articles were at low concern regarding inclusion and exclusion of participants and none of them included participants who already had the outcome (menopausal women). Inclusion of participants who already have the outcome may result in biased and overestimated predictive performance. Most of the articles suffer from serious analytical limitations. For instance, 10 articles excluded participants with missing data from the final analysis, which may cause biased predictor-outcome associations and result in misleading model performance. In addition, three articles only presented classification measures, including sensitivity and specificity, to show model predictive performance without providing C-statistics (22,27,33). It is worth mentioning that the estimation of classification requires the consideration of one or more thresholds in the range of the model-predicted probabilities, which typically leads to misinformation since the whole range of predicted probabilities of the model is not fully considered, and also, the selection of thresholds might be data-driven rather than prespecified clinical cut-offs. The introduction of thresholds helps reporting the performance of predictive models at probability thresholds

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

that may be clinically relevant, which does not include the entire range of the model-predicted probabilities. This approach may result in critical bias in the estimated classification measures, especially when thresholds are selected to show the maximum performance (37,38). Calibration was not assessed in nine studies, and it is worth mentioning that only the study by Gohari et al. (29) performed calibration for all models, while the remaining four studies reported calibration for a limited number of models (16,27,31,32), which indicates that the performance of their prediction models to provide accurate probabilities are doubtful. Poorly calibrated or not calibrated risk prediction models can be misleading and potentially harmful for clinical decision-making guidelines. Moreover, while over half of the included articles presented some form of internal validation that accounts for the overfitting of the developed model and optimism in its predictive performance, the evidence from the external validation of the presented models is still unsatisfactory. Only one article performed by Tehrani et al. 2014 (32) replicated their models into another cohort in a cross-validation approach. In addition, most of the presented models did not provide 95% CIs of the estimates. The majority of articles did not test whether the differences between models in predicted estimates (e.g., AUC) were significantly different, thus precluding the interpretation of results.

Strengths and limitations

To our best knowledge, the current study is the first study summarizing the risk prediction models of menopause onset. Due to a large number of prediction models (n = 180) and the limited number of included articles (n = 14), we were not able to perform a meta-analysis. In addition, only eight articles reported a 95% confidence interval of sensitivity, specificity, AUC and C-statistics. While we applied a comprehensive search strategy and screened references of the final included articles to capture all relevant studies, we cannot exclude the possibility there has been underreporting of negative results. Although all of the included studies were at high ROB, their applicability according to PROBAST was satisfactory. Applicability concerns whether the review question and

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

the included studies are a good match regarding its domains (participants, predictors, outcome). Eight out of 14 studies were rated as low concern (16,19,21,22,27,29,31,34) and two (30,33) and four studies (17,18,20,32) were rated as unclear and high concern with regard to applicability, respectively. Reviewing the risk prediction models with high applicability indicates that most of them have included age and AMH in their models. Thus, such predictors can be used by clinicians to get insights on prediction of ONM to make better clinical judgment for health-related consequences of menopause onset. In case of limited resources to measure biomarkers, the models presented by Taylor et al. (32) may be used by clinicians to predict ONM considering that the models were based on simple questions related to menstrual variability which can be obtained during the physical examination. However, it is important that clinicians and patients understand the limitation of such prediction models when interpreting the results.

Future prospective

Due to increased longevity, there is a trend of postponing childbearing; therefore, prediction of menopause onset could improve early informed decision-making regarding family planning. Predicted early menopause could emphasize the need for timely prevention of bone demineralization and cardiovascular and neurological disease, whereas the prediction of late menopause would open options for preventive management of breast, endometrial and ovarian cancer. Future studies should be performed in different countries and include premenopausal women of different ethnicities with wider age ranges (that accounts for both early and late menopause) to improve the model generalizability. Further, increased knowledge of the variation of menopausal onset and performance of prediction models across different ethnicities and medical history will allow for better personalized prediction, that is, providing an accurate prediction for each single patient. Efforts are required to avoid poor calibration when developing prediction models and evaluate calibration when validating them, and improvement of methods that deal with analysis. Currently, there is limited evidence in the prediction of early onset of menopause, and/or long-term prediction of menopause onset.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

CONCLUSION

This systematic review summarizes the available literature trying to predict the risk of ONM in different settings and highlights numerous methodological deficiencies of the current prediction models. While the predictive performance of the presented models is limited, their applicability is satisfactory. Age and AMH were the most used predictors in the presented models and in general showed to have high predictability for menopause onset. The current study can help guide future studies to develop methodologically sound predictors of ONM.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Source of funding: None

Acknowledgment: HR-D has received funding for his PhD studies by the Swiss National Science Foundation under the "Spirit grant" (SNSF IZSTZ0_190277). PFR has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 801076, through the SSPH+ Global Ph.D. Fellowship Programme in Public Health Sciences (GlobalP3HS) of the Swiss School of Public Health. The funding agencies had no role in the study (conceptualization, design, data collection, analysis, and writing).

Data availability: Data available within the main article and its supplementary materials.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

References

1. Te Velde ER PP. The variability of female reproductive ageing. *Human reproduction update.* Vol 8(2)2002:141-154.
2. Soules MR, Sherman, S., Parrott, E., Rebar, R., Santoro, N., Utian, W., & Woods, N. Executive summary: stages of reproductive aging workshop (STRAW). *Menopause.* Vol 8(6). Park City, Utah2001:402-407.
3. WH. O. Research on the menopause in the 1990s: report of a WHO scientific group: World Health Organization. 1996.
4. Shuster LT RD, Gostout BS, Grossardt BR, Rocca WA. Premature menopause or early menopause: long-term health consequences. *Maturitas*2010:161-166.
5. Jung SJ SA, Kang D. Menarche age, menopause age and other reproductive factors in association with post-menopausal onset depression: Results from Health Examinees Study (HEXA). *Journal of affective disorders*2015:127-135.
6. Svejme O AH, Nilsson JÅ, Karlsson M. Early menopause and risk of osteoporosis, fracture and mortality: a 34- year prospective observational study in 390 women. *BJOG: An International Journal of Obstetrics & Gynaecology*2012:810-816.
7. Muka T O-WC, Kunutsor S, Laven JS, Fauser BC, Chowdhury R, et al. Association of age at onset of menopause and time since onset of menopause with cardiovascular outcomes, intermediate vascular traits, and all-cause mortality: a systematic review and meta-analysis. *JAMA cardiology*2016:767-776.
8. Appiah D SP, Demerath EW, Loehr LR, Chang PP, Folsom AR. . ., :e003769. Association of age at menopause with incident heart failure: a prospective cohort study and meta- analysis. *Journal of the American Heart Association*2016.
9. EB. G. The timing of the age at which natural menopause occurs. *Obstetrics and Gynecology Clinics*:425-440.
10. Allen RH CC. Contraception for midlife women. *Menopause*2016 111-113.
11. MacArthur T BG, Ayers C. Menopausal women requesting egg/embryo donation: examining health screening guidelines for assisted reproductive technology. *Menopause*2016:799-802.
12. Bradley LD GN-A. The medical management of abnormal uterine bleeding in reproductive-aged women. *American journal of obstetrics and gynecology*2016:31-44.
13. Aydogan B, & Mirkin, S. The utility of measuring anti-Müllerian hormone in predicting menopause. *Climacteric*2015:777-789.
14. Gold EB BJ, Crawford S, Samuels S, Greendale GA, Harlow SD, et al. Factors associated with age at natural menopause in a multiethnic sample of midlife women. *American journal of epidemiology*2001:865-874.
15. Li L WJ, Pu D, Zhao Y, Wan C, Sun L, et al. Factors associated with the age of natural menopause and menopausal symptoms in Chinese women. *Maturitas*2012:354-360.
16. De Kat AC, Van Der Schouw YT, Eijkemans MJ, Broer SL, Verschuren WM, Broekmans FJ. Can menopause prediction be improved with multiple AMH measurements? results from the prospective Doetinchem cohort study. *The Journal of Clinical Endocrinology & Metabolism*. 2019;104(11):5024-5031.
17. Broer S, Eijkemans M, Scheffer G, Van Rooij I, De Vet A, Themmen A, Laven J, De Jong F, Te Velde E, Fauser B. Anti-Müllerian hormone predicts menopause: a long-term follow-up study in normoovulatory women. *The Journal of Clinical Endocrinology & Metabolism*. 2011;96(8):2532-2539.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

18. Depmann M, Eijkemans M, Broer S, Scheffer G, Van Rooij I, Laven J, Broekmans F. Does anti-Müllerian hormone predict menopause in the general population? Results of a prospective ongoing cohort study. *Human reproduction*. 2016;31(7):1579-1587.
19. van Rooij IA, den Tonkelaar I, Broekmans FJ, Looman CW, Scheffer GJ, de Jong FH, Themmen AP, te Velde ER. Anti-müllerian hormone is a promising predictor for the occurrence of the menopausal transition. *Menopause*. 2004;11(6 Part 1 of 2):601-606.
20. Dolleman M, Depmann M, Eijkemans M, Heimensem J, Broer S, Van Der Stroom E, Laven J, Van Rooij I, Scheffer G, Peeters P. Anti-Müllerian hormone is a more accurate predictor of individual time to menopause than mother's age at menopause. *Human reproduction*. 2014;29(3):584-591.
21. Dólleman M, Verschuren WM, Eijkemans MJ, Broekmans FJ, Van Der Schouw YT. Added value of anti-Müllerian hormone in prediction of menopause: results from a large prospective cohort study. *Human Reproduction*. 2015;30(8):1974-1981.
22. Finkelstein JS, Lee H, Karlamangla A, Neer RM, Sluss PM, Burnett-Bowie S-AM, Darakananda K, Donahoe PK, Harlow SD, Prizand SH. Antimullerian hormone and impending menopause in late reproductive age: the study of women's health across the nation. *The Journal of Clinical Endocrinology & Metabolism*. 2020;105(4):e1862-e1871.
23. Muka T, Glisic M, Milic J, Verhoog S, Bohlius J, Bramer W, Chowdhury R, Franco OH. A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research. *European journal of epidemiology*. 2020;35(1):49-60.
24. Moher D, et al. "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement.". *Systematic reviews* 2015:1-9.
25. Raeisi-Dehkordi H, Kummer S, Raguindin PF, Dejanovic G, Taneri PT, Cardona I, Kastrati L, Minder B, Voortman T, Marques-Vidal P, Dhana K, Glisic M, Muka T. Data from: Supplementary Materials for: Risk prediction models of natural menopause onset: a systematic review. Uploaded June 24, 2022. BORIS (Bern Open Repository and Information System). https://boris.unibe.ch/170883
26. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, Reitsma JB, Collins GS. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS medicine*. 2014;11(10):e1001744.
27. Greendale GA, Ishii S, Huang M-H, Karlamangla AS. Predicting the timeline to the final menstrual period: the study of women's health across the nation. *The Journal of Clinical Endocrinology & Metabolism*. 2013;98(4):1483-1491.
28. Wolff RF, et al. "PROBAST: a tool to assess the risk of bias and applicability of prediction model studies.". *Annals of internal medicine*2019.
29. Gohari MR, Ramezani Tehrani F, Chenouri S, Solaymani-Dodaran M, Azizi F. Individualized predictions of time to menopause using multiple measurements of antimüllerian hormone. *Menopause*. 2016;23(8):839-845.
30. Kim C, Slaughter JC, Wang ET, Appiah D, Schreiner P, Leader B, Calderon-Margalit R, Sternfeld B, Siscovick D, Wellons M. Anti-Müllerian hormone, follicle stimulating hormone, antral follicle count, and risk of menopause within 5 years. *Maturitas*. 2017;102:18-25.
31. Ramezani Tehrani F, Bidhendi Yarandi R, Solaymani-Dodaran M, Tohidi M, Firouzi F, Azizi F. Improving prediction of age at menopause using multiple anti-Müllerian hormone measurements: the Tehran Lipid-Glucose Study. *The Journal of Clinical Endocrinology & Metabolism*. 2020;105(5):1589-1598.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

32. Ramezani Tehrani F, Dolleman M, van Disseldorp J, Broer S, Azizi F, Solaymani-Dodaran M, Fauser B, Laven J, Eijkemans M, Broekmans F. Predicting menopausal age with anti-Müllerian hormone: a cross-validation study of two existing models. *Climacteric*. 2014;17(5):583-590.
33. Taylor SM, Kinney AM, Kline JK. Menopausal transition: predicting time to menopause for women 44 years or older from simple questions on menstrual variability. *Menopause*. 2004;11(1):40-48.
34. Ramezani Tehrani F, Sheidaei A, Firouzi F, Tohidi M, Azizi F, Behboudi-Gandevani S. Does Anti Mullerian Hormone Decline Rate Improve the Predication of Age at Menopause? *Frontiers in Endocrinology*. 2021:1104.
35. De Kat AC, Broekmans FJ, Lambalk CB. Role of AMH in Prediction of Menopause. *Frontiers in Endocrinology*. 2021:1078.
36. Tanbo TG, Fedorcsak PZ. Can time to menopause be predicted? *Acta obstetricia et gynecologica Scandinavica*. 2021;100(11):1961-1968.
37. Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. *Clinical chemistry*. 2008;54(4):729-737.
38. Ewald B. Post hoc choice of cut points introduced bias to diagnostic research. *Journal of clinical epidemiology*. 2006;59(8):798-801.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Legend of figures

Figure 1. Study selection procedure.

Figure 2. Risk of bias assessment (ROB) summary of the included studies.

Figure 3. Risk of bias graph of the included studies.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Table 1. Most common predictors of menopause onset.

Study	Sample size	Number and percentage of women reaching menopause	Follow up period (years)	Time period	Prediction time (years)	Model	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	C-Statistics (95% CI)
Broer, 2011	185	48 (25.9%)	11	1996-2010	n.r.	Age	-	-	-	87
						AMH	-	-	-	86
						FSH	-	-	-	70
De Kat, 2019	2434 [a]	1298 (53%)	20	1987-2019	n.r.	AMH	-	-	-	70 (68-72)
Depmann, 2016	155	81 (52.2%)	14	1992-2013	n.r.	Age	-	-	-	85
						AMH	-	-	-	78
						FSH	-	-	-	66
Dolleman, 2014 (**study group 2**)	150	46 (30.6%)	12	n.r.	n.r	Age	-	-	-	84 (78-90)
						AMH	-	-	-	86 (81-91)
Dolleman, 2015	1163 [b]	527 (45.3%)	10	1993-2007	n.r	Age	-	-	-	88
Kim, 2016	426 [c]	n = 55 (13%)	10	2001-2011	5	Age	-	-	-	83

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

	359 [d]	n.r.				Age	-	-	-	82
	327 [e]					Age	-	-	-	81
	306 [f]					Age	-	-	-	86
Taylor, 2004	326	n.r.	4.8	1993-1997	Within 2 years	Age ≥ 50 years	42.4 (26.8-60.0)	87.4 (82.9-91.1)	-	-
					Within 4 years	Age ≥ 50 years	40 (28.3-52.3)	94.8 (90.2-97.4)	-	-
Van Rooij, 2004	81	N = 14 had cycle fluctuations	5	1996-?	4	Age	-	-	82 (71-93)	-
						AMH	-	-	87 (79-96)	-
						FSH	-	-	72 (56-88)	-

Abbreviations: CI, Confidence interval; AUC, Area under the curve; FSH, Follicle stimulation hormone; AMH, Anti-Mullerian Hormone; AFC, Antral follicle count; ANM, age at natural menopause; OC, oral contraceptive; ANM, age at natural menopause; n.r., not reported

[a] Women who were using oral contraceptives were also included in the final analysis.

[b] Pregnant women, women who were taking hormonal replacement therapy and OC users were also included in the final analysis.

[c] All women

[d] Women not using estrogen

[e] Women with regular menses only

[f] No women with dominant follicle

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Table 2. Final menopause prediction models.

Study	Sample size	Number and percentage of women reaching menopause	Follow up period (years)	Time period	Prediction time (years)	Model	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	C-Statistics (95% CI)
Broer, 2011	185	48 (25.9%)	11	1996-2010	n.r.	AMH + age*	-	-	-	90
De Kat, 2019[d]	2434 [a]	1298 (53%)	20	1987-2019	n.r.	AMH at age 35*	-	-	-	71 (69-72)
Depmann, 2016	155	81 (52.2%)	14	1992-2013	n.r.	AMH + age + smoking*	-	-	-	87
Dolleman, 2014 (**study group 2**)	150	46 (30.6%)	12	n.r.	n.r	Daughter's age + mother's ANM + daughter's AMH*	-	-	-	92 88-96
Dolleman, 2015[d]	1163 [b]	527 (45.3%)	10	1993-2007	n.r	AMH + age + BMI + packyears of smoking + menstrual cycle status	-	-	-	91
Finkelstein, 2020[d]	1537	n.r	2	1996-1998	3	AMH + age + BMI*	-	-	89.6 (88.9-90.3)	-
Gohari, 2016[f]	266	63 (23.7 %)	6.5	1999-2006	3	AMH + Age at making prediction was 43*	-	-	98.7	-

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Greendale, 2013[d]	552	n.r.	10	1996-2007	2 years before FMP [c]	FSH + Estradiol + age + menopause transition stage + race/ethnicity + time of venipuncture (in early follicular phase = cycle days 2-5 = in-window)	-	-	90.2	-
					1 year before FMP [d]	FSH + Estradiol + age + menopause transition stage + race/ethnicity + time of venipuncture (in early follicular phase = cycle days 2-5 = in-window)	-	-	92.6	-
					FMP [e]	FSH + Estradiol + age + menopause transition stage + race/ethnicity + time of venipuncture (in early follicular phase = cycle days 2-5 = in-window)	-	-	94.5	-
Kim, 2016[d]	306 [c]	n.r.	10	2001-2011	5	Age + AMH + FSH + AFC*	-	-	-	95
Taylor, 2004	326	n.r.	4.8	1993-1997	Within 2 years	60 or more days of amenorrhea during the previous year*	93.9 (80.6-98.9)	90.8 (86.7-94.0)	-	-
Tehrani, 2014[e]	266	n = 63, 23.7%	6	n.r.	n.r.	Age + AMH (TLGS model on TLGS data)*	-	-	-	88 (83-94)

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

Tehrani, 2020[d]	959	n = 529, 55.2%	14	1998-?	n.r.	Age + AMH + annual AMH decline rate + sample storage time	-	-	-	81 (79-83)
Tehrani, 2021	901	522	18	1998-2016	n.r.	AMH value + Log (annual AMH decline rate) (adjusted for age at baseline and storage time of AMH)	-	-	-	0.85 (0.83–0.87)
Van Rooij, 2004[d]	81	n = 14 had cycle fluctuations	5	1996-?	4	Age + AMH + Inhibin B	-	-	92 (86-99)	-

Abbreviations: CI, Confidence interval; AUC, Area under the curve; FSH, Follicle stimulation hormone; AMH, Anti-Mullerian Hormone; AFC, Antral follicle count; ANM, age at natural menopause; OC, oral contraceptive; ANM, age at natural menopause; n.r., not reported

[a] Women who were using oral contraceptives were also included in the final analysis.

[b] Pregnant women, women who were taking hormonal replacement therapy and OC users were also included in the final analysis.

[c] No women with dominant follicle

[d] Studies that performed internal validation for their final models.

[e] Studies that performed external validation.

[f] Studies that performed calibration for their final models.

* These models were selected for their best performance. Models without labelling were called final models by the authors.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

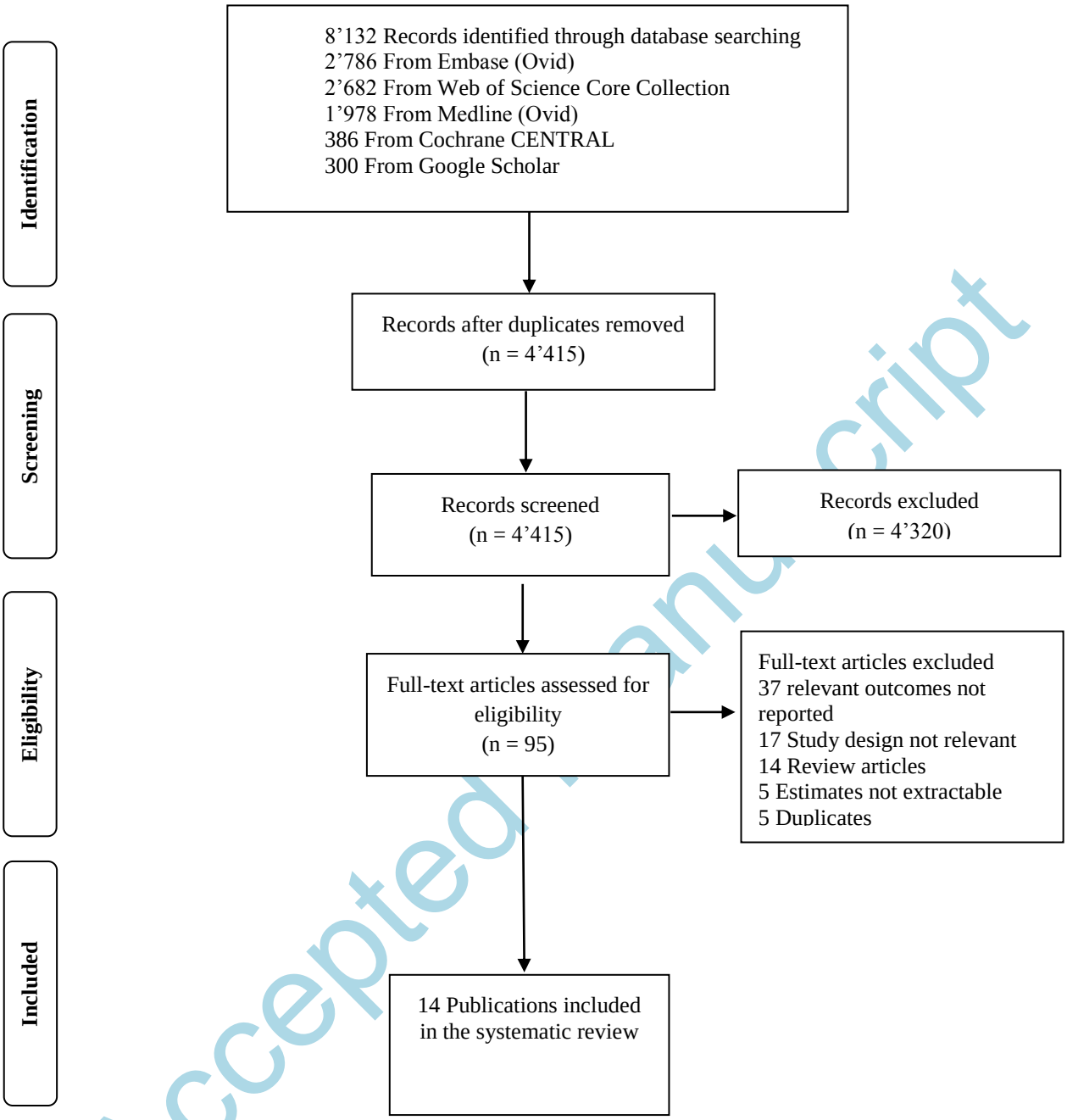

Figure 1. Study selection procedure.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

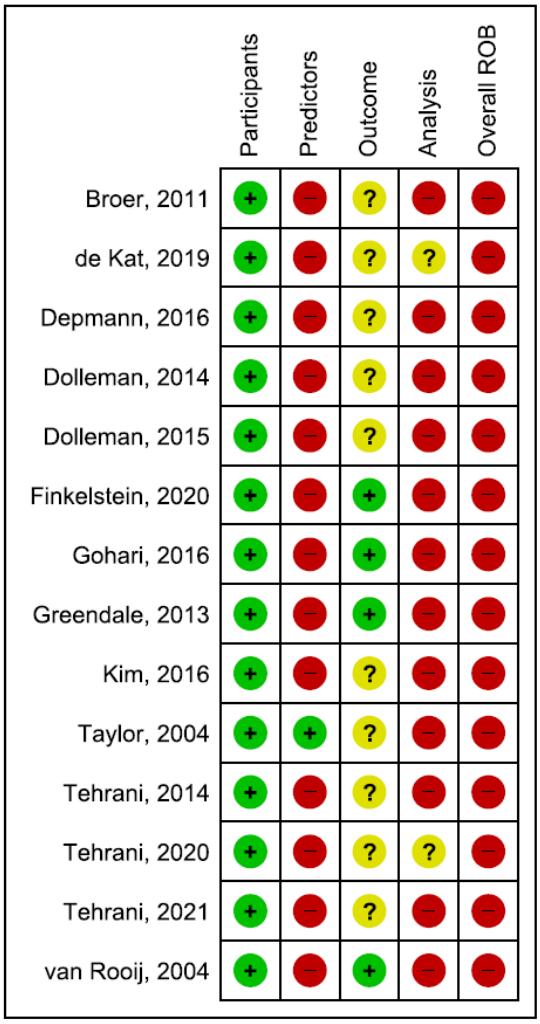

Figure 2. Risk of bias assessment (ROB) summary of the included studies.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022

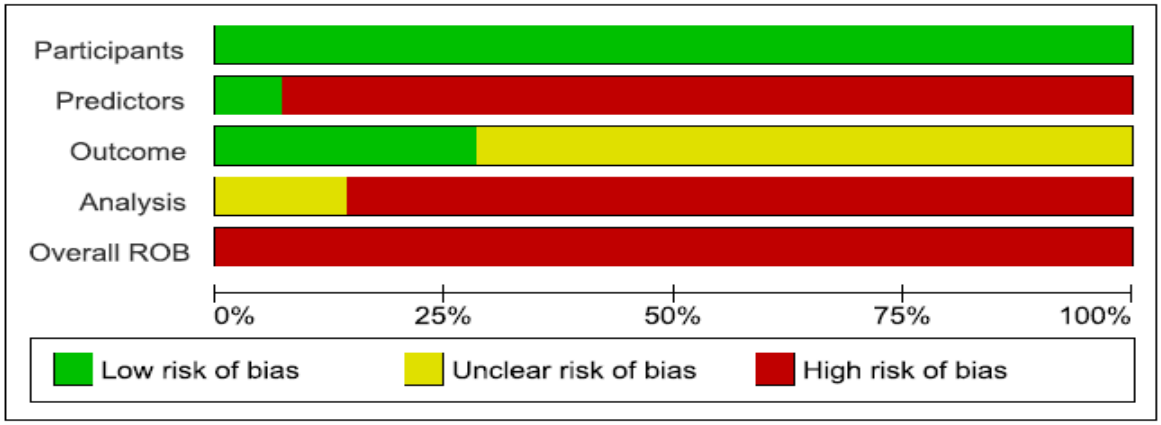

Figure 3. Risk of bias graph of the included studies.

Accepted Manuscript

Downloaded from https://academic.oup.com/jcem/advance-article/doi/10.1210/clinem/dgac461/6652568 by Universitaetsbibliothek Bern user on 04 August 2022