Article type : Original Article - Cardiovascular Medicine

Predictive performance of the CHA2DS2-VASc rule in atrial fibrillation: a systematic review and meta-analysis.

The CHA2DS2-VASc rule, a meta-analysis

Sander van Doorn, MD; Thomas P.A. Debray, PhD; Femke Kaasenbrood, MD; Arno W. Hoes, MD, PhD; Frans H. Rutten, MD, PhD; Karel G.M. Moons, PhD; Geert-Jan Geersing, MD, PhD

Julius Center for Health Sciences and Primary care, University Medical Center Utrecht, the Netherlands

Correspondence:

Sander van Doorn, MD

Julius Center for Health Sciences and Primary care,

University Medical Center Utrecht

PO box 85500

3508 AB Utrecht

The Netherlands

Email: S.vanDoorn@umcutrecht.nl

Phone: (+31) (0)88 756 8174, Fax: (+31) (0)88 756 9028

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jth.13690

Essentials

- The widely recommended CHA2DS2-VASc shows conflicting results in contemporary validation studies.
- We performed a systematic review and meta-analysis of 19 studies validating CHA2DS2-VASc.
- There was high heterogeneity in stroke risks for different CHA2DS2-VASc scores.
- This was not explained by differences between setting of care, or by performing meta-regression.

Abstract

BACKGROUND: The CHA2DS2-VASc decision rule is widely recommended for estimating stroke risk in patients with atrial fibrillation (AF) though validation studies show ambiguous and conflicting results.

OBJECTIVES: We aimed to (1) review existing studies validating CHA2DS2-VASc in AF patients not (yet) anticoagulated, 2) meta-analyze estimates of stroke risk per score, and 3) explore sources of heterogeneity across the validation studies.

METHODS: We performed a systematic literature review and random effects meta-analysis of studies externally validating CHA2DS2-VASc in AF patients not on anticoagulants. To explore between-study heterogeneity in stroke risk, we stratified studies to the clinical setting in which patient enrollment started, and performed meta-regression.

RESULTS: In total 19 studies were evaluated with over two million person-years of follow-up. In studies recruiting AF patients in hospitals, stroke risk for a score of zero, one and two were 0.4% (approximate 95% prediction interval (PI) 0.2 to 3.2%), 1.2% (95% PI 0.1 – 3.8%) and 2.2% (95% PI 0.03 – 7.8%), respectively. This was consistently higher than studies

recruiting patients from the open general population, with risks of 0.2% (95% PI 0.0 – 0.9%), 0.7% (0.3 – 1.2%) and 1.5% (95% PI 0.4 – 3.3%) for score zero to two respectively. Heterogeneity as reflected by the wide prediction intervals could not be fully explained by meta-regression.

CONCLUSIONS: Studies validating CHA2DS2-VASc demonstrate high heterogeneity in predicted stroke risks for different scores.

Keywords: atrial fibrillation, CHA2DS2-VASc, clinical prediction rule, meta-analysis, systematic review

Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia[1]and a major risk factor for ischemic stroke.[2] Anticoagulants – such as vitamin K antagonists (VKA) or direct oral anticoagulants (DOACs) – can effectively reduce stroke risk,[3,4] but their relative benefits and harms depend on the absolute risk of stroke while off treatment, given that they inherently carry a risk of (major) bleeding complications. Hereto, clinical decision rules have been developed to estimate stroke risk in AF patients, with the CHA2DS2-VASc rule as the most well-known example.[5] Published as an update to the CHADS2 rule[6], the CHA2DS2-VASc was first recommended in the 2010 ESC practice guideline.[7] A swift uptake in clinical practice was followed, but subsequent validation studies showed ambiguous and conflicting results. This is for instance exemplified by the ongoing debate on the optimal threshold below which stroke risk is low enough to omit anticoagulation.[8-12]

Therefore, the aim of the present study was to evaluate the current evidence-base of using CHA2DS2-VASc for predicting stroke in AF patients. Hereto, we performed the following steps: 1) review existing studies validating CHA2DS2-VASc for AF patients not (yet) anticoagulated, 2) meta-analyze estimates of the c-statistic and stroke risk per score, and 3) explore sources of heterogeneity across the validation studies.

Methods

Throughout the planning and conducting of this systematic review we followed the CHARMS recommendations for framing the review question, critical appraisal, and data extraction for systematic reviews of prediction modelling studies.[13] See Table 1 for details.

The CHA2DS2-VASc score

The CHA2DS2-VASc clinical decision rule was developed in 2010 by Lip et al. as an update to the original CHADS2 rule[6] by including additional predictors for stroke. Patients were assigned points for congestive heart failure (1 point), hypertension (1 point), age above 75 years (2 points), diabetes (1 point) and prior stroke (2 points), age above 65 (1 point), female sex (1 point) and vascular disease (1 point). Risk categories were defined using the total sum of scored points, and consisted of 'low' (0 points), 'intermediate' (1 point) and 'high (≥ 2 points). Using these categories, the c-statistic was 0.61 (0.51-0.70) in the derivation cohort.[5] See Data S1. No efforts were made to adjust the c-statistic for potential over-optimism.[14]

Data sources and search strategy

We performed a systematic search to identify all studies that validated the CHA2DS2-VASc rule in patients with non-valvular AF. Medline and Embase were searched from January 1st 2001 till March 1st 2017. The search syntax was based on the broad Ingui search filter for identifying prediction studies[15], and augmented with the filter by Geersing et al.[16] and the term 'Atrial Fibrillation' with its MeSH heading (Data S2). Cross-reference checks were performed using the reference lists of each selected article.

Study selection

As CHA2DS2-VASc was specifically developed to guide anticoagulant decision making, notably for selecting AF patients in whom anticoagulant therapy can be safely withheld, we focused on studies validating this decision rule in AF patients not already treated with anticoagulants. To identify articles eligible for this review, the following inclusion criteria were used:

- Original research articles on the external validation of CHA2DS2-VASc (i.e. validation in patients not used for the derivation of the score);
- Including adults aged > 18 years with non-valvular AF;
- AF patients not yet treated with anticoagulation, or data presented separately for those not anticoagulated. Treatment with antiplatelet therapy was allowed.
- Allowing for extraction of c-statistic and/or absolute stroke or thrombo-embolic risks at different risk scores of CHA2DS2-VASc.

Studies including patient populations that dictate specific treatment decisions regardless of the score on a clinical decision rule, e.g. those after cardiac surgery, with mechanical heart valves or mitral valve stenosis, after ablation or left appendage closure, were excluded.

A single reviewer (SvD) performed the study selection and included all eligible articles after consensus with a second reviewer (GJG).

If different articles used subsets of the same data source, the article studying the patient population most representative for our study domain was included after consensus between reviewers, or after consultation with the corresponding authors where needed.

Critical appraisal and risk of bias assessment

We critically appraised the selected studies according to the Checklist for critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) guidelines.[13] From the checklist we identified twelve items relevant for external validation

studies (see Table S1 for an overview). Two independent reviewers (SvD, FK) scored the risk of bias for each item (no risk of bias, risk of bias or unclear) and decided on a summary risk of bias estimate where studies without high risk on any item were considered at low risk of bias. Any disagreements were resolved by consensus with a third reviewer (GJG).

Data extraction and quantitative synthesis

Data extraction was independently performed by two reviewers (SvD, FK), disagreements were resolved by discussion. We stratified studies in those recruiting patients from an unselected general population setting – e.g. primary care databases or healthcare insurance data – and studies enrolling a selected subsample of AF patients, e.g. recruited during a hospital admission or during a visit to an outpatient cardiology department. We subsequently extracted the following information from each validation study, if reported:

- **Setting**: setting (e.g. general population or hospital care setting), locations and periods of recruitment;
- **Study characteristics**, i.e. the study design and the source of the data, the number of patients and total duration of follow-up, geographic region;
- **Outcomes**: type of outcomes studied (ischemic stroke; all strokes including hemorrhagic stroke; or thromboembolism, commonly defined as ischemic stroke, TIA, systemic embolism or a combination thereof);
- **Population characteristics**: the annual incidence of the main outcome, mean patient age, proportion of patients with (congestive) heart failure, hypertension, diabetes, a history of prior stroke or TIA, vascular disease, and proportion of females; the distribution across individual scores of CHA2DS2-VASc rule; and the proportion of patients using a platelet inhibitor;
- **Validation study results**: the c-statistic(s), the annual outcome risk per score, and corresponding estimates of uncertainty.

Data preparation

In accordance with previous recommendations, we rescaled the extracted c-statistic by applying the logit transformation.[17] If more than one c-statistic was reported, e.g. when calculated using aforementioned risk categories or one for a continuous score, the highest c-statistic was used. The error variance of the logit c-statistic was estimated from the reported confidence interval[18] or standard error (Delta method). If no information on uncertainty was reported, we used the approximation as reported by Debray et al.[17]
Furthermore, we rescaled annual stroke risk estimates by applying the square rooted transformation.[19] The corresponding variance was estimated using Poisson approximations and, again, applying the Delta method.[17]

Data analysis

We applied random effects meta-analysis using restricted maximum likelihood estimation (REML) to summarize estimates of model discrimination (logit c-statistic) and annual risk per score (square root risks).[17,20] In accordance with recent guidelines, confidence intervals were calculated using the Hartung-Knapp-Sidik-Jonkman method.[21] We calculated approximate 95% prediction intervals (95% PI) to ascertain the potential impact of between-study heterogeneity. These intervals indicate the range of performances (e.g. c-statistic or stroke rates per CHA2DS2-VASc score) that can be expected in future validation studies with similar characteristics as the ones included in our review. Additionally, we calculated the probability that the annual stroke risk was below a certain threshold if in 'real life' practice the CHA2DS2-VASc rule assigned an AF patient with a score 0, 1 or 2.[22]
Finally, we performed random effects meta-regression to investigate potential sources of heterogeneity. For study characteristics, we included the outcome under study, the number of person years of follow-up and the dichotomized risk of bias as covariates. For summarized

patient characteristics, covariates of interest were mean age of the study population, proportion of females, mean CHA2DS2-VASc score, prevalence of heart failure and prevalence of platelet inhibitor use. See Table S3 and Fig S3a-d. All analyses were performed with the package *metafor* (univariate meta-analysis) version 1.9-8 in R 3.3.0.

Results

Included studies

The process of study selection is shown in Figure 1. The initial search yielded 17,667 results, from which 8096 duplicates were discarded. After reading title and abstract we excluded 19,245 articles, primarily because these included patients outside the domain of interest (~60%), did not address risk prediction (~30%) or did not externally validate CHA2DS2-VASc (~5%).

In total, 126 studies were subjected to full text evaluation. Of these, inclusion criteria were not met in 107 studies, resulting in a final selection of 19 validation studies.

The key characteristics of each study are presented in Table 2. Seven studies (in total 163,610 AF patients with in total 365,501 person years of follow-up) were performed in AF patients recruited from the general (unselected) population, and twelve studies (in total 683,138 AF patients and 1,738,930 person years follow-up) included a subsample of patients from the hospital care setting.

The outcomes under study consisted of 1) ischemic stroke (ten studies), or 2) all thromboembolic events in eight studies, that is, ischemic strokes and systemic thromboembolism (defined as peripheral embolism in six studies and peripheral embolism and/or pulmonary embolism in two studies), or 3) all types of stroke (ischemic plus

hemorrhagic stroke) in one study. Most studies originated from Europe and North America, and five were performed in East Asia. The number of included patients ranged from 154 to 198,697 and the follow-up time from 11 to 53 months (Table 2).

Risk of bias

The risk of bias of the included studies is summarized in Fig S1. For details on individual studies, see Table S2. Overall, two of seven studies performed in the general population were considered at risk of bias, as was the majority of studies enrolling patients from a hospital care setting. In general, the source of the data and the eligibility criteria caused no concern for bias. Some studies did not provide information on the use of antiplatelet therapy. This could induce biased results on the predictive accuracy of CHA2DS2-VASc since, albeit to a limited extent, antiplatelet therapy may reduce the occurrence of stroke and thus underestimate the predictive accuracy of the rules.[3] The definition and measurement of the predictors – i.e. the variables included in the CHA2DS2-VASc rule – and the outcome under study frequently differed across studies. Mostly, these were clearly defined. Predictors were mostly assessed blinded for the outcome. No study explicitly reported whether the outcomes were assessed blinded for the initial CHA2DS2-VASc score, potentially introducing bias for outcomes requiring subjective interpretation such as TIA.[13]

In six studies (two in unselected and four in selected patients) the number of outcome events was lower than the generally recommended ~100 events for validation of a decision rule.[42] In addition, the amount and handling of missing data was unclear in the majority of studies. As data are seldom missing completely at random, inadequate handling of missing data could introduce bias.[43-45]

Meta-analysis of discriminative ability

In both populations, there was substantial between-study heterogeneity. In studies enrolling patients from the general population, we found an average c-statistic of 0.64 (95% CI 0.56 to 0.71). The variation in discriminative performance of CHA2DS2-VASc across studies is indicated in Fig S2a-b and reflected by the wide approximate 95% prediction interval (95% PI) that ranged from 0.45 to 0.79.

In studies recruiting from a hospital care setting, we found a somewhat higher average c-statistic of 0.71 (95% CI 0.62 to 0.79). Again, the 95% PI was wide, ranging from 0.40 to 0.90. Based on 'eye-balling, we identified three outlying studies. When we excluded two studies with high c-statistics,[24,30] summary estimates were similar to the results of studies in general population settings. After excluding one study with a low c-statistic[31] or all three outlying studies, discrimination remained highest in studies recruiting from hospitals, with a lower point estimate but more narrow 95% prediction intervals (data not shown).

Meta-analysis of stroke risk per score

Figure 2 shows the forest plots with the annual stroke risks and/or systemic thromboembolism for the scores 0 to 3.

For every score on CHA2DS2-VASc, there was substantial heterogeneity in both settings of care with wide approximate 95% prediction intervals. In studies enrolling patients from the general population for example, the stroke risk for a CHA2DS2-VASc score of zero in a new validation study could lie between 0.0% and 0.9%. For score one and two, these were 0.3% and 1.2%, and 0.4 and 3.3%, respectively. See Fig. 2a.

Studies recruiting from a hospital care setting showed a more diverse distribution of risks per score, with higher pooled annual risks for all scores. For instance, the annual risk for a CHA2DS2-VASc score of 1 was 1.4% (approximate 95% PI 0.04 to 6.5%) in these studies, compared to 0.7%

(approximate 95% PI 0.3 to 1.2%) in studies enrolling AF patients from the general population. See Fig. 2b. Although excluding one outlier[31] of the hospital-based studies in a sensitivity analysis resulted in a lower risk for CHA2DS2-VASc score 1 of 1.2%, this was still nearly twice as high as compared to the pooled estimate of studies enrolling patients from the general population. Furthermore, the prediction interval remained wide, ranging from 0.06 to 3.8. The differences between study populations did not sufficiently explain heterogeneity. See Fig. 2c.

Excluding one study that recruited patients all 75 years of age and older[23] in an additional sensitivity analysis did not change the results. Two studies sampled from the same CPRD data source and including the same patients multiple times in our meta-analysis cannot be ruled-out. Excluding either the study by van den Ham[35] or by Allan[40]did not change our results (data not shown).

To further illustrate the interpretation of pooled stroke risks and their uncertainty (due to estimation error and heterogeneity), we calculated the probability that patients with a certain CHA2DS2-VASc score have an annual stroke risk below 1%. For patients recruited from the general population, this probability was 98% (score 0), 91% (score $\leq$ 1) and 19% (score $\leq$ 2). For patients recruited from a hospital care setting, these probabilities dropped to 71%, 39% and, respectively, 17%.

Meta-regression and best available evidence

To further explore sources of heterogeneity in both discrimination and the stroke risk per score, we performed several meta-regression analyses. These demonstrate that it was difficult to identify any relevant sources, as regression coefficients for risk of bias, study characteristics or summarized patient characteristics were not statistically significant. Furthermore, we explored heterogeneity between studies considered as 'best available evidence'. See Table S3 and Fig S3a-d.

Discussion

This systematic review and meta-analysis thoroughly explores heterogeneity in the results of all currently available validation studies of CHA2DS2-VASc. Our analysis confirms that most validation studies of CHA2DS2-VASc yield conflicting results, with highly variable estimates for stroke risk per score. This heterogeneity partly appears to arise from population or case-mix differences across the validation studies, as stratified analyses showed lower stroke risk estimates for studies enrolling patients from the open general population as compared to studies using hospital-based recruitment strategies. Yet, substantial between-study heterogeneity remained and could not be resolved by adjusting for differences in study characteristics, differences in risk of bias, or other differences in population characteristics.

Strengths and limitations

A major strength of this study is that we applied rigorous and state-of-the art systematic review and quantitative synthesis. Previous studies[46,47] found a similar modest discrimination of CHA2DS2-VASc, but did not provide stroke risk for each score or explored potential sources for heterogeneity such as case-mix differences. Another recent publication [48] also found substantial variation in stroke risks, but could not identify any source of this heterogeneity. Our systematic literature review and meta-analysis adds the following additional inferences. First, the study by Quinn does not provide a summarized estimate per CHA2DS2-VASc score, and the observed heterogeneity was summarized in indices of an I^2-value or the Q-statistic. We did provide such summarized estimates per CHA2DS2-VASc-score and report the heterogeneity around these point estimates with 95% prediction intervals that are easily appreciated in clinical practice. Furthermore, we used the CHARMS checklist specifically designed to appraise prediction modelling studies and assess their risk of bias. Lastly, in contrast to previous studies, we additionally calculated a summarized estimate of the c-statistic as a measure of discrimination.

Also, the NICE guidelines[49] provide a formal evaluation of CHA2DS2-VASc, and the score provided in a cost-effective method to estimate stroke risk and indicate anticoagulant treatment. However, stroke risks were based on a single study,[50] no summary estimate was provided and subsequently a measure of uncertainty was lacking.

Nevertheless, for full appreciation of our results several issues should be considered. The CHA2DS2-VASc rule has been advocated as superior to its precursor CHADS2 in identifying AF patients at 'truly low' risk of stroke, in particular for low scores on CHA2DS2-VASc (0 or 1). However, prompted by validation studies of CHA2DS2-VASc showing widely varying results, there has been much debate on what score (in particular 0 or 1) truly defines low risk. This is reflected in our meta-analysis. Indeed, we found that future validation studies where patient enrollment starts in hospital care may observe a very low stroke risk for CHA2DS2-VASc score zero (well below 1%; the threshold above which it is often advocated that the benefits of anticoagulants outweigh the bleeding risk[51,52]). However, patients with a score of one, two or even three may also found to have such a low stroke risk. Conversely, future studies may also find an in fact high stroke risk (e.g. above 3%) already in patients with score zero, which we believe explains the confusion and recent debate on what score denotes a low risk.[8,9,53-55]

We undertook many efforts to explore sources of this large extent of between-study heterogeneity, and several issues require further inspection. First, our results suggest that differences in risk of bias do not play an important role when summarizing estimates of prediction model performance. Although several validation studies showed shortcomings, differences were small and even the most homogeneous group of studies at low risk of bias showed conflicting results in stroke risk per score.

Second, some heterogeneity could be explained by differences in case-mix across the validation studies. We observed a clinically relevant higher stroke risk in studies recruiting AF patients from a hospital care setting compared to those enrolling from the general population. It is possible that AF patients recruited from the hospital care setting represent more diseased patients with a higher baseline risk of stroke, independent of their overall CHA2DS2-VASc score. As an example, the type and burden of AF (paroxysmal, persistent or permanent) may have an association with stroke risk,[56] as may the severity and duration of a patient's individual CHA2DS2-VASc risk factors, or risk factors not included in the CHA2DS2-VASc score such as renal failure.[57] Case-mix differences in such risk factors between patients sampled from hospital care and from the general population could, at least partly, explain the variation in validation studies and the observed difference in stroke risk in our results.

However, third, although we stratified studies to the clinical setting where patient enrollment started, this strategy did not sufficiently help to explain all heterogeneity across the validation studies. Adjustment for differences in study characteristics through meta-regression did not much affect the extent of between-study heterogeneity. For instance, the definition of the outcome under study – only ischemic stroke; ischemic stroke and thromboembolism or indeed even any type of stroke including intra-cranial haemorrhage – could potentially influence the risks for each CHA2DS2-VASc score but including the different outcomes as a covariate in the meta-regression model did not affect the results. Similarly, additional summarized population characteristics such as mean CHA2DS2-VASc score or use of platelet inhibitors did not account for the heterogeneous results. Importantly, we could not evaluate which combination of predictors contributed to a patient's CHA2DS2-VASc score. It is believed that not all predictors in the CHA2DS2-VASc decision rule carry the same stroke risk[58,59] although this is not acknowledged in the rule as almost all its predictors

contribute 1 point. Likewise, females with a score 1 (i.e. no other risk factor) are likely to be at lower risk than males with 1 risk factor. However, the included validation studies often do not report stratified analyses for males and females and thus in this meta-analysis of aggregated data we were unable to account for this. Individual patient data meta-analysis would be needed to fully clarify issues such as stroke risk in females with no additional stroke risk factors.

Fourth, the age categories of the CHA2DS2-VASc are broad and thus a patient aged 65 will receive the same score as a 74 years old person for the 'Age' category, though stroke risk will likely not be equal. This also results in heterogeneity and variation found in the results of validation studies

Fifth, it has previously been shown that ethnicity has an effect on stroke risk[60,61] and stroke mortality[62]. Unfortunately we did not have enough data on ethnicity to consider this variable in our meta-analysis. We did, however, include the geographic region as a covariate in the meta-regression model (data not shown), and in line with the findings by Quinn et al.[48] this could not sufficiently explain the large heterogeneity.

Sixth, we were only able to meta-analyze the stroke risk per score and the c-statistic, as these were the measures most often and consistently reported in current validation studies. Other measures such as decision curve analysis have been proposed to better investigate the clinical value of using a decision rule.[63,64] Unfortunately, these aspects of model performance are rarely consistently quantified. This renders meta-analyzing such measures difficult, if not (yet) impossible, and can be considered a limitation to this current systematic review and meta-analysis of pooled c-statistics and strokes risks per score.

Furthermore, it is important to consider that we only included validation studies where patients did not (yet) use anticoagulants. By itself, the choice of including only patients not on anticoagulation could result in a selected subtype of AF patients as this may have led to confounding 'by contra-indication'[65] when anticoagulation is withheld due to (for instance)

severe illness or bleeding risk. We however deliberately chose not to include populations already on anticoagulants as CHA2DS2-VASc is intended to be used for stroke risk prediction *prior* to anticoagulant treatment decisions in AF patients.

Finally, it should be noted though that, essentially, heterogeneity is not uncommon for studies validating diagnostic or prognostic decision rules, also in the field of thrombotic disorders. Clinical decision rules like CHA2DS2-VASc are popular in daily clinical practice because they are helpful and easy-to-apply methods to tailor subsequent treatment decisions. Indeed, for younger patients without additional stroke risk factors (score of 0) and for more 'high-risk' patients (score of around 2-3 or above), CHA2DS2-VASc clearly facilitates in anticoagulant treatment recommendations. However, based on synthesis of the current evidence in the literature, tailoring treatment for patients at 'low to intermediate risk' (i.e. CHA2DS2-VASc scores 1 to 2) remains ambiguous, as is also reflected by the discussion in literature on the optimal threshold for initiating anticoagulant treatment.[8-12]

Clinical implications and future research questions

In the treatment of atrial fibrillation, adequate identification of different stroke risk groups is pivotal for anticoagulant treatment decisions. The main inference of our meta-analysis is that – albeit a simple, effective and easy-to-use tool at truly low and truly high risk patients – CHA2DS2-VASc may have difficulties in tailoring anticoagulant treatment adequately in AF patients at intermediate risk of stroke (roughly those with a score of 1 or 2). Differences in stroke risks between studies recruiting from hospitals or from the general population indicate that possible case-mix differences between populations should be taken into account in clinical decision-making but further uncertainty remains. Future research should focus on this issue, with further model revision, and considering additional co-morbidity items (e.g. renal impairment), (novel) biomarkers or imaging such as left atrial wall remodeling patterns in addition to existing prediction models for quantifying the thrombotic risk in AF patients.[66-68]. These additional tests may ultimately result in a better clinical decision. Whether this can be achieved should be the focus of further investigation.

Conclusions

Studies validating CHA2DS2-VASc demonstrate high heterogeneity in predicted stroke risks for different scores.

Acknowledgements

S. van Doorn had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. S. van Doorn performed the literature search. S. van Doorn, F. Kaasenbrood, and G.-J. Geersing performed critical appraisal. S. van Doorn and F. Kaasenbrood performed data extraction and risk of bias assessment. T. P. A. Debray and S. van Doorn performed data analysis. All authors interpreted the data. S. van Doorn, T. P. A. Debray, A. W. Hoes, F. H. Rutten, K. G. M. Moons, and G.-J. Geersing wrote and critically reviewed the manuscript.

Disclosures

G.-J. Geersing and F. H. Rutten received an unrestricted institutional grant from Boehringer Ingelheim. K. G. M. Moons received a grant from The Netherlands Organization for Scientific Research (ZONMW 918.10.615 and 91208004). GJG and TPAD are supported by a VENI grant from The Netherlands Organization for Scientific Research (ZONMW 016.166.030 and, respectively, 916.17.050). All funding sources had no role in the design, conduct, analyses, or reporting of the study or in the decision to submit the manuscript for publication. All other authors declared no conflict of interest.

Table 1: Framing the review aim using CHARMS key items.

Item	Comment
1. Type of model	CHA2DS2-VASc rule
2. Intended scope of review	1) review existing studies validating CHA2DS2-VASc for AF patients not (yet) anticoagulated, 2) meta-analyze estimates of c-statistic and stroke risk per score, and 3) explore sources of heterogeneity across the validation studies.
3. Type of modelling studies	External validation studies of CHA2DS2-VASc
4. Target population to whom the model applies	Patients with non-valvular AF not already treated with oral anticoagulants.
5. Outcomes to be predicted	Any of ischemic stroke and/or TIA; all types of stroke; systemic thromboembolism or a combination thereof.
6. Time span	One year risk of the outcome
7. Intended moment of using the model	At the time of diagnosis of AF, and annually when revising anticoagulant treatment indication.

Table 2. Overview of the included studies

Author	Year	Location	Setting	Outcome	N	% on antiplatelet therapy	Months of follow-up	Incidence per 100 person-years
Hobbs [23]	2011	UK	Non-selected	ischemic stroke	665	55.3	26.4	3.6
Olesen [24]	2011	Denmark	Hospital	TE	73538	34.7	12.0	7.6
Sandhu [25]	2011	Canada	Hospital	all stroke incl. ICH	4476	n.r.	12.0	6.2
Abu-Assi [26]	2013	Spain	Non-selected	TE	186	81.6	42.7	2.2
Guo [27]	2013	China	Hospital	TE	885	79.0	22.8	3.7
Singer [28]	2013	USA	Non-selected	TE	10927	n.r.	35.8	2.1
Forslund [29]	2014	Sweden	Non-selected	ischemic stroke	9959	.0	12.0	2.0
Komatsu [30]	2014	Japan	Hospital	TE	332	33.0	53.0	2.1
Siu [31]	2014	China	Hospita	ischemic	3881	.0	38.4	9.3

			l	stroke				
Abumuaileq [32]	2015	Spain	Hospital	TE	154	97.4	11.0	5.8
Saliba [33]	2015	Israel	Non-selected	ischemic stroke	41140	n.r.	11.1	4.5
Suzuki [34]	2015	Japan	Hospital	ischemic stroke	3588	41.8	16.8	1.3
van den Ham [35]	2015	UK	Non-selected	ischemic stroke	60594	n.r.	25.2	3.0
Aspberg [36]	2016	Sweden	Hospital	ischemic stroke	152153	n.r.	26.4	3.3
Chao [37]	2016	Taiwan	Hospital	ischemic stroke	186570	.0	40.8	3.7
Nielsen [38]	2016	Denmark	Hospital	ischemic stroke	198697	n.r.	29.0	3.2
Xing [39]	2016	China	Hospital	TE	413	68.3	23.9	14.3
Allan [40]	2017	UK	Non-selected	ischemic stroke	40139	n.r.	26.4	3.8
McAlister [41]	2017	Canada	Hospital	TE	58451	n.r.	31.0	4.2

TE=thrombo-embolism; ICH=intra-cranial hemorrhage; n.r.=not reported

References

1 Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, Singer DE. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *JAMA* 2001; **285**: 2370–5.

2 Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke* 1991; **22**: 983–8.

3 Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Ann Intern Med* 2007; **146**: 857–67.

4 Lin L, Lim WS, Zhou HJ, Khoo AL, Tan KT, Chew AP, Foo D, Chin JJ, Lim BP. Clinical and Safety Outcomes of Oral Antithrombotics for Stroke Prevention in Atrial Fibrillation: A Systematic Review and Network Meta-analysis. *J Am Med Dir Assoc* 2015; **16**: 1103.e1–19.

5 Lip GYH, Nieuwlaat R, Pisters R, Lane DA, Crijns HJGM. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest* 2010; **137**: 263–72.

6 Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of Clinical Classification Schemes for Predicting Stroke: Results From the National Registry of Atrial Fibrillation. *JAMA* 2001; **285**: 2864–70.

7 European Heart Rhythm Association, European Association for Cardio-Thoracic Surgery, Camm AJ, Kirchhof P, Lip GYH, Schotten U, Savelieva I, Ernst S, Van Gelder IC, Al-Attar N, Hindricks G, Prendergast B, Heidbuchel H, Alfieri O, Angelini A, Atar D, Colonna P, De Caterina R, De Sutter J, Goette A, et al. Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *Eur Heart J* 2010; **31:** 2369–429.

8 Savino JA, Halperin JL. Should Patients With Atrial Fibrillation and 1 Stroke Risk Factor (CHA2DS2-VASc Score 1 in Men, 2 in Women) Be Anticoagulated?Response to Savino and Halperin. *Circulation* 2016; **133**: 1504–11.

9 Lip GYH, Nielsen PB. Should Patients With Atrial Fibrillation and 1 Stroke Risk Factor (CHA2DS2-VASc Score 1 in Men, 2 in Women) Be Anticoagulated?Response to Lip and Nielsen. *Circulation* 2016; **133**: 1498–503.

10 Friberg L, Skeppholm M, Terént A. Benefit of Anticoagulation Unlikely in Patients With Atrial Fibrillation and a CHA2DS2-VASc Score of 1. *J Am Coll Cardiol* 2015; **65**: 225–32.

11 Huisman MV. Patients With Atrial Fibrillation and a CHA2DS2-VASc Score of 1: Are They at Low or High Stroke Risk? *J Am Coll Cardiol* 2015; **65**: 1395–7.

12 Olesen JB, Torp-Pedersen C. Stroke risk in atrial fibrillation: Do we anticoagulate CHADS2 or CHA2DS2-VASc ≥1, or higher? *Thromb Haemost* 2015; **113**: 1165–9.

13 Moons KGM, de Groot JAH, 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 Med* 2014; **11**: e1001744.

14 Steyerberg E. Clinical Prediction Models. 1st ed. Springer Science & Business Media; 2008.

15 Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc* 2001; **8**: 391–7.

16 Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KGM. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS One* 2012; **7**: e32844.

17 Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, Riley RD, Moons KGM. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;

356: i6460.

18 Altman DG, Bland JM. How to obtain the P value from a confidence interval. *BMJ* 2011; **343**: d2304–4.

19 Trikalinos TA, Trow P, Schmid CH. Simulation-Based Comparison of Methods for Meta-Analysis of Proportions and Rates. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013.

20 van Klaveren D, Steyerberg EW, Perel P, Vergouwe Y. Assessing discriminative ability of risk models in clustered data. *BMC Med Res Methodol* 2014; **14**: 5.

21 IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol* 2014; **14**: 1.

22 Snell KIE, Hua H, Debray TPA, Ensor J, Look MP, Moons KGM, Riley RD. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol* 2016; **69**: 40–50.

23 Hobbs FD, Roalfe AK, Lip GYH, Fletcher K, Fitzmaurice DA, Mant J. Performance of stroke risk scores in older people with atrial fibrillation not taking warfarin: comparative cohort study from BAFTA trial. *BMJ* 2011; **342**: d3653.

24 Olesen JB, Lip GYH, Hansen ML, Hansen PR, Tolstrup JS, Lindhardsen J, Selmer C, Ahlehoff O, Olsen A-MS, Gislason GH, Torp-Pedersen C. Validation of risk stratification schemes for predicting stroke and thromboembolism in patients with atrial fibrillation: nationwide cohort study. *BMJ* 2011; **342**: d124.

25 Sandhu RK, Bakal JA, Ezekowitz JA, McAlister FA. Risk stratification schemes, anticoagulation use and outcomes: the risk--treatment paradox in patients with newly diagnosed non-valvular atrial fibrillation. *Heart* 2011; **97**: 2046–50.

26 Abu-Assi E, Otero-Ravina F, Allut Vidal G, Coutado Mendez A, Vaamonde Mosquera L, Sanchez Loureiro M, Caneda Villar MC, Fernandez Villaverde JM, Maestro Saavedra FJ, Gonzalez-Juanatey JR, Grupo Barbanza researchers. Comparison of the reliability and validity of four contemporary risk stratification schemes to predict thromboembolism in non-anticoagulated patients with atrial fibrillation. *Int J Cardiol* 2013; **166**: 205–9.

27 Guo Y, Apostolakis S, Blann AD, Wang H, Zhao X, Zhang Y, Zhang D, Ma J, Wang Y, Lip GYH. Validation of contemporary stroke and bleeding risk stratification scores in non-anticoagulated Chinese patients with atrial fibrillation. *Int J Cardiol* 2013; **168**: 904–9.

28 Singer DE, Chang Y, Borowsky LH, Fang MC, Pomernacki NK, Udaltsova N, Reynolds K, Go AS. A new risk scheme to predict ischemic stroke and other thromboembolism in atrial fibrillation: the ATRIA study stroke risk score. *J Am Heart Assoc* 2013; **2**: e000250–0.

29 1 Forslund T, Wettermark B, Wandell P, Euler von M, Hasselstrom J, Hjemdahl P. Risks for stroke and bleeding with warfarin or aspirin treatment in patients with atrial fibrillation at different CHA(2)DS(2)VASc scores: experience from the Stockholm region. *Eur J Clin Pharmacol* 2014; **70**: 1477–85.

30 Komatsu T, Sato Y, Ozawa M, Kunugita F, Yoshizawa R, Morino Y, Nakamura M. Comparison between CHADS2 and CHA2DS2-VASc score for risk stratification of ischemic stroke in Japanese patients with non-valvular paroxysmal atrial fibrillation not receiving anticoagulant therapy. *Int Heart J* 2014; **55**: 119–25.

31 Siu C-W, Lip GYH, Lam K-F, Tse H-F. Risk of stroke and intracranial hemorrhage in 9727 Chinese with atrial fibrillation in Hong Kong. *Heart Rhythm* 2014; **11**: 1401–8.

32 Abumuaileq RRY, Abu-Assi E, López-López A, Raposeiras-Roubin S, Rodríguez-Mañero M, Martínez-Sande L, García-Seara J, Fernandez-López XA, Peña-Gil C, González-Juanatey JR. Comparison between CHA2DS2-VASc and the new R2CHADS2 and ATRIA scores at predicting thromboembolic event in non-anticoagulated and anticoagulated patients with non-

valvular atrial fibrillation. *BMC Cardiovasc Disord* 2015; **15**: 156.

33 Saliba W, Barnett-Griness O, Elias M, Rennert G. The Association Between Red Cell Distribution Width and Stroke in Patients with Atrial Fibrillation. *Am J Med* 2015; **128**: 192.e11–8.

34 Suzuki S, Yamashita T, Okumura K, Atarashi H, Akao M, Ogawa H, Inoue H. Incidence of Ischemic Stroke in Japanese Patients With Atrial Fibrillation Not Receiving Anticoagulation Therapy. *Circ J* 2015; **79**: 432–8.

35 van den Ham HA, Klungel OH, Singer DE, Leufkens HGM, van Staa TP. Comparative Performance of ATRIA, CHADS2, and CHA2DS2-VASc Risk Scores Predicting Stroke in Patients With Atrial Fibrillation: Results From a National Primary Care Database. *J Am Coll Cardiol* 2015; **66**: 1851–9.

36 Aspberg S, Chang Y, Atterman A, Bottai M, Go AS, Singer DE. Comparison of the ATRIA, CHADS2, and CHA2DS2-VASc stroke risk scores in predicting ischaemic stroke in a large Swedish cohort of patients with atrial fibrillation. *Eur Heart J* 2016; **37**: 3203–10.

37 Chao T-F, Liu C-J, Tuan T-C, Chen S-J, Wang K-L, Lin Y-J, Chang S-L, Lo L-W, Hu Y-F, Chen T-J, Chiang C-E, Chen S-A. Comparisons of CHADS2 and CHA2DS2-VASc scores for stroke risk stratification in atrial fibrillation: Which scoring system should be used for Asians? *Heart Rhythm* 2016; **13**: 46–53.

38 Nielsen PB, Larsen TB, Skjøth F, Overvad TF, Lip GYH. Stroke and thromboembolic event rates in atrial fibrillation according to different guideline treatment thresholds: A nationwide cohort study. *Sci Rep* 2016; **6**: 27410.

39 Xing Y, Ma Q, Ma X, Wang C, Zhang D, Sun Y. CHADS2 score has a better predictive value than CHA2DS2-VASc score in elderly patients with atrial fibrillation. *Clin Interv Aging* 2016; **11**: 941–6.

40 Allan V, Banerjee A, Shah AD, Patel R, Denaxas S, Casas J-P, Hemingway H. Net clinical benefit of warfarin in individuals with atrial fibrillation across stroke risk and across primary and secondary care. *Heart* 2017; **103**: 210–8.

41 McAlister FA, Wiebe N, Jun M, Sandhu R, James MT, McMurtry MS, Hemmelgarn BR, Tonelli M. Are Existing Risk Scores for Nonvalvular Atrial Fibrillation Useful for Prediction or Risk Adjustment in Patients With Chronic Kidney Disease? *Can J Cardiol* 2017; **33**: 243–52.

42 Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005; **58**: 475–83.

43 Gorelick MH. Bias arising from missing data in predictive models. *J Clin Epidemiol* 2006; **59**: 1115–23.

44 Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Med Res Methodol* 2010; **10**: 7.

45 Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006; **59**: 1087–91.

46 Chen J-Y, Zhang A-D, Lu H-Y, Guo J, Wang F-F, Li Z-C. CHADS2 versus CHA2DS2-VASc score in assessing the stroke and thromboembolism risk stratification in patients with atrial fibrillation: a systematic review and meta-analysis. *J Geriatr Cardiol* 2013; **10**: 258–66.

47 Joundi RA, Cipriano LE, Sposato LA, Saposnik G, Group OBOTSORW. Ischemic Stroke Risk in Patients With Atrial Fibrillation and CHA2DS2-VASc Score of 1 Systematic Review and Meta-Analysis. *Stroke* 2016; **47**: 1364–7.

48 Quinn GR, Severdija ON, Chang Y, Singer DE. Wide Variation in Reported Rates of Stroke Across Cohorts of Patients With Atrial Fibrillation. *Circulation* 2017; **135**: 208–19.

49 NICE. Atrial fibrillation: the management of atrial fibrillation. (Clinical guideline 180). NICE; 2014.

50 Friberg L, Rosenqvist M, Lip GYH. Evaluation of risk stratification schemes for ischaemic stroke and bleeding in 182 678 patients with atrial fibrillation: the Swedish Atrial Fibrillation cohort study. *Eur Heart J* 2012; **33**: 1500–10.

51 Eckman MH, Singer DE, Rosand J, Greenberg SM. Moving the tipping point: the decision to anticoagulate patients with atrial fibrillation. *Circ Cardiovasc Qual Outcomes* 2011; **4**: 14–21.

52 Singer DE, Chang Y, Fang MC, Borowsky LH, Pomernacki NK, Udaltsova N, Go AS. The net clinical benefit of warfarin anticoagulation in atrial fibrillation. *Ann Intern Med* 2009; **151**: 297–305.

53 Cairns JA, Healey JS, Macle L, Mitchell LB, Verma A. The New Canadian Cardiovascular Society Algorithm for Antithrombotic Therapy of Atrial Fibrillation Is Appropriately Based on Current Epidemiologic Data. *Can J Cardiol* 2015; **31**: 20–3.

54 Lip GYH, Nielsen PB, Skjøth F, Rasmussen LH, Larsen TB. Atrial Fibrillation Patients Categorized as "Not for Anticoagulation" According to the 2014 Canadian Cardiovascular Society Algorithm Are Not 'Low Risk'. *Can J Cardiol* 2015; **31**: 24–8.

55 Nielsen PB, Larsen TB, Skjøth F, Gorst-Rasmussen A, Rasmussen LH, Lip GYH. Restarting Anticoagulant Treatment After Intracranial Hemorrhage in Patients With Atrial Fibrillation and the Impact on Recurrent Stroke, Mortality, and Bleeding A Nationwide Cohort Study. *Circulation* 2015; **132**: 517–25.

56 Steinberg BA, Hellkamp AS, Lokhnygina Y, Patel MR, Breithardt G, Hankey GJ, Becker RC, Singer DE, Halperin JL, Hacke W, Nessel CC, Berkowitz SD, Mahaffey KW, Fox KAA, Califf RM, Piccini JP, ROCKET AF Steering Committee and Investigators. Higher risk of death and stroke in patients with persistent vs. paroxysmal atrial fibrillation: results from the ROCKET-AF

Trial. *Eur Heart J* 2015; **36**: 288–96.

57 Olesen JB, Lip GYH, Kamper A-L, Hommel K, Kober L, Lane DA, Lindhardsen J, Gislason GH, Torp-Pedersen C. Stroke and bleeding in atrial fibrillation with chronic kidney disease. *N Engl J Med* 2012; **367**: 625–35.

58 Chao T-F, Liu C-J, Wang K-L, Lin Y-J, Chang S-L, Lo L-W, Hu Y-F, Tuan T-C, Chen T-J, Lip GYH, Chen S-A. Should Atrial Fibrillation Patients With 1 Additional Risk Factor of the CHA2DS2-VASc Score (Beyond Sex) Receive Oral Anticoagulation? *J Am Coll Cardiol* 2015; **65**: 635–42.

59 Huang D, ANGUO L, YUE W-S, YIN L, Tse H-F, Siu C-W. Refinement of ischemic stroke risk in patients with atrial fibrillation and CHA2 DS2 -VASc score of 1. *Pacing Clin Electrophysiol* 2014; **37**: 1442–7.

60 Chang K-C, Wang Y-C, Ko P-Y, Wu H-P, Chen Y-W, Muo C-H, Sung F-C, Li T-C, Hsu CY. Increased Risk of First-Ever Stroke in Younger Patients With Atrial Fibrillation Not Recommended for Antithrombotic Therapy by Current Guidelines: A Population-Based Study in an East Asian Cohort of 22 Million People. *Mayo Clinic Proceedings* 2014; **89**: 1487–97.

61 Hajat C, Tilling K, Stewart JA, Lemic-Stojcevic N, Wolfe CDA. Ethnic differences in risk factors for ischemic stroke: a European case-control study. *Stroke* 2004; **35**: 1562–7.

62 Ayala C, Greenlund KJ, Croft JB, Keenan NL, Donehoo RS, Giles WH, Kittner SJ, Marks JS. Racial/ethnic disparities in mortality by stroke subtype in the United States, 1995-1998. *Am J Epidemiol* 2001; **154**: 1057–63.

63 Vickers AJ. Decision analysis for the evaluation of diagnostic tests, prediction models, and molecular markers. *Am Stat* 2008; **62**: 314-20.

64 Steyerberg EW, Pencina MJ, Lingsma HF, Kattan MW, Vickers AJ, Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. *Eur J Clin*

Invest 2012; **42**: 216–28.

65 Walker AM. Confounding by indication. *Epidemiology* 1996; **7**: 335–6.

66 Hijazi Z, Oldgren J, Siegbahn A, Granger CB, Wallentin L. Biomarkers in atrial fibrillation: a clinical review. *Eur Heart J* 2013; **34**: 1475–80.

67 Wu N, Chen X, Cai T, Wu L, Xiang Y, Zhang M, Li Y, Song Z, Zhong L. Association of Inflammatory and Hemostatic Markers With Stroke and Thromboembolic Events in Atrial Fibrillation: A Systematic Review and Meta-analysis. *Can J Cardiol* 2015; **31**: 278–86.

68 Providencia R, Botelho A, Trigo J, Quintal N, Nascimento J, Mota P, Leitão-Marques A. Possible refinement of clinical thromboembolism assessment in patients with atrial fibrillation using echocardiographic parameters. *EP Europace* 2012; **14**: 36–45.

Figs. Legends

Fig. 1: *Process of study selection*

Fig. 2a: *Stroke risk (events per 100 person-years) per score in studies recruiting from the general population*

Solid bars represent 95% confidence intervals

Grey bars represent 95% approximate prediction intervals

Summary estimate is stroke risk as events per 100 per years (95% C.I.) [95% P.I.]

Fig. 2b*: Stroke risk (events per 100 person-years) per score in studies recruiting from hospitals*

Solid bars represent 95% confidence intervals

Grey bars represent 95% approximate prediction intervals

Summary estimate is stroke risk as events per 100 per years (95% C.I.) [95% P.I.]

Fig. 2c: *Stroke risk (events per 100 person-years) per score in studies recruiting from hospitals, excluding relative outliers*

Solid bars represent 95% confidence intervals

Grey bars represent 95% approximate prediction intervals

Summary estimate is stroke risk as events per 100 per years (95% C.I.) [95% P.I.]

Figure 1
Process of study selection

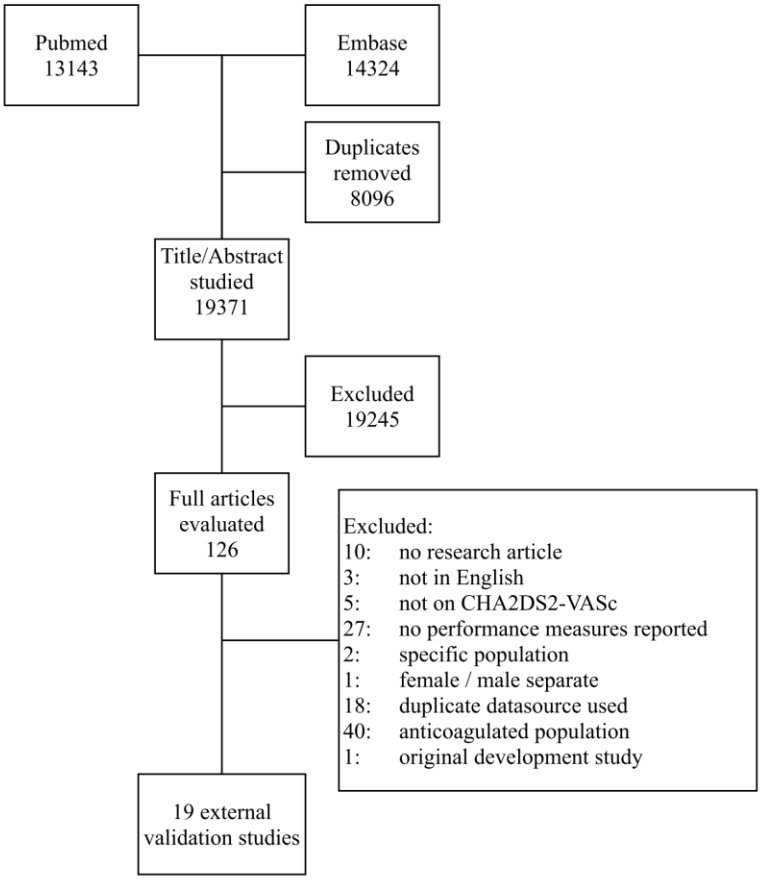
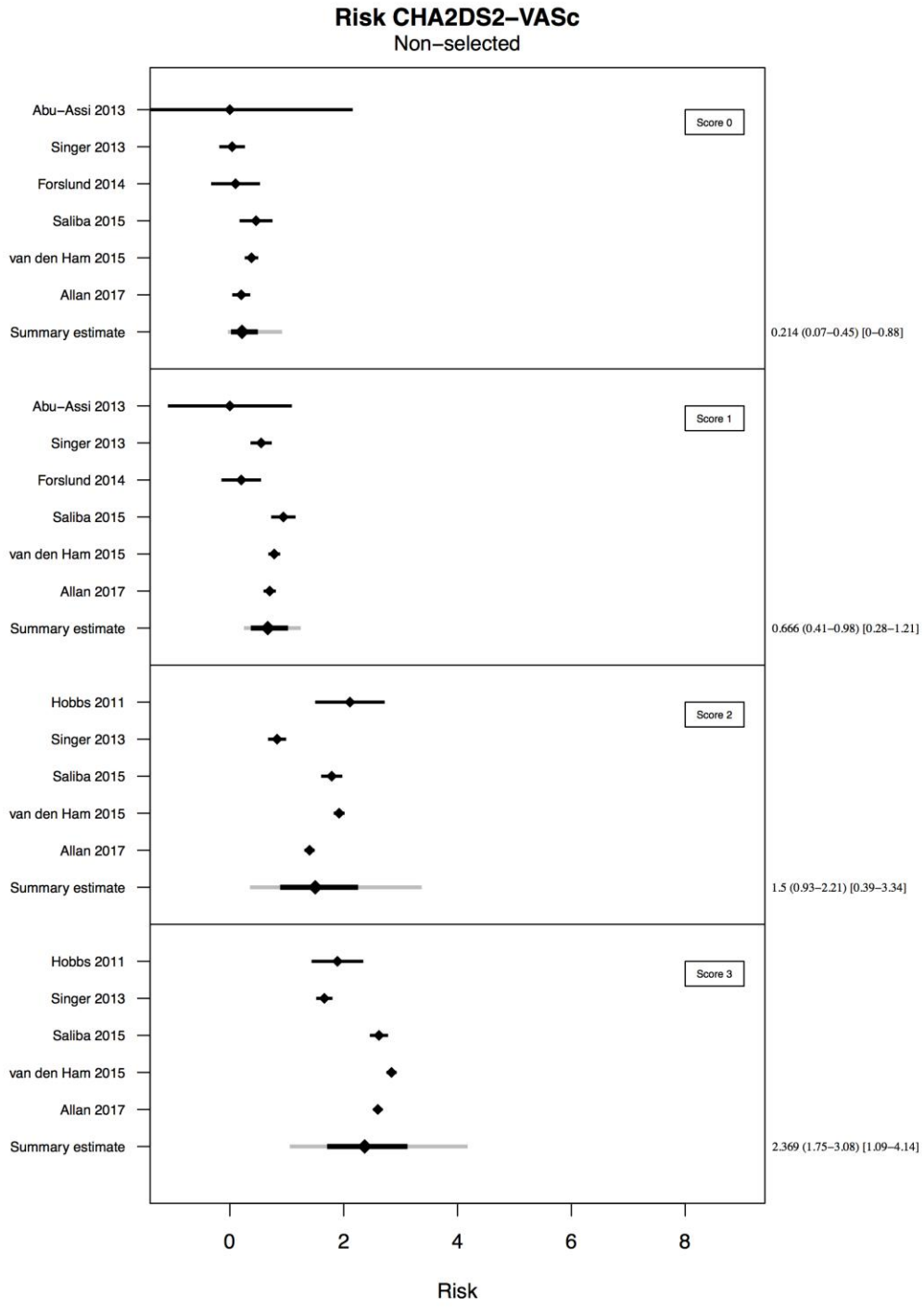
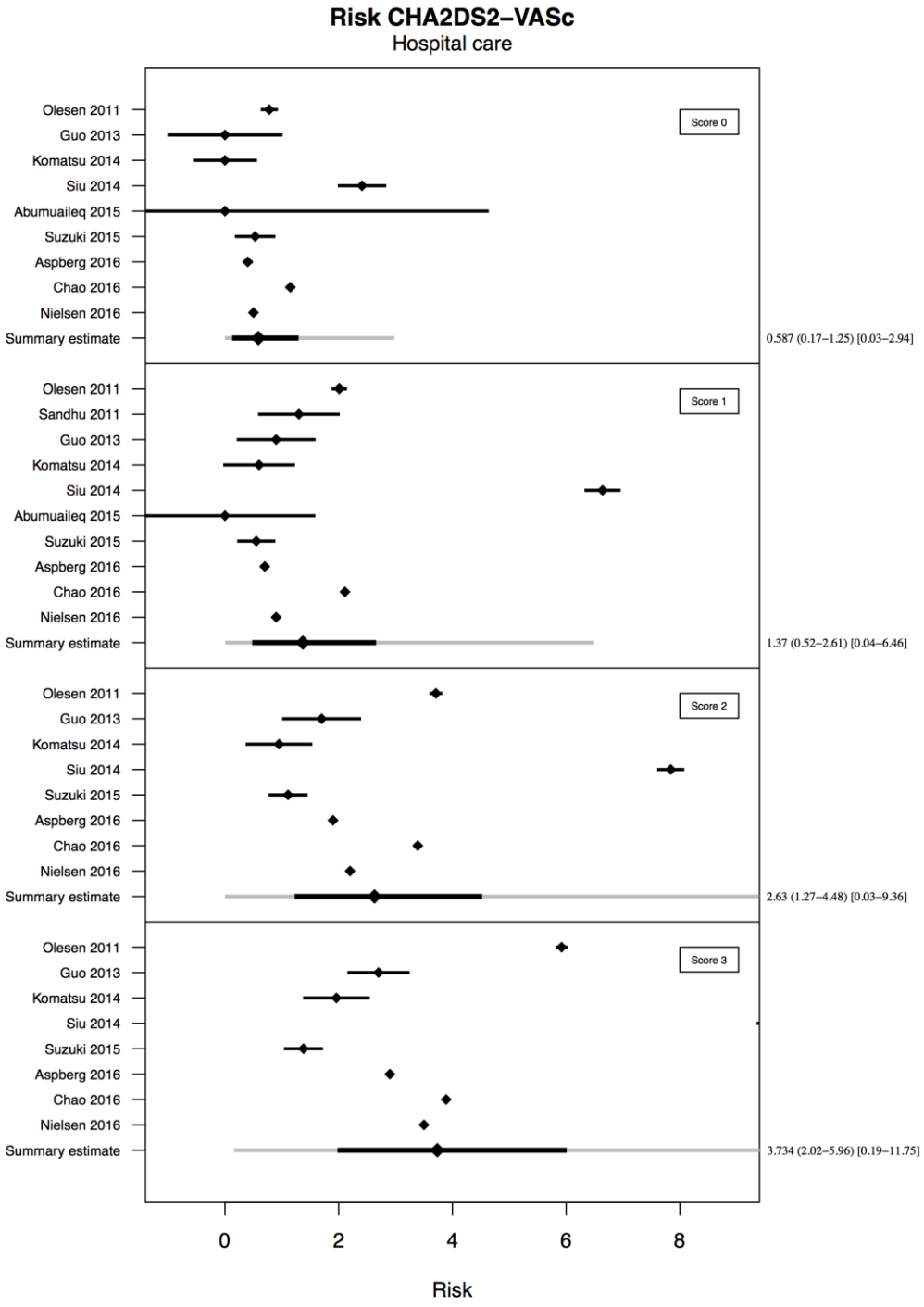


Figure 2a
Stroke risk (events per 100 person-years) per score in studies recruiting from the general population



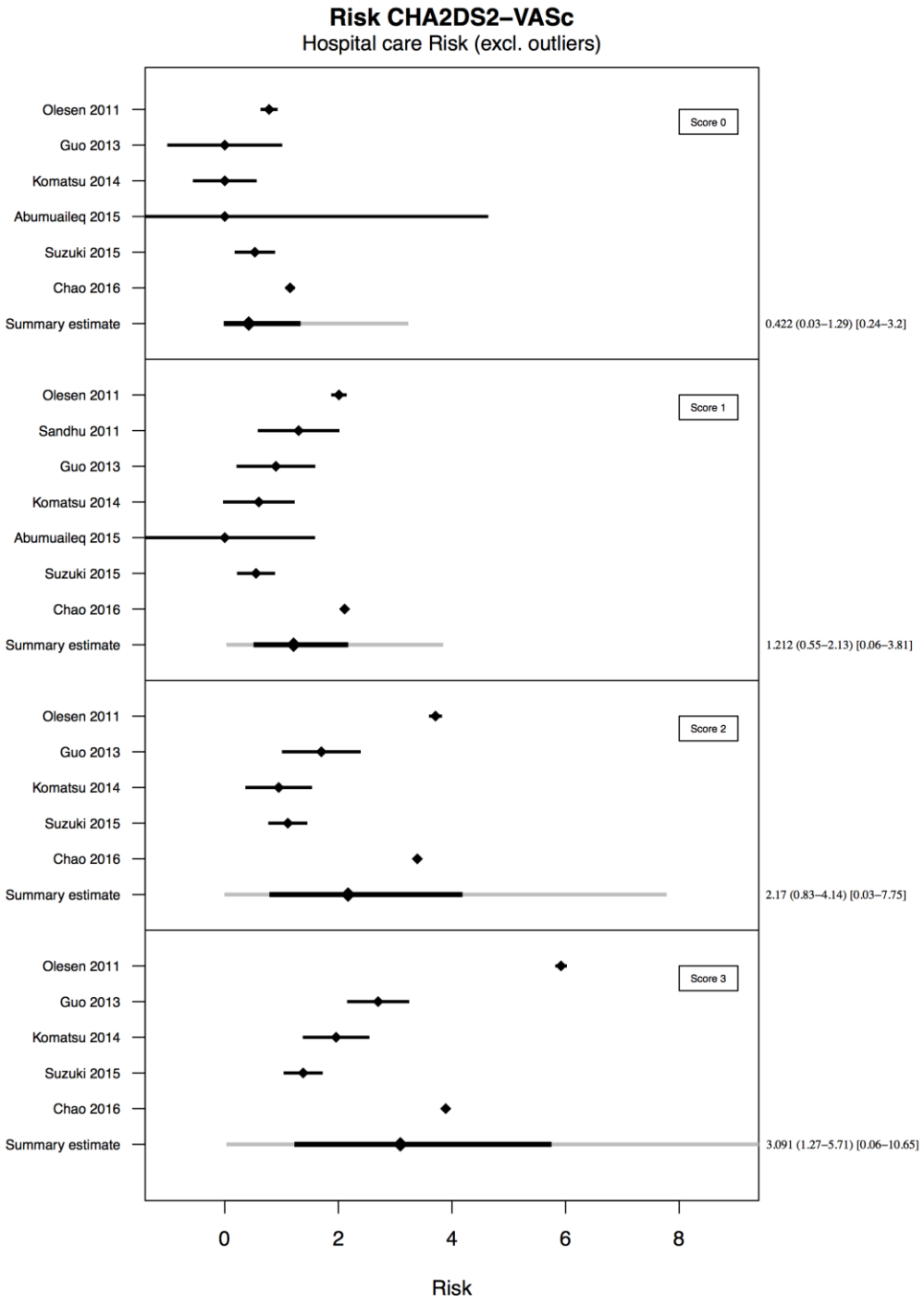
Solid bars represent 95% confidence intervals
Dashed bars represent 95% confidence intervals, estimated
Grey bars represent 95% prediction intervals
Summary estimate is c-statistic (95% C.I.) [95% P.I.]

Figure 2b
Stroke risk (events per 100 person-years) per score in studies recruiting from hospitals



Solid bars represent 95% confidence intervals
Dashed bars represent 95% confidence intervals, estimated
Grey bars represent 95% prediction intervals
Summary estimate is c-statistic (95% C.I.) [95% P.I.]

Figure 2c
Stroke risk (events per 100 person-years) per score in studies recruiting from hospitals, excluding outliers



Solid bars represent 95% confidence intervals
Dashed bars represent 95% confidence intervals, estimated
Grey bars represent 95% prediction intervals
Summary estimate is c-statistic (95% C.I.) [95% P.I.]