Validation of the CHADS2 clinical prediction rule to predict ischaemic stroke. A systematic review and meta-analysis.

AUTHOR(S)

Claire Keogh, Emma Wallace, Ciara Dillon, Borislav D. Dimitrov, Tom Fahey

Validation of the $CHADS_2$ clinical prediction rule to predict ischaemic stroke: A systematic review and meta-analysis[1]

Claire Keogh[1] (PhD)*, Emma Wallace[1] (MB), Ciara Dillon[1] (MB), Borislav D Dimitrov[1] (DM/PhD), Tom Fahey[1] (MD)

[1] Royal College of Surgeons in Ireland

***Corresponding author:** Dr Claire Keogh

Correspondence and reprints to: HRB Centre for Primary Care Research, Department of General Practice, Royal College of Surgeons in Ireland, 123 St. Stephens Green, Dublin 2, Republic of Ireland.

Fax: +353 1 402 2764

Tel: +353 1 402 2317

Email: clairekeogh@rcsi.ie

Running title: Validation of the $CHADS_2$ score for ischaemic stroke

Nomenclature and abbreviations: CPR = clinical prediction rules; CI = confidence intervals; RR = risk ratios

Words: 5118 (excluding tables, figure legends and references)

[1] This work was supported by the Health Research Board (HRB) of Ireland through the HRB Centre for Primary Care Research under Grant HRC/2007/1.

Abstract

Background: The $CHADS_2$ predicts annual risk of ischaemic stroke in non-valvular atrial fibrillation. This systematic review and meta-analysis aims to determine the predictive value of $CHADS_2$.

Methods: The literature was systematically searched from 2001 to October 2010. Data was pooled and analysed using discrimination and calibration statistical measures, using a random effects model.

Results: Eight data sets (n=2815) were included. The diagnostic accuracy suggested a cut-point of ≥1 has higher sensitivity (92%) than specificity (12%) and a cut-point of ≥4 has higher specificity (96%) than sensitivity (33%). Lower summary estimates were observed for cut-points ≥2 (sensitivity 79%, specificity 42%) and ≥3 (specificity 77%, sensitivity 50%). There was insufficient data to analyse cut-points ≥5 or ≥6. Moderate pooled *c* statistic values were identified for the classic (0.63, 95% CI 0.52-0.75) and revised (0.60, 95% CI 0.43-0.72) view of stratification of the $CHADS_2$. Calibration analysis indicated no significant difference between the predicted and observed strokes across the three risk strata for the classic or revised view. All results were associated with high heterogeneity and conclusions should be made cautiously.

Conclusions: The pooled *c* statistic and calibration analysis suggests minimal clinical utility of both the classic and revised view of the $CHADS_2$ in predicting ischaemic stroke across all risk strata. Due to high heterogeneity across studies and low event rates across all risk strata, the results should be interpreted cautiously. Further validation of $CHADS_2$ should perhaps be undertaken, given the methodological differences between many of the available validation studies and the original $CHADS_2$ derivation study.

Keywords: atrial fibrillation, cerebral infarct, risk factors, risk prediction, $CHADS_2$

What is known about this topic

- The $CHADS_2$ clinical prediction rule (CPR) is used to predict the annual stroke risk for patients with non-valvular atrial fibrillation.
- Patients can be stratified according to their stroke risk into low, moderate or high risk groups, depending on how many risk factors they have. However, no consensus exists as to where the cut-off points for each strata should be.
- Although the $CHADS_2$ score has previously been validated, there exist methodological differences between many of these validation studies and the original $CHADS_2$ derivation study.

What this paper adds

- This systematic review and meta-analysis validates the $CHADS_2$ score, whilst accounting for the methodological differences between the derivation and validation studies. Specifically, the current study controls for stroke type and adjusts for the net clinical benefit of treatment.
- This paper compares the predictive ability of two different stratification classification methods, the classic view versus the revised view, using both discrimination and calibration statistical methods.
- Overall, the pooled data suggests only reasonable utility of the classic and revised view of the $CHADS_2$ score in predicting ischaemic stroke. However, caution should be applied when interpreting the results due to high heterogeneity across studies and low event rates across all risk strata.
- Further validation of the $CHADS_2$ score is necessary, given the methodological differences between many of the derivation and validation studies.

Introduction

Non-valvular atrial fibrillation (NVAF) is the most common cardiac arrhythmia. Its prevalence is age-dependent rising from 1% in patients under 60 years to 15% in those over 85 years. (1) The presence of NVAF increases the risk of stroke more than four-fold and approximately one third of all strokes are associated with atrial fibrillation. (2, 3) The $CHADS_2$ score is commonly used by clinicians to predict the annual stroke risk in patients with NVAF. The score was derived by expert consensus and consists of risk factors that were found to increase the risk of ischaemic stroke in patients with NVAF in the Atrial Fibrillation Investigators (AFI) and Stroke Prevention in Atrial Fibrillation (SPAF) trials. (4) It was tested in a non-warfarin population and the risk scores were adjusted for the impact of aspirin therapy. It consists of six clinical features and assigns one point for each of congestive heart failure, hypertension, age $\geq$ 75, and diabetes mellitus, and two points for prior history of stroke or transient ischaemic attack (TIA). (4) The original CPR forms a cumulative risk score for each patient based on the number of risk factors present. Total scores range from 0 to 6 and annual stroke rate for each of the seven categories range from 1.9 (score 0) to 18.2 (score 6) (rates adjusted for treatment effect). The CPR can be used to stratify patients according to low, moderate and high risk of stroke according to either a classic or a revised view of stratification (see Table 1 for summary). The classic view defines low risk by a $CHADS_2$ score of 0, moderate risk by a score of 1-2 and high risk as a score of 3-6. The revised view defines low risk by a $CHADS_2$ score of 0, moderate risk by a score of 1 and high risk as a score of 2-6.

A number of studies validating the $CHADS_2$ score have been published. However, there are methodological differences between many of these validation studies and the original study in which the CPR was derived and validated. Firstly, many studies have not restricted the outcome to ischaemic stroke and include non-central nervous system emboli and haemorrhagic stroke, resulting in an overestimation of the ischaemic stroke rate. (5) Secondly, a number of the validation studies are criticised due to being conducted in treated populations. (6) As such, there is limited validation of the original $CHADS_2$ for non-treated populations.

This systematic review aims to determine the accuracy of the $CHADS_2$ in predicting ischaemic stroke in patients with NVAF across three risk strata (low, moderate, and high) according to both the classic and revised view of stratification. The study focuses on one stroke type, ischaemic stroke, ICD 9 CM code:434, and excludes all other types of stroke such as transient ischaemic attack, haemorrhagic stroke and non-central nervous system emboli. Due to the unethical nature of withholding treatment from patients, there exists limited placebo data. As such, each of the studies included in the current work contain patients receiving oral anticoagulation or antiplatelet therapy, with the observed numbers of strokes in each study risk-adjusted for the corresponding treatment benefits.

Methods

Search strategy

A search string was developed to search the Medline database using the PubMed search engine to identify NVAF ('atrial fibrillation'), stroke ('cerebrovascular accident', 'cerebral infarction', 'cva', 'venous thromboembolism') and the $CHADS_2$ score ('CHADS2', 'CHADS 2') and type of study ('prognosis', 'indicators', 'risks'). The search string was restricted to humans and certain publication types were removed (editorial, letter, case reports, comments, dictionary or news). No restrictions were placed on language. The search period ranged from January 2001 to October 2010 (as the $CHADS_2$ was published in 2001). The Cochrane Library, EMBASE, Cinahl and MEDION databases were searched in a similar manner. The references of each relevant article were searched. Where necessary, authors of published studies were contacted to request additional data.

Study selection

Inclusion criteria for the systematic review were: (1) Patient population: adult patients with a diagnosis of NVAF (including chronic, paroxysmal, persistent, permanent and new onset); (2) Outcome measure: ischaemic stroke (ICD-9-CM code 434); (3) Study design: prospective or retrospective cohort; (4) Treatment: either warfarin (adjusted dose) or aspirin prescribed alone; (5) Explanatory variables: $CHADS_2$ score and; (6) Setting of care: primary care, hospital, other specialist settings. Exclusion criteria included: (1) studies that included patients with rheumatic/valvular heart disease and other patients in specific populations

(e.g. patients undergoing ablation procedures); (2) studies that could not separate all ischaemic strokes from other forms of stroke/emboli including transient ischaemic attack, haemorrhagic stroke and non-cerebral emboli; (3) studies where the end point was focused on mortality; (4) studies that prescribed treatment other than warfarin or aspirin or that prescribed combinations of these treatments.

Data extraction

The titles and abstracts for each article retrieved by the electronic search were independently screened by two researchers (EW and CK). The full text article was retrieved for any study that was considered potentially relevant. Each full text article was independently read and considered for inclusion by two researchers (CK and CD). Disagreements were resolved by a third reviewer (EW). Additional data was requested from the authors where necessary. For each study, data was extracted for each of the inclusion criteria variables.

Quality assessment

Quality assessment was independently performed by two researchers (CK and CD) following the methodological standards reported by McGinn for validation studies of CPRs. (7) This quality checklist comprises of five questions that assess the internal and external validity of studies.

Statistical methods

Diagnostic accuracy of the $CHADS_2$ score

The diagnostic accuracy of the $CHADS_2$ at different cut-points was assessed. Data was extracted and 2x2 tables were constructed for each cut-point. For example, the cut-point of ≥1 was constructed by comparing the data below the cut-point (score 0-1) with all data above the cut point (score 2-6). Results are presented as summary sensitivities and specificities and corresponding 95% confidence intervals, calculated using a random effect bivariate model (Stata package metandi). This method accounts for variation in study size and heterogeneity beyond chance as a result of clinical and methodological differences between the studies. (8) Pooled estimates can only be calculated using this model with four or more studies.

The individual and summary estimates of sensitivity and specificity for each cut-point was plotted in a summary receiver operating characteristics (sROC) plot, with the associated sensitivity (true positive) on the y axis against 1-specificity (false negative) on the x axis. Results of interest are the 95% confidence region (which illustrates the precision with which pooled values are estimated) and the 95% prediction region (which illustrates the amount of between study variation) were plotted around the pooled estimates. Heterogeneity was assessed visually (using the sROC plots) and statistically (using the variance of logit transformed sensitivity and specificity), where smaller values indicate less heterogeneity across studies.

As all of the data included in the current study consists of patients treated with either warfarin or aspirin, the observed stroke rates were adjusted for the net benefit of protection against stroke by each treatment. Warfarin is associated with an overall risk reduction of 68% (95% CI, 50 to 79), (9) while aspirin is associated with an overall risk reduction of 21% (95% CI, 0 to 38). (10) Therefore, the observed number of strokes in each included study in the current work was increased by 68% or 21%, depending on the treatment received by each study population.

Predictive value of the $CHADS_2$ score

The predictive value of both the classic and revised view of stratification of the $CHADS_2$ score was measured using the *c* statistic, to determine if either method of classification performed significantly better than chance. A *c* statistic score of 0.5 or above indicates that the classification method performs significantly better than chance. The *c* statistic for both the classic and revised view of the $CHADS_2$ score was calculated for each individual data set to allow for direct comparison between the two views of stratification. The individual *c* statistics were then pooled for each of the classic and revised view of stratification. The observed stroke rates were adjusted for the net benefit of protection offered by warfarin or aspirin treatment, as described above.

Calibration of the $CHADS_2$ score

The data presented in the original $CHADS_2$ study was used as the predictive model. (4) The data observed in each of the validation data sets was compared against this. The number of strokes predicted by the $CHADS_2$ score and those observed within each data set were derived across three risk strata – low, moderate and high according to both the classic and revised view of stratification of the $CHADS_2$ score. To calculate the predicted stroke risk for each of the validation data sets, we applied the adjusted stroke rate from the original $CHADS_2$ study for the classic view (low: 1.9; moderate: 3.4; high 11.3) and the revised view (low: 1.9; moderate: 2.8; high: 9.8) (see Table 1). The observed stroke rates were adjusted for the net benefit of protection offered by warfarin or aspirin treatment, as described above.

Results are presented as risk ratios (RR) with 95% confidence intervals for each of the three risk strata for both the classic and revised view. A RR <1 indicates that the $CHADS_2$ score under-predicts the risk of stroke (i.e. the predicted number of strokes is less than the observed number of strokes). A RR >1 indicates that the $CHADS_2$ score over-predicts the risk of stroke (i.e. the predicted number of strokes is greater than the observed number of strokes). A RR$=1$ indicates a perfect calibration between the observed and predicted number of strokes. Review Manager 5 software from the Cochrane collaboration was used to perform the statistical analyses. Risk ratios were calculated using the Mantel-Haenzel statistical method. A random effects analysis was performed and heterogeneity was described by the I^2 statistic, where lower values indicate less heterogeneity.

Results

Overview of included studies

An overview of the search strategy is presented in Figure 1. The initial search retrieved 3145 articles, of which 3060 were excluded on the basis of their title or abstract. Of the remaining 85 studies, 60 were excluded as not relevant. Of the remaining 25 articles, one study met all the inclusion criteria. The remaining articles were identified as potentially having relevant data. For example, the published study may have combined both ischaemic and haemorrhagic stroke data but met all other inclusion criteria. In these cases, the authors of the remaining articles were contacted. Relevant data was received from four authors. The authors from the remaining studies either (1) responded to say that they could not provide us with the relevant data or (2) did not respond. The five studies resulted in a total of eight sets of data (i.e. three studies provided information for two individual treatment groups).

Study description

Table 2 summarises the characteristics of the included studies. All five publications were in English. The data from only one study had been previously published. (11) Additional data was provided by four authors because the data published in the original articles contained both ischaemic and non-ischaemic stroke types. (12-15) Two studies collected data prospectively, (11, 15) and three retrospectively. (12-14) Data was obtained from different settings including primary care, (13) outpatient (11) and specialist settings,(15) as well as established databases. (12, 14) Different types of NVAF were included, including chronic,(12, 13) paroxysmal, (11, 14) persistent, (14) permanent (14, 15) and new onset. (11, 14) Annual

stroke data was available for four data sets (i.e. two studies), (13, 14) with the remaining four data sets (i.e. three studies) (11, 12, 15) adjusted for follow-up to derive annual stroke data. (16) A total of 2815 participants are included in the analysis. Of these, 2558 participants were receiving warfarin and 257 were receiving aspirin. A detailed description of the number of patients and number of ischaemic strokes observed in each study for each of the three risk strata for both the classic and the revised view is presented in Table 3. The low risk category ($CHADS_2$ score 0 for both the classic and revised view) has a low number of events relative to the moderate and high risk categories. When the data is classified according to the classic view, the majority of patients are classified as moderate risk (low 20.39%; moderate 60.46%; high 19.15%). However, when the data is classified according to the revised view, the majority of patients are classified as high risk (low 21.54%; moderate 33.21%; high 45.25%).

Study quality

The methodological quality of the studies is presented in Figure 2. The external validity was mixed. Although the studies generally represented a wide spectrum of the disease, the patients were generally not selected in an unbiased fashion. Similarly, mixed results were reported in terms of internal validity. The follow-up of patients was adequate. However, issues relating to blinding were generally unreported.

Diagnostic test accuracy of the $CHADS_2$ score

The sensitivity, specificity and associated transformed variance of logit are presented in Table 4. Cut-points are assessed from $CHADS_2 \geq 1$ through $CHADS_2 \geq 4$. There was insufficient data to complete this analysis for scores ≥5 or ≥6. The results indicate that an increase in $CHADS_2$ score is associated with an increased specificity and decreased sensitivity. A cut-point of ≥1 is associated with a higher sensitivity (92%) than specificity (12%), suggesting that this is useful for ruling out the likelihood of stroke when a $CHADS_2$ score of ≥1 is absent (i.e. a $CHADS_2$ score of 0). A cut-point of ≥4 is associated with a higher specificity (96%) than sensitivity (33%) suggesting that a $CHADS_2$ score of 4 or more is useful for ruling in the likelihood of stroke. Lower summary estimates were observed for a cut-point of ≥2 and ≥3. A cut-point of ≥2 (i.e. a $CHADS_2$ score of 0 or 1) is associated with higher sensitivity (79%) than specificity (42%), suggesting that this is better for ruling out stroke when a score of ≥2 is absent (i.e. a $CHADS_2$ score of 0 or 1), while a cut-point of ≥3 is associated with higher specificity (77%) than sensitivity (50%) suggesting that a $CHADS_2$ score of 4 or more is better at ruling in a stroke. However, there is wide heterogeneity associated with each of the cut-points as indicated by the visual inspection of the confidence and prediction ellipse around the mean (see on-line version) and statistically by the variance of logit transformed sensitivity and specificity (Table 4).

Predictive value of the $CHADS_2$ score

The predictive value of the $CHADS_2$ score for both the classic and revised view of classification for each individual study is presented in Table 5. The *c* statistic indicates wide variability across studies in the predictive value of both the classic and revised views of

stratification. For the classic view, the *c* statistic ranges from poor predictability (0.44, 95% CI 0.17-0.70) to good predictability and performing better than chance (0.79, 95% CI 0.76-0.82). A similar pattern emerged for the revised view, with the *c* statistic ranging from 0.43 (0.22-0.65) to 0.75 (0.73-0.78). For two studies, the classic view resulted in a higher *c* statistic than the revised view and for three studies the opposite pattern was true. However, in general similar *c* statistic values were reported for both the classic and revised view within each individual study, with the majority of values indicating that the CPR performed at or around the level of chance.

The *c* statistic values were pooled for the classic and revised views. The results indicate that both the classic (0.63, 95% CI 0.52-0.75) and revised (0.60, 95% CI 0.43-0.72) views of stratification offered limited predictability, with the classic view performing slightly better than chance, and less so for the revised view of the $CHADS_2$ score. The pooled analysis for the c statistic values indicated high levels of heterogeneity. This should be considered when interpreting the results.

Calibration of annual ischaemic stroke risk

The pooled estimates indicate that there is good calibration between the predicted and observed events in each of the $CHADS_2$ score risk strata, for both the classic and revised view (Fig. 3). However, these results should be interpreted with caution as visual inspection of the forest plots coupled with the I^2 values (ranging from 33% to 88%) suggest high levels of heterogeneity.

In an attempt to account for the heterogeneity, a subgroup analysis based on prevalence was conducted for each of the three risk strata for the classic and revised views. Prevalence of stroke in each validation data set ranged from 0% to 22.5%. A study was classified as high prevalence if the associated prevalence was higher than that reported by the $CHADS_2$ original study (4.9%). Four data sets were classified as high prevalence and the remaining four were classified as low prevalence. Overall, the subgroup analysis reduced the heterogeneity, with most risk strata associated with low heterogeneity. The pooled estimates suggest that the calibration for the high prevalence group for both the classic and revised view remained unchanged relative to the original analysis. The pooled estimates from the low prevalence group suggest that the $CHADS_2$ score over predicts the risk of stroke in the moderate (RR 5.27, 95%CI 1.54-18.02) and high (RR 3.97, 95%CI 1.35-11.85) risk categories for the classic view of stratification for the $CHADS_2$ score and for the moderate risk category with the revised view of the $CHADS_2$ score, although the confidence intervals were very wide and included 1.0 (RR 2.72, 95%CI 0.50-14.62). Once again, these results should be interpreted with caution due to the small number of studies in each prevalence group and the wide confidence intervals around the pooled estimates.

Discussion

Summary of results

The diagnostic test accuracy analysis of the different cut-points produced mixed results in terms of the strength of the observations. A cut-point of ≥1 (i.e. a $CHADS_2$ score of 0) or ≥2 (i.e. a $CHADS_2$ score of 0 or 1) is useful for ruling out the likelihood of stroke when absent and a cut-point of ≥3 or ≥4 is useful for ruling in the likelihood of stroke when present. However, only the cut-points of ≥1 and ≥4 were associated with high summary estimate values. There was insufficient data to calculate a score for a cut-point of either ≥5 or ≥6. The pooled analysis of the *c* statistic values indicates that both the classic and revised views of stratification offer limited predictability. The pooled analysis from the calibration of the $CHADS_2$ score suggests good calibration for both the classic and revised view of stratification. The results from all of these analyses should be judged in the context of high heterogeneity and low event rates across all risk strata.

The classic view of stratification of the $CHADS_2$ score classified the majority of patients as moderate risk, while the revised view of stratification classified the majority of patients as high risk. Quality assessment of the internal and external validity of the included studies produced mixed results.

Strengths and weaknesses

The strengths include: the restriction of stroke type to only one type of stroke (ischaemic stroke); adjusting for the net benefit of protection offered with warfarin or aspirin; comparing the utility of the classic and revised view of risk stratification models; the inclusion of original data from authors; the inclusion of real-world, non-trial data; and pooling the results of studies to conduct a formal quantitative validation of the $CHADS_2$ score.

We acknowledge that our review has several limitations. There is modest to high heterogeneity across the studies included in this review. This was problematic for the diagnostic test accuracy, as well as the validation of the $CHADS_2$ score. Heterogeneity can be caused by a number of factors including chance, variation in the pre-test probability of having an ischaemic stroke and other factors. However, the current study controlled for one potential source of heterogeneity in the calibration element of the study by conducting a sub-group analysis according to the prevalence of ischaemic strokes in each data set.

A further potential source of heterogeneity is the variability in the intensity of anticoagulation across the individual studies. Research indicates that patients outside the therapeutic INR range are at an increased risk of thrombotic events, with patients in a sub-therapeutic INR range associated with more frequent and more severe strokes. (17, 18) In the current work, some studies failed to report the achieved international normalised ratio (INR) levels, (12, 15) whilst another study reported variation in acceptable INR levels for different age groups within the study. (14) This may potentially introduce some bias in

terms of the number of strokes occurring in each study. There is insufficient data in the current study to perform a subgroup analysis in this regard. Nevertheless, intensity of anticoagulation should be considered in all future meta-analyses in this area.

Only five studies (eight data sets) were included in the current study. This figure does not reflect the large number of papers that have been published on the $CHADS_2$ score, which may bias the results. However, as outlined in the introduction, there are a number of methodological disparities across these validation studies. The current study has attempted to limit the impact of this type of heterogeneity by restricting the stroke type to ischaemic stroke only (excluding transient ischaemic attacks, haemorrhagic strokes and all other non-central nervous system emboli). We view this as an advantage to the current work, which is designed to test the validity of the $CHADS_2$ score in predicting ischaemic stroke.

The pooled calibration method of analysis used to pool the data sets in the validation of the $CHADS_2$ score is based on a comparative approach previously reported in a single validation study. (19) The current method extends this method and employs the absolute risk from the derivation study as a model to generate predicted values in multiple validation studies. The absolute risk is presented according to the relevant risk strata of the CPR so that the clinical value of the CPR according to these strata can be assessed.

One potential limitation is the need to adjust for the net benefit of protection offered by warfarin or aspirin in each of the data sets included. Limited placebo data exist in patients with NVAF and current treatment guidelines recommend that all patients with NVAF be provided with some form of antithrombotic therapy. (20, 21) As a result, it was necessary to adjust for the treatment effect according to the risk reduction offered by warfarin and aspirin. This value does not account for the variation associated with the upper and lower confidence intervals.

It was not possible to determine the diagnostic accuracy of a cut-point of ≥ 5 or ≥ 6 in the current study due to insufficient data. As such, caution should be applied when interpreting these results to prevent the incorrect assumption that a lack of data equates to a lack of evidence to support the clinical validity of these cut-points.

There was wide variability across studies for the c statistic analysis. This may result from methodological differences, including differences between study populations so that c-statistics from one study cannot necessarily be directly compared to that derived from a different study. The results from the pooled *c* statistic values reported here are lower than the original $CHADS_2$ study. (4) However, the lower values are consistent with more recent literature and this is attributed to a change in the risk factor profile in current patient populations. (22)

Implications for practice

The results from the diagnostic accuracy suggest that two cut-points are helpful in terms of clinical decision making. A cut-point of ≥1 is associated with high sensitivity and is useful for ruling out the likelihood of stroke when a $CHADS_2$ score of ≥1 is absent (i.e. a $CHADS_2$ score of 0). A cut-point of ≥4 (i.e. a $CHADS_2$ score of 4, 5 or 6) is associated with high specificity and is therefore useful for ruling in the likelihood of stroke. The results indicate that cut-points of ≥2 and ≥3 were associated with relatively lower levels of sensitivities and specificities. It is not possible to comment on the clinical validity of individual cut-points of ≥5 or ≥6 as there was insufficient data available for analysis. Clinicians should therefore not score out the utility of an individual $CHADS_2$ score of 5 or 6 in predicting stroke.

The pooled results from the calibration of the $CHADS_2$ score suggest that is a good predictor of ischaemic stroke risk in people with NVAF across all three risk strata of low, moderate and high risk, particularly in high prevalence populations. The revised view of the risk strata (low 0; moderate 1; and high 2-6) offers marginally better prediction than the classic view (low 0; moderate 1-2; and high 3-6). The revised view of classification is helpful in terms of therapeutic decision making as it results in more people being classified as high risk and fewer patients classified as moderate risk. This means that more patients (i.e. those in the high risk stratum) will receive with a clear recommendation for anticoagulation therapy and fewer patients (i.e. those in the moderate risk stratum) will receive an ambiguous recommendation for either antiplatelet or anticoagulation therapy. However, high heterogeneity between studies was observed for both the discrimination and calibration analyses. Therefore, results should be interpreted cautiously.

Future research

Different risk factors are associated with different incidence of stroke. (23, 24) Prior history of stroke or TIA is a strongly weighted risk factor in the $CHADS_2$ score and is associated with a total of 2 points, double that of other risk factors. However, a $CHADS_2$ score of 2 may be comprised of this single risk factor or may be comprised of two individual risk factors associated with 1 point (e.g. age and hypertension). Without individual patient level data it is not possible to test for this in the current study. This suggests one avenue for future research.

The current study offered a comparison between the predictive ability of the classic and revised view of risk stratification for the $CHADS_2$ score. However, the $CHADS_2$ score is only one of a number of CPRs that have been developed to predict stroke in patients with atrial fibrillation. This remains a potential source of confusion for the clinician in terms of selecting the optimal CPR to predict stroke in this population. Future research should validate the performance of individual CPRs across different studies and data sets, as well as compare different rules within the same data set. Furthermore, each of these CPRs should undergo formal impact analysis to determine if their application impacts physician behaviour or changes patient outcomes.

Recent European guidelines have recommended against artificially stratifying patients into low, moderate and high stroke risk strata. (25) Instead, these guidelines suggest a risk factor-based approach to inform recommendations for antithrombotic therapy. The

CHA_2DS_2-VASc (a modified version of the $CHADS_2$ score) offers one such approach. (6) Research suggests that the CHA_2DS_2-VASc accurately identifies those patients with atrial fibrillation who are truly at low risk of thromboembolism. (24) This is particularly important given that recent data suggests that new and safer anticoagulation drugs could now be offered at a substantially lower risk of ischaemic stroke (>0.9% per year). (26) Therefore, it is necessary to be able to accurately discriminate between those patients who are truly low risk and do not need antithrombotic therapy and those with one or more risk factors that would benefit from oral anticoagulation.

Conclusion

The pooled *c* statistic and calibration analysis suggests minimal clinical utility of both the classic and revised view of the $CHADS_2$ score in predicting ischaemic stroke across all risk strata. Due to high heterogeneity across studies and low event rates across all risk strata, the results presented here should be interpreted cautiously. Further validation of $CHADS_2$ score should perhaps be undertaken, given the methodological differences between many of the available validation studies and the original $CHADS_2$ derivation study.

Acknowledgements: We would like to acknowledge and thank Yutao Guo, Ole Hauch*, Judith Stephenson, Laurie Jacobs and Mart•n Ruiz Ortiz for providing additional information. *In the interests of full disclosure, we would like to disclose that Ole Hauch in an employee of AstraZenca and owns AstraZenca stock options.

References

1. Heeringa J, van der Kuip DA, Hofman A, et al. Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. Eur Heart J. 2006;27:949-53.

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. Hannon N, Sheehan O, Kelly L, et al. Stroke associated with atrial fibrillation--incidence and early outcomes in the north Dublin population stroke study. Cerebrovasc Dis. 2010;29:43-9.

4. Gage BF, Waterman AD, Shannon W, et al. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA. 2001;285:2864-70.

5. Hart RG, Pearce LA. Current status of stroke risk stratification in patients with atrial fibrillation. Stroke. 2009;40:2607-10.

6. Lip GY, Nieuwlaat R, Pisters R, et al. 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.

7. McGinn TG, Guyatt GH, Wyer PC, et al. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. JAMA. 2000;284:79-84.

8. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58:982-90.

9. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomized controlled trials. Arch Intern Med. 1994;154:1449-57.

10. van Walraven C, Hart RG, Singer DE, et al. Oral anticoagulants vs aspirin in nonvalvular atrial fibrillation: an individual patient meta-analysis. JAMA. 2002;288:2441-8.

11. Masaki N, Suzuki M, Iwatsuka R, et al. Effectiveness of risk stratification according to CHADS2 score in Japanese patients with nonvalvular atrial fibrillation. Int Heart J. 2009;50:323-9.

12. Boccuzzi SJ, Martin J, Stephenson J, et al. Retrospective study of total healthcare costs associated with chronic nonvalvular atrial fibrillation and the occurrence of a first transient ischemic attack, stroke or major bleed. Curr Med Res Opin. 2009;25:2853-64.

13. Jacobs LG, Billett HH, Freeman K, et al. Anticoagulation for stroke prevention in elderly patients with atrial fibrillation, including those with falls and/or early-stage dementia: a single-center, retrospective, observational study. Am J Geriatr Pharmacother. 2009;7:159-66.

14. Guo Y, Wu Q, Zhang L, et al. Antithrombotic therapy in very elderly patients with atrial fibrillation: is it enough to assess thromboembolic risk? Clin Interv Aging. 2010;5:157-62.

15. Ruiz Ortiz M, Romo E, Mesa D, et al. Oral anticoagulation in nonvalvular atrial fibrillation in clinical practice: impact of CHADS(2) score on outcome. Cardiology. 2010;115:200-4.

16. Rossi A. Statistics of observational studies: basic concepts for medical writers The Journals of the European Medical Writers Association. 2009;18:165-8.

17. Rietbrock S, Plumb JM, Gallagher AM, et al. How effective are dose-adjusted warfarin and aspirin for the prevention of stroke in patients with chronic atrial fibrillation? An analysis of the UK General Practice Research Database. Thromb Haemost. 2009;101:527-34.

18. Hylek EM, Rose AJ. Understanding low INR in clinical practice. Thromb Haemost. 2009;101:417-8.

19. Bont J, Hak E, Hoes AW, et al. Predicting death in elderly patients with community-acquired pneumonia: a prospective validation study reevaluating the CRB-65 severity assessment tool. Arch Intern Med. 2008;168:1465-8.

20. National, Collaborating, Centre, et al. Atrial fibrillation: National clinical guideline for management in primary and secondary care. London: Royal College of Physicians. 2006.

21. Singer DE, Albers GW, Dalen JE, et al. Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest. 2008;133:546S-92S.

22. Karthikeyan G, Eikelboom JW. The CHADS2 score for stroke risk stratification in atrial fibrillation--friend or foe? Thromb Haemost. 2010;104:45-8.

23. Gorin L, Fauchier L, Nonin E, et al. Antithrombotic treatment and the risk of death and stroke in patients with atrial fibrillation and a CHADS2 score=1. Thromb Haemost. 2010;103:833-40.

24. Olesen JB, Lip GY, Hansen ML, et al. Validation of risk stratification schemes for predicting stroke and thromboembolism in patients with atrial fibrillation: nationwide cohort study. BMJ. 2011;342:d124.

25. Camm AJ, Kirchhof P, Lip GY, 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.

26. Eckman MH, Singer DE, Rosand J, et al. Moving the tipping point: the decision to anticoagulate patients with atrial fibrillation. Circ Cardiovasc Qual Outcomes. 2011;4:14-21.

Legends to Figures and Tables

Figure 1: Overview of search strategy. This is an overview of the number of articles included during each stage of the systematic review process.

Figure 2: Methodological quality of studies included in the review (McGinn Criteria). This provides an overview of the methodological quality of the final set of articles included in the systematic review and meta-analysis. The methodological quality criteria concern both internal and external validity.

Figure 3: Results from classic view [low (0), moderate (1-2) and high (3-6)] and revised view [moderate (1) and high (2-6)] of the $CHADS_2$ score. This provides a graphical representation of the calibration of the $CHADS_2$ score. A RR <1 indicates that the $CHADS_2$ score under-predicts the risk of stroke. A RR >1 indicates that the $CHADS_2$ score over-predicts the risk of stroke. A RR$=1$ indicates a perfect calibration between the observed and predicted number of strokes.

ON-LINE FIGURE: Summary receiver operating characteristics (sROC) plots for each cut-off point of the $CHADS_2$ CPR* This provides a graphical view of pooled estimates for each of the cut-points of the $CHADS_2$ score. The 95% confidence region and 95% prediction region illustrate the precision with which pooled values are estimated and to illustrate the amount of between study variation, respectively.

Table 1: Summary of the $CHADS_2$ clinical prediction rule stratified according to low, moderate and high risk (associated stroke rate). This provides a detailed overview of the different stroke risk stratification schemes, the classic view and the revised view. Details included are the associated $CHADS_2$ score, stroke rate and treatment recommendations.

Table 2: Characteristics of studies included in the review. This provides specific details for each of the data sets from each study included in the final systematic review and meta-analysis.

Table 3: Number of patients, number of ischaemic strokes and proportion of total sample categorised as low, moderate and high risk for each included study for both the classic and revised view of the $CHADS_2$ score at one year follow-up. This provides details on the number of patients and number of ischaemic strokes observed in each study for each of the three risk strata for both the classic and the revised view. Total number of patients and the representative percentage of overall number of patients for low, moderate and high risk group for both the classic and revised view of stratification is also presented.

Table 4: Summary estimates of sensitivity and specificity for the $CHADS_2$ score. This provides details of the sensitivity and specificity and associated variance logit results for each of the individual $CHADS_2$ score cut-points ≥ 1 through ≥ 4. There was insufficient data to calculate values for a cut-point of ≥ 5 or ≥ 6.

Table 5: Predictive value of the $CHADS_2$ score for the classic and revised view of stratification measured using the *c* statistic (95% confidence intervals). This presents the *c* statistic for each of the individual studies, as well as the overall pooled estimate *c* statistic.

Table 1: Summary of the $CHADS_2$ clinical prediction rule stratified according to low, moderate and high risk (associated stroke rate)

Risk stratification	**Low risk**	**Moderate risk**	**High risk**
Classic view: $CHADS_2$ score	**0**	**1-2**	**3-6**
Associated stroke rate	1.9	3.4*	11.3*
Revised view: $CHADS_2$ score	**0**	**1**	**2-6**
Associated stroke rate	1.9	2.8	9.8*
Guideline recommendations (20, 21)	Antiplatelet treatment	Antiplatelet treatment or anticoagulation†	Anticoagulation recommended

Note: $CHADS_2$ score forms a cumulative score based on six clinical features. 1 point for each of congestive heart failure, hypertension, age $\geq$ 75, and diabetes mellitus and 2 points for prior history of stroke or transient ischaemic attack (TIA)

* Rates are the mean of the original probability intervals reported by Gage (4)

† Choice of treatment depends on physician and patient preferences

Table 2: Characteristics of studies included in the review

Authors	Study setting	Study type	Participants	Type of NVAF*	Treatment received	Outcome event	Duration of follow-up	Published Data
Boccuzzi 2009 (12) (warfarin)	Medical and pharmacy claims database; USA	Retrospective cohort	N=1724	Chronic	Warfarin (adjusted dose)	Ischaemic stroke	Mean = 12.94 months	No
Jacobs 2009 (13) (warfarin)	Geriatric ambulatory, urban primary care practice; USA	Retrospective cohort	N=90	Chronic (≥3 months)	Warfarin: adjusted dose	Ischaemic Stroke	1 year	No
Jacobs 2009 (13) (aspirin)	Geriatric ambulatory, urban primary care practice; USA	Retrospective cohort	N=16	Chronic (≥3 months)	Aspirin	Ischaemic stroke	1 year	No
Masaki et al 2009 (11) (warfarin)	Outpatient clinic; Japan	Prospective cohort	N=182	Paroxysmal	Warfarin: adjusted dose (Average INR†=1.89 ±0.31)	Cerebral infarction	Average = 703 ±88 days	Yes
Guo 2010 (14) (warfarin)	Clinical information systems database; China	Retrospective cohort	N=6	Paroxysmal, persistent,	Warfarin: adjusted dose.	Ischaemic stroke	1 year	No

				permanent and new onset	Target INR was 2.0 (range 1.6-2.5) but in those >80 years INR 1.5-1.8 accepted			
Guo 2010 (14) (aspirin)	Clinical information systems database; China	Retrospective cohort	N=42	Paroxysmal, persistent, permanent and new onset	Aspirin	Ischaemic stroke	1 year	No
Ruitz Ortiz 2010 (15) (Warfarin)	Outpatient cardiology clinics at University hospitals; Spain	Prospective cohort	N=556	Permanent	Warfarin: Adjusted dose. Target INR 2-3	Ischaemic stroke	Mean=2.4 years	No
Ruitz Ortiz 2010 (15) (Aspirin)	Outpatient cardiology clinics at University hospitals; Spain	Prospective cohort	N=199	Permanent	Aspirin	Ischaemic stroke	Mean=2.4 years	No

*NVAF=non-valvular atrial fibrillation †INR= international normalised ratio

Table 3: Number of patients, number of ischaemic strokes and proportion of total sample categorised as low, moderate and high risk for each included study for both the classic and revised view of the $CHADS_2$ score at one year follow-up

Study	Classic view			Revised view		
	Low	Moderate	High	Low	Moderate	High
	N ischaemic strokes*/N patients					
Boccuzzi 2009 (12) (warfarin)**	9/449	81/1047	121/228	9/449	34/624	168/651
Jacobs 2009(13) (warfarin)	0/1	3/41	3/48	0/1	0/12	6/77
Jacobs 2009 (13) (aspirin)	0/1	0/9	0/6	0/1	0/4	0/11
Masaki 2009 (11)(warfarin)**	3/32	16/106	22/44	3/32	Unable to compile†	Unable to compile†
Guo 2010 (14) (warfarin)	0/0	0/0	0/6	0/0	0/0	0/6
Guo 2010 (14) (aspirin)**	0/2	4/20	3/20	0/2	1/10	5/30
Ruiz Ortiz 2010 (15) (warfarin)**	3/42	2/363	3/115	3/42	1/189	5/325
Ruiz Ortiz 2010 (15) (aspirin)	1/47	1/116	1/36	1/47	1/46	3/106
Total events/total N	16/574	107/1702	153/539	161/574	37/885	187/1206
% of total population in each strata	20.39%	60.46%	19.15%	20.39%	33.21%	45.25%

*The number of strokes is adjusted for the net benefit of protection offered by warfarin or aspirin treatment. ** The data is adjusted for follow-up to derive annual stroke data †Unable to compile as the published data was originally presented as stratified according to the classic view scoring system.

Table 4: Summary estimates of sensitivity and specificity for the $CHADS_2$ score*

$CHADS_2$ score	No. of studies	Sensitivity (95% CI)	Variance logit (sensitivity)	Specificity (95% CI)	Variance logit (specificity)
≥1	6	0.92 (0.82-0.96)	0.31	0.12 (0.06-0.24)	0.85
≥2	4	0.79 (0.64-0.89)	0.01	0.42 (0.24-0.63)	0.74
≥3	6	0.50 (0.37-0.63)	0.88	0.77 (0.59-0.88)	0.99
≥4	5	0.33 (0.21-0.47)	0.00	0.96 (0.66-0.10)	7.28

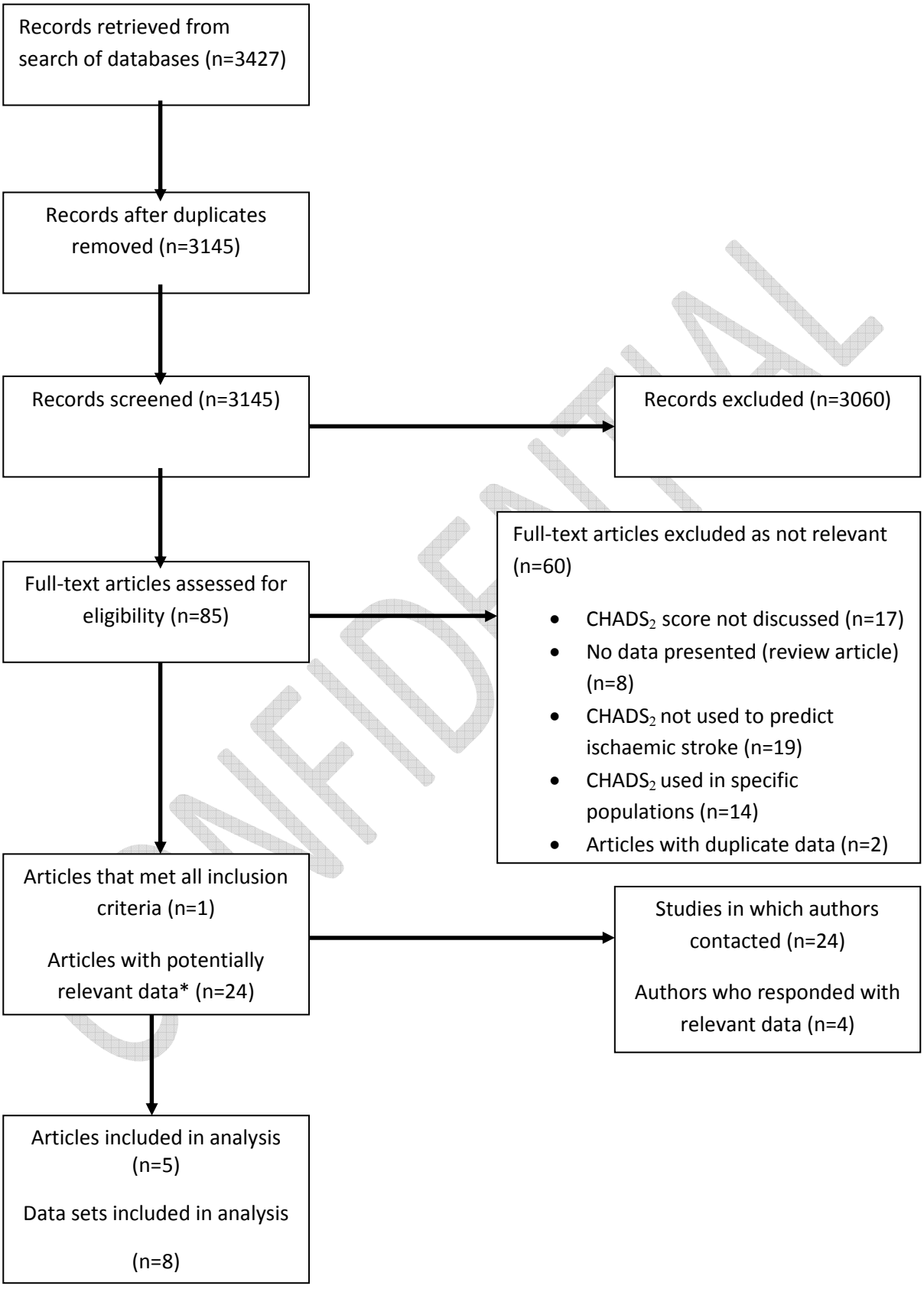
*There was insufficient data to examine the $CHADS_2$ score for ≥5 or ≥6

Table 5: Predictive value of the $CHADS_2$ score for the classic and revised view of stratification measured using the *c* statistic (95% confidence intervals)*

Study	Classic view c statistic (95% confidence interval)	Revised view c statistic (95% confidence interval)
Boccuzzi 2009 (12) (warfarin)	0.79 (0.76-0.82)	0.75 (0.73-0.78)
Guo 2010 (14) (aspirin)	0.53 (0.32-0.74)	0.57 (0.40-0.75)
Guo 2010 (14) (warfarin)	Insufficient data	Insufficient data
Jacobs 2009 (13) (aspirin)	Insufficient data	Insufficient data
Jacobs 2009 (13) (warfarin)	0.49 (0.26-0.71)	0.58 (0.54-0.62)
Masaki 2009 (11) (warfarin)	0.71 (0.62-0.79)	Data not available in this format
Ruiz Ortiz 2010 (15) (aspirin)	0.52 (0.07-0.97)	0.54 (0.28-0.79)
Ruiz Ortiz 2010 (15) (warfarin)	0.44 (0.17-0.70)	0.43 (0.22-0.65)
Pooled analysis	0.63 (0.52-0.75)	0.60 (0.43-0.72)

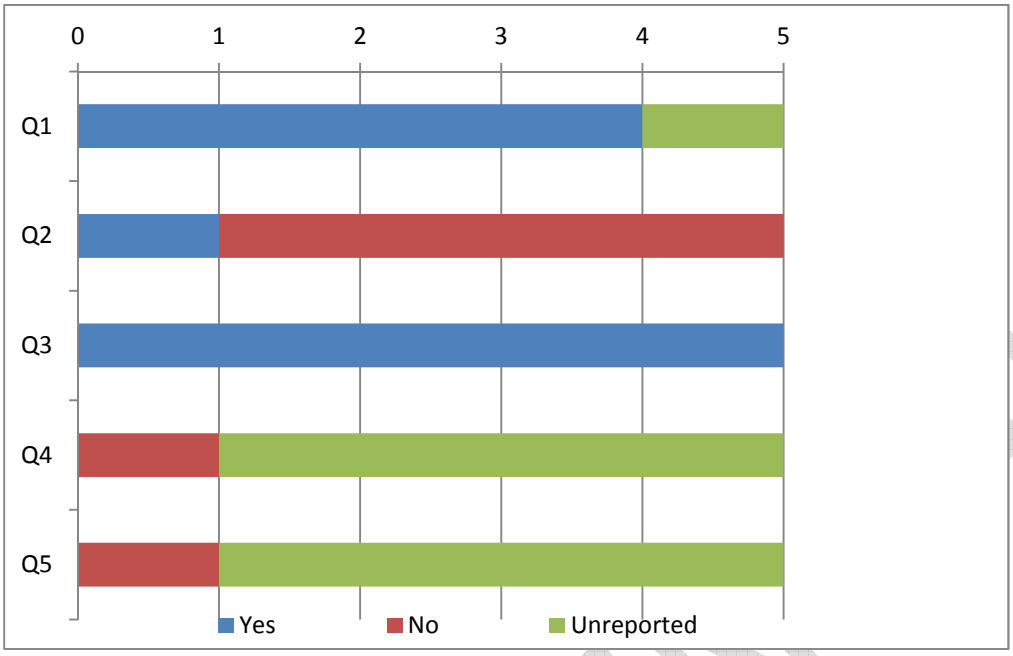
*The *c* statistic is calculated based on the number of strokes that have been adjusted for the net benefit of protection offered by warfarin or aspirin treatment.

Figure 1: Overview of search strategy



* These studies were selected on the basis that the author may have been able to provide data (e.g. data analysed in published paper included ischaemic stroke combined with haemorrhagic stroke)

Figure 2: Methodological quality of studies included in the review (McGinn Criteria)



Q1 (External validity): Do patients represent a wide spectrum of severity of disease?

Q2 (External validity): Were patients selected in an unbiased fashion?

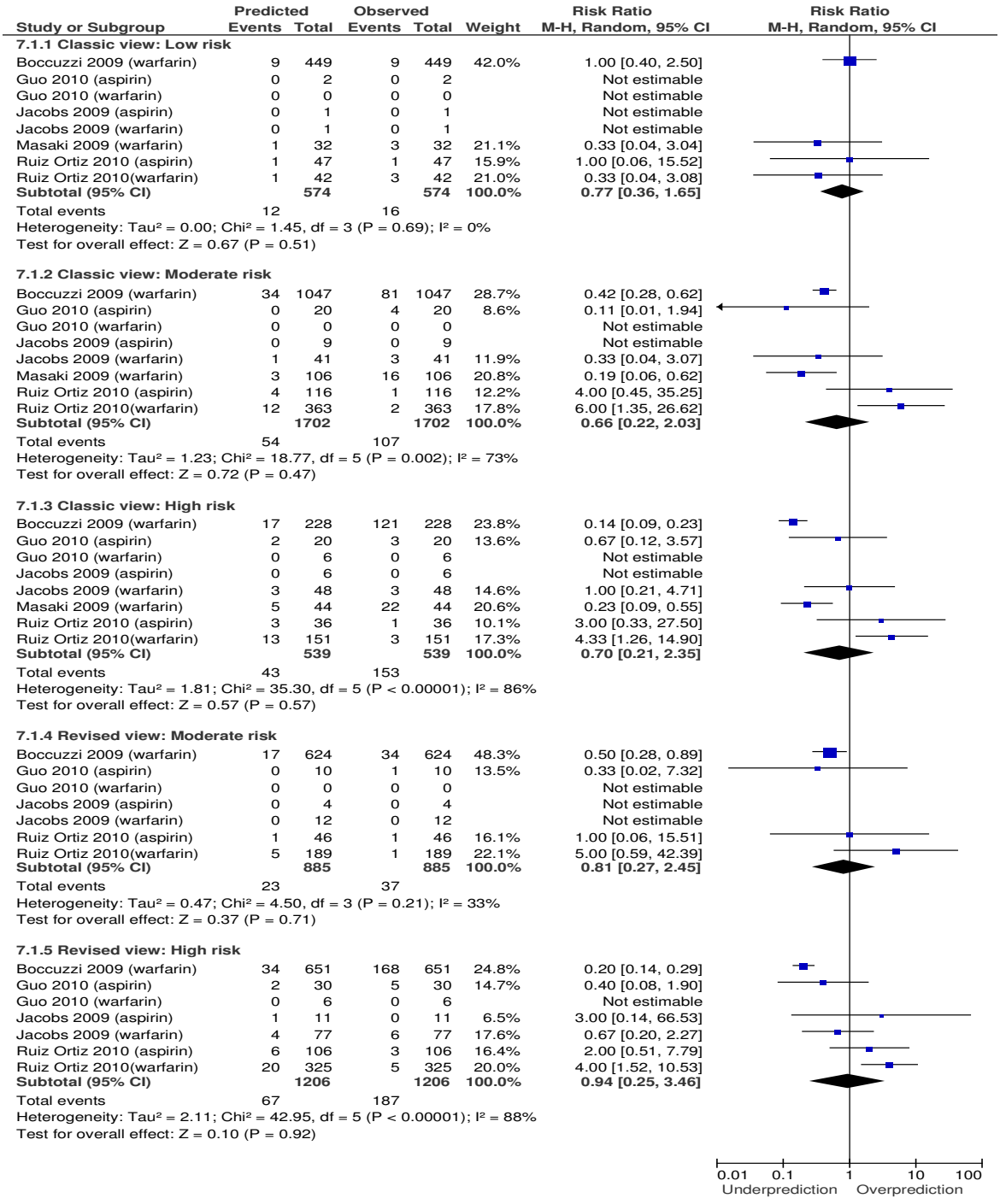
Q3 (Internal validity): Was there >80% follow-up of those enrolled?

Q4 (Internal validity): Were those assessing the presence of predictors blinded to the outcome event?

Q5 (Internal validity): Were those assessing the outcome event blinded to the presence of predictors?

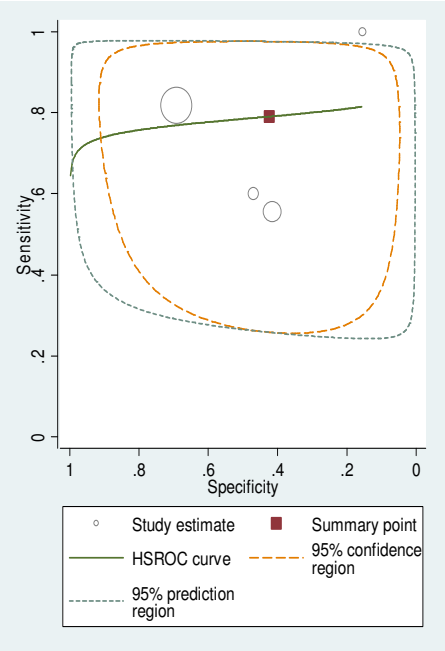
This article is not an exact copy of the original published article in Thrombosis and Haemostasis. The definitive publisher-authenticated version of Keogh C, Wallace E, Dillon C, Dimitrov BD, Fahey T. Validation of the CHADS2 clinical prediction rule to predict ischaemic stroke. A systematic review and meta-analysis. Thromb Haemost, 2011 Aug 31:106(3):528-38. Epub 2011 Jul 28 is available online at: http://www.schattauer.de/en/magazine/subject-areas/journals-a-z/thrombosis-and-haemostasis/contents/archive/issue/1439/manuscript/16384.html

Figure 3: Results from classic view [low (0), moderate (1-2) and high (3-6)] and revised view [moderate (1) and high (2-6)] of the $CHADS_2$ score

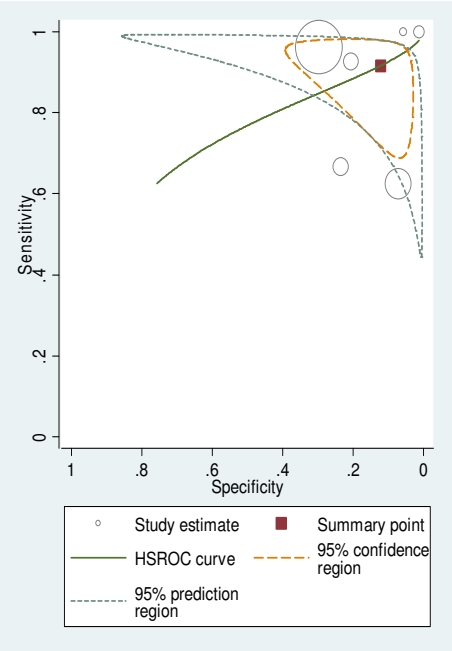


On-line version: Summary receiver operating characteristics (sROC) plots for each cut-off point of the $CHADS_2$ score*

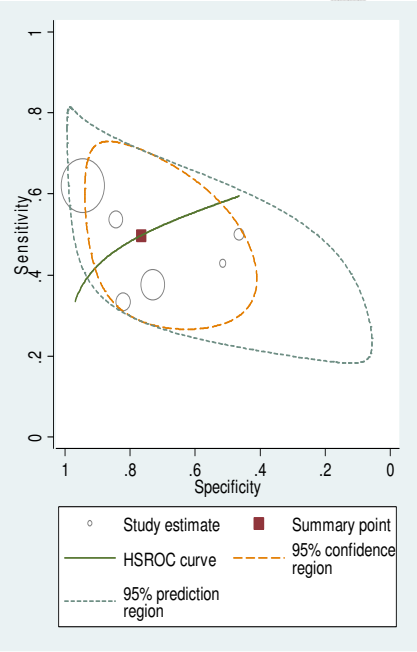
Cut point ≥1

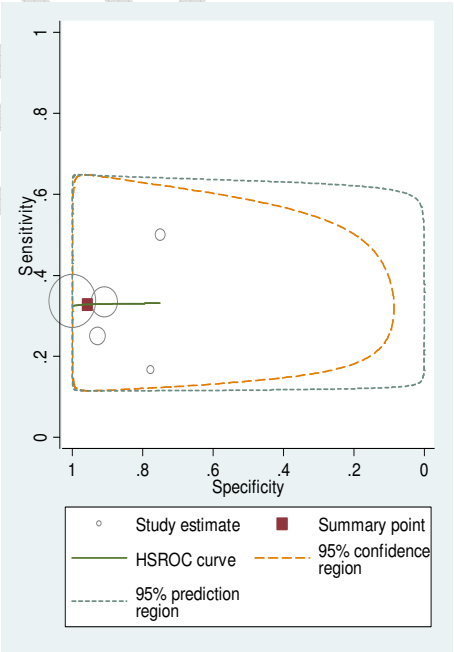


Cut point ≥2



Cut point ≥3





*There was insufficient data to complete this analysis for ≥5 or ≥6