Prognostic indices for early mortality in ischaemic stroke - meta-analysis

Katharina Mattishent[1]

Chun Shing Kwok[1]

Ashwin Mahtani[1],

Kelum Pelpola[2]

Phyo Kyaw Myint[3]

Yoon K Loke[1,]

[1]Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK
[2] Southend University Hospital Trust, Westcliff-on-Sea, Essex SS0 0RY, UK

[3] Institute of Applied Health Sciences, School of Medicine & Dentistry, University of Aberdeen, Aberdeen, Scotland, United Kingdom

Abstract word count: 249

Word count: 2800

Correspondence to:
Katharina Mattishent
Norwich Medical School,
University of East Anglia
Norwich NR4 7TJ
Telephone: 44 1603 591 234
Fax: 44 1603 593 752
Email: k.mattishent@uea.ac.uk

Short title: Prognostic scores for early mortality in ischaemic stroke

Contributors: KM, CSK, YKL and PKM conceptualized the review and developed the protocol. KM, CSK, YKL, KP and AM selected studies and abstracted the data; KM and YKL carried out the synthesis of the data and wrote the manuscript with critical input from all authors. YKL acts as guarantor for the paper.

Abstract

Objectives: Several models have been developed to predict mortality in ischaemic stroke. We aimed to evaluate systematically the performance of published stroke prognostic scores.

Methods: We searched MEDLINE and EMBASE in February 2014 for prognostic models (published between 2003-2014) used in predicting early mortality (< 6 months) after ischaemic stroke. We evaluated discriminant ability of the tools through meta-analysis of the area under the curve receiver operating characteristic (AUC) or c-statistic. We evaluated the following components of study validity: collection of prognostic variables, neuroimaging, treatment pathways, and missing data.

Results: We identified 18 articles (involving 163 240 patients) reporting on the performance of prognostic models for mortality in ischaemic stroke, with 15 articles providing AUC for meta-analysis. Most studies were either retrospective, or posthoc analyses of prospectively collected data; all but three reported validation data. The iSCORE had the largest number of validation cohorts (five) within our systematic review and showed good performance in four different countries, pooled AUC 0.84 (95% CI 0.82 – 0.87). We identified other potentially useful prognostic tools that have yet to be as extensively validated as iSCORE. - these include SOAR (2 studies, pooled AUC 0.79, 95% CI 0.78-0.80), GWTG (2 studies, pooled AUC 0.72, 95% CI 0.72-0.72)) and PLAN (1 study, pooled AUC 0.85, 95% CI 0.84 – 0.87).

Conclusions: Our meta-analysis has identified and summarized the performance of several prognostic scores with modest to good predictive accuracy for early mortality in ischaemic stroke, with the iSCORE having the broadest evidence base.

Key words: Mortality; Prognostic scores; Risk prediction model; Stroke

Introduction

Strokes are one of the leading causes of mortality and morbidity world-wide. Annually, 15 million people worldwide suffer a stroke; of these, 5 million die and another 5 million are left permanently disabled. (1) Mortality from stroke is particularly prominent in the first 30 days following the event. (2) A number of studies in recent years have focused on deriving and validating prognostic scores for early mortality after ischaemic stroke in the acute setting, (3-5) with one study demonstrating that prognostic scoring had substantially greater predictive accuracy than physicians' judgments. (6) Availability of reliable prognostic tools could improve clinical care, guide shared decision-making and enhance communication between clinicians and patients. The possibility of matching patients according to prognostic score also enables stroke physicians to do comparative evaluations of different models of stroke care, whether as part of quality improvement projects or clinical trials. However, absence of uniformly accepted prognostic tool amongst the myriad of options is an important barrier.

We are not aware of any recent meta-analyses of stroke prognosis tools, but there has been a previous systematic review published by Counsell in 2001. (7) This systematic review critically appraised 83 separate prognostic models and identified serious deficiencies in the statistical validity, generalizability and validity of the evidence at that time. There has since emerged a plethora of publications reporting on different stroke prognosis scores. (4, 5, 8-10)

Hence, we aimed to synthesize recent evidence on prognostic models in patients presenting acutely with ischaemic strokes, and to assess comparative performance of different scores so that clinicians and researchers can make informed decisions on use of such tools.

Methods

Eligibility criteria

We selected studies that used clinical variables (or groups of variables) in multivariate clinical prognostic models for overall mortality (< 6 months) in adult patients presenting with stroke. Eligible studies had to have a majority of participants with ischaemic stroke, with reporting of test performance through sensitivity/ specificity or area under receiver operating characteristic (AUC) or c-statistic. As our main aim was to produce a synthesis of up to date evidence, we restricted our selection to studies published from 2003 onwards.

We excluded studies that were designed solely to correlate mortality with laboratory (e.g. albumin, white cell count, copeptin, etc.) or radiological variables (such as size of lesion). We did not consider studies that reported only on functional outcomes, or were based only on patients requiring intensive care. As our main focus was on stroke patients presenting to healthcare facilities, we excluded studies that focused on mortality in specific subsets of patients e.g. following a particular intervention (i.e. after thrombolysis or thrombectomy), or those that specifically examined stroke in a particular brain area (e.g. thalamic, or basilar).

Search strategy

We searched MEDLINE and EMBASE (OvidSP interface, February 2014) using the search terms listed in Supplementary Material 1, without any language restriction. We also checked the bibliographies of included studies for any potentially suitable studies.

Study selection and data extraction

We allocated study screening and data extraction to pairs of reviewers (KM, CSK, KP, AM, YKL) who independently scanned all titles and abstracts for potentially relevant articles, before proceeding to obtain full text versions for further checking. Any uncertainties and discrepancies were resolved through discussion and with a third reviewer. We also contacted authors if there were any areas that required further clarification.

We used a standardized form for data collection which included details on the setting, date of study, country of origin, selection criteria, participant characteristics, and outcome measures.

Assessment of Study Validity

For the assessment of study validity, pairs of reviewers independently checked whether there was clear reporting of neuroimaging, time of patient assessment, missing or incomplete data, and treatment protocols.

Data analysis

We focused on the Area under the Receiver Operating Curve (AUC) or c-statistic (which are equivalent measures of the discriminant ability for binary outcomes). (11) Here, the discriminant ability reflects how well the model separates patients who die during follow-up as opposed to those who survive. For studies that reported on both derivation and validation components, we chose to analyze data relating to the validation portion. If different mortality time-points were reported in a single study, we focused on 30-day as the first choice, inpatient mortality as the second choice, and where neither were available, we accepted a time point (< 6 months) for analysis. If a number of AUC values were available for a particular prognostic tool, we calculated a weighted pooled average using random effects inverse variance meta-analysis. If the

AUCs were listed without standard errors, we imputed these values from the 95% confidence intervals or through Hanley's method. (12)

We assessed heterogeneity through the I^2 statistic and visual inspection of the Forest plots. The performance of the prognostic score was judged according to AUC thresholds that have been described by other researchers: excellent (AUC≥0.90), good (AUC≥0.80 and <0.90), fair (AUC≥0.70 and <0.80) and poor (AUC <0.70). (13)

Results

We included 18 relevant studies from 2374 hits that were retrieved through the electronic database search. (3-5, 8-10, 14-25) The flow chart of study selection is shown in Figure 1. Characteristics and results of the included studies are shown in Supplementary Table A1, while assessment of study validity is reported in Supplementary Table A2.

The included studies had a total of 163240 participants (sample sizes from 75 – 109995), with mean age 71 years, while 54% of the participants were male. There were 10 multi-centre studies that recruited patients from more than two healthcare sites. (3-5, 8-10, 18, 19, 22, 25) Geographical locations were diverse, and included North America, Europe, Egypt and Asia. All the studies evaluated score validation, except for three that were mainly derivation studies. (8, 19, 23)

Validity assessment

As the majority of studies were retrospective in design, or posthoc analyses of prospectively collected clinical data, we were seldom able to judge if the prognostic variables were collected early in the course of the presentation. Treatment pathways were seldom reported, with only three studies explicitly stating that participants received similar care. (14, 18, 19) We recorded more complete reporting of the modality used in neuroimaging (12 studies), as well as amount of missing data (10 studies). (Supplementary Table A2) In view of the lack of detail in methodological reporting, we have not attempted to categorize studies into either a high or low quality subgroup.

Quantitative comparison of AUC

We were able to evaluate the following prognostic models in the comparative quantitative analysis: iSCORE (five cohorts in four articles) (4, 14, 16, 22), NIHSS (three cohorts), GWTG (two cohorts), (5, 15, 25) Essen Stroke Risk Score (two cohorts), (18, 20) SOAR (2 cohorts) (3, 8) and PLAN (one cohort). (9)

The AUCs from individual studies, as well as pooled mean AUC across studies (and heterogeneity statistic) are shown in Figure 2. A summary of the information required in the calculation of each prognostic tool is available in Supplementary Table 3.

iSCORE

The performance of the iSCORE in predicting 30-day mortality has been evaluated in five cohorts with a total of 12833 participants from Canada, France, Greece and Korea. (4, 14, 16, 22) Point estimates of the AUC ranged from 0.79 to 0.86, with a weighted pooled average of 0.84 (0.82 – 0.87).

NIHSS

The performance of the NIHSS was reported in three cohorts with a total of 50864 participants from India (30-day mortality), North America and China (both focusing on inpatient mortality). (5, 15, 25) Point estimates of the AUC ranged from 0.83 to 0.89, with a weighted pooled average of 0.85 (0.82 – 0.88).

Essen Stroke Risk Score (ESRS)

The performance of the ESRS in predicting 90-day or inpatient mortality was reported in two cohorts with a total of 7570 participants from multiple centres. (18, 20) Point estimates of the AUC were identical in both studies, and yielded a weighted pooled average of 0.71 (0.69 – 0.72).

GWTG, with or without NIHSS

There were two studies reporting on the performance of the GWTG score on its own for predicting inpatient mortality. (5, 25) The studies enrolled at total of 117010 participants in North America and China. Both studies demonstrated consistent findings for the GWTG, with a weighted pooled average AUC of 0.72 (0.72 – 0.72). When the GWTG was considered together with NIHSS, the pooled AUC was markedly improved to 0.85 (0.84 – 0.87).

SOAR

The performance of the SOAR score with regards to predicting inpatient mortality was evaluated in two UK cohorts with a total of 15902 participants. Point estimates of the AUC ranged from 0.79 to 0.80, with a weighted pooled average of 0.79 (0.78 – 0.80). (3, 8)

PLAN

We identified only one study reporting on the PLAN score. (9) This study recruited 4904 participants in Canada and reported an AUC of 0.87 (0.85-0.88) for those who were not thrombolysed, and 0.72 (0.69- 0.75) for those who were thrombolysed. We estimated a weighted pooled average AUC of 0.85 (0.84-0.87) for the whole cohort.

Prognostic scores with AUC from single cohorts not included in comparative meta-analysis

We identified only one study reporting on the prognostic value of the GCS. This study recruited 1217 participants in Scotland and reported an AUC of 0.78 (0.75-0.81). (24)

Roquer et al. evaluated the prognostic value of the VRS II in 1527 patients in Spain and found an AUC of 0.71 (0.67-0.75) for inpatient mortality. (23)

In addition to testing the ESRS, Maier et al. also studied the RRE-90 and ABCD scores in predicting inpatient mortality, with respective AUCs of 0.64 (0.56–0.73) and 0.66 (0.59-0.73) (20)

Finally, one study reported an AUC of 0.73 without 95% confidence intervals for the Six Simple Variable model. (19)

Studies not suitable for quantitative AUC analysis

Three studies reported only on sensitivity and specificity of the prognostic model. (10, 17, 21)

There were two studies that enrolled small sample sizes (<100 patients) in single centres, which means that the data may have limited generalizability or applicability. El Sheikh et al. reported on the APACHE III score in 93 patients in Egypt and found a sensitivity of 0.89 and specificity of 0.70 for 30-day mortality. (17) Martinsson et al. evaluated 90-day mortality with the Barthel Index and Activities of Daily Living in 75 patients in Stockholm and reported a sensitivity of 0.81 and specificity of 0.53 for a Barthel Index of >10. (21)

A multicenter study of 1217 patients in Germany for the purposes of deriving and validating the ESRS found a sensitivity of 0.58 and specificity of 0.92 based on the threshold of 0.289 for the score. (10)

Discussion

Although the prediction of mortality in ischaemic stroke is complex, our review has identified several promising developments with moderate to good performance that can help clinicians and researchers decide which score to use. One of the frontrunners is the iSCORE. This prognostic model has the largest number of validation studies within our systematic review and has been tested in different countries (Canada, France, Greece and Korea) with consistently good results. (4, 14, 16, 22) An important barrier to the use of iSCORE by non-specialists is the need to calculate a neurological subscale, either the Canadian Neurological Scale (CNS) or NIHSS score beforehand. This additional step is potentially laborious and may require additional training. However, it is possible to calculate the iSCORE online [http://www.sorcan.ca/iscore/] or via a mobile application that has some guidance on the CNS score, and there appears to be less of a problem with missing data items with the CNS than with NIHSS. (26)

A further point to consider in relation to the iSCORE is that if the NIHSS score is already available, then that alone may be sufficient to provide prognostic accuracy similar to that of the iSCORE. We found that the NIHSS score has been reported in three cohorts from India, North America and China with a very similar weighted pooled average AUC (good predictive accuracy) to the iSCORE. Nevertheless, we also recognize that NIHSS scoring can be complex for non-specialists or difficult to obtain (missing in 60% of participants from a North American cohort), (5)and there are problems with inter-rater reliability. (26)

Based on the pooled average AUC, we would consider the GWTG and SOAR scores to have moderate performance in predicting mortality after ischaemic stroke; the major advantage being ease of use by non-specialists because neither of GWTG nor SOAR requires use of neurological severity subscales such as the NIHSS. However, each of these scores has been evaluated in only

two studies and we feel that there is a need to validate these further; a direct comparison with the iScore would be desirable. It is also important to note that one of the elements needed for the SOAR score is the Oxford Community Stroke Project classification (OCSP), which requires greater depth of knowledge, and may not always correlate well with findings on brain imaging. (27) There are issues arising from variation in inter-rater reliability with the OCSP and modified Rankin score (both of which are components of SOAR), (28) thus potentially leading to inconsistent estimates in the final score . An advantage of the GWTG score is that is does not require such pre-knowledge (of the NIHSS or OCSP for example) in order to complete it. However, two studies that directly compared GWTG with NIHSS found that NIHSS offered greater discriminant ability than GWTG alone. (5, 25)

The PLAN score also has a similar weighted pooled average AUC (good predictive accuracy) compared to the iSCORE and NIHSS, however, we found only one study reporting it. (9) In this study, the performance of PLAN in patients who received thrombolysis was weaker, for reasons which are as yet unclear. It does appear to be promising though in that it only uses few clinical variables which can be used as a bedside assessment tool and does not need specialist pre-knowledge of other subscales and classifications.

Back in 2001, Counsell's systematic review commented on the overall poor quality and lack of improvement in stroke prognosis research over a time period of two decades. (7) In comparison to Counsell's findings, our updated systematic review of studies published in the last ten years has identified larger, more rigorous studies that may have been previously lacking. Unlike the previous systematic review, we were able to carry out meta-analysis that reported appropriate

statistical measures from a variety of validation sets. We believe that the information from our systematic review will be very useful in helping researchers stratify and match patients when comparing mortality outcomes in observational studies of stroke care (e.g. between different healthcare centres, or different times of presentation such as weekends or weekdays).

However, the available studies do not report on acceptability and uptake of current prognostic scores in the day to day management of stroke patients. While good performance of a prediction rule is an important pre-requisite, patients will not gain any benefits from the profusion of prognostic scoring models if the uptake and implementation is patchy. There are parallels here with prognostic indices in community-acquired pneumonia, where an Australian survey found that only 12% of respiratory physicians and 35% emergency physicians reported regular use of the highly sensitive Pneumonia Severity Index. (29) The complexity of calculation proved challenging and many physicians were unable to accurately estimate the Pneumonia Severity Index scores. (29) Ideally, a prognostic score should be easy to use (without requiring specialist training or additional steps in having to calculate a subscale beforehand), memorable and accurate.

Our systematic review has limitations. We have focused only on research carried out over the last ten years and we chose not to emphasize functional outcomes because they are assessed in diverse ways, and determined to some extent by pre-stroke status. We aimed to specifically evaluate overall mortality as a hard outcome, bearing in mind findings from a recent systematic review where existing prognostic models in stroke had poor discriminant ability for recurrent stroke and myocardial infarction. (30) The majority of our included studies were retrospective, or

posthoc analyses of prospectively collected clinical data and we have not categorized studies into either high or low quality subgroups. We selected published studies which used the AUC or c-statistic as their primary measure; it is possible that studies that found poor discriminant ability may have been unpublished or unreported. The aetiology and severity of stroke can vary considerably across different geographical and ethnic populations, and a model that performs well in one hospital may perform less accurately in another setting without further re-calibration. We appreciate that prognostic models are imperfect, and should only be interpreted together with clinical information and judgment.

Conclusions

There are now a number of stroke prognostic scores showing moderate to good performance in predicting mortality after ischaemic stroke, and our review suggests that the iSCORE has the broadest supporting evidence base amongst the available prognostic tools.

We feel that the most promising recently validated models should all be compared directly in a large, prospective multi-centre international cohort measuring clinician uptake and ensuring treatment on the same pathway.

Declaration of competing interests and sources of funding: No funding was received for this study. The authors have no conflicts of interest to declare.

Acknowledgements: None

Figures and Tables

Figures

Figure 1: Flow chart of study selection

Figure 2: Meta-analysis of AUC for prognostic models

Supplementary material

Supplementary Material 1: Search strategy

Supplementary Tables 1,2 and 3

Table 1: Characteristics of included studies

Table 2: Validity assessment of included studies

Table 3: Variables required for estimation of prognostic score

Figure1: Flow Chart of Study Selection

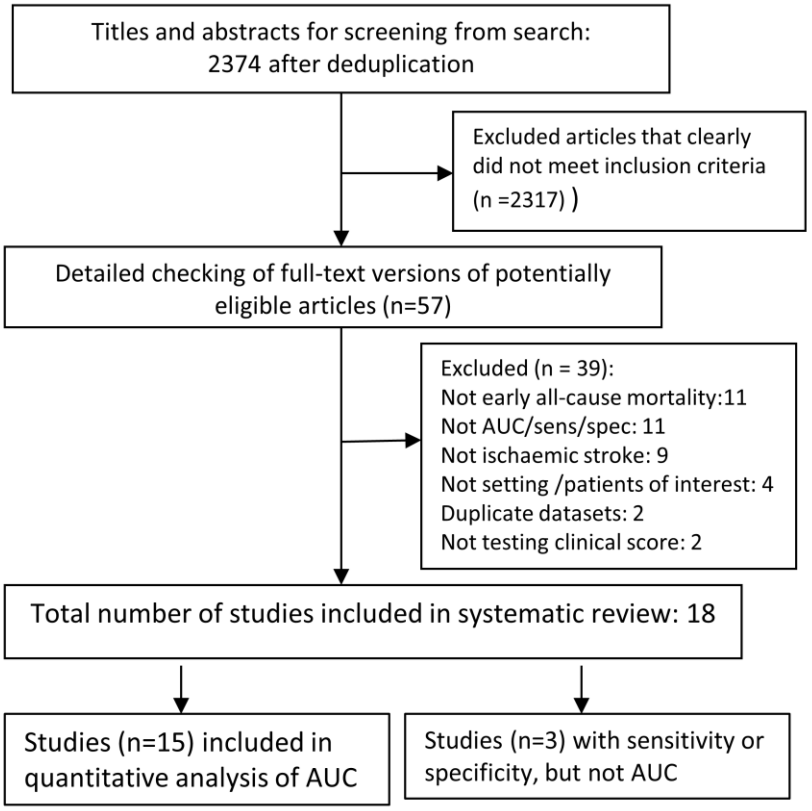
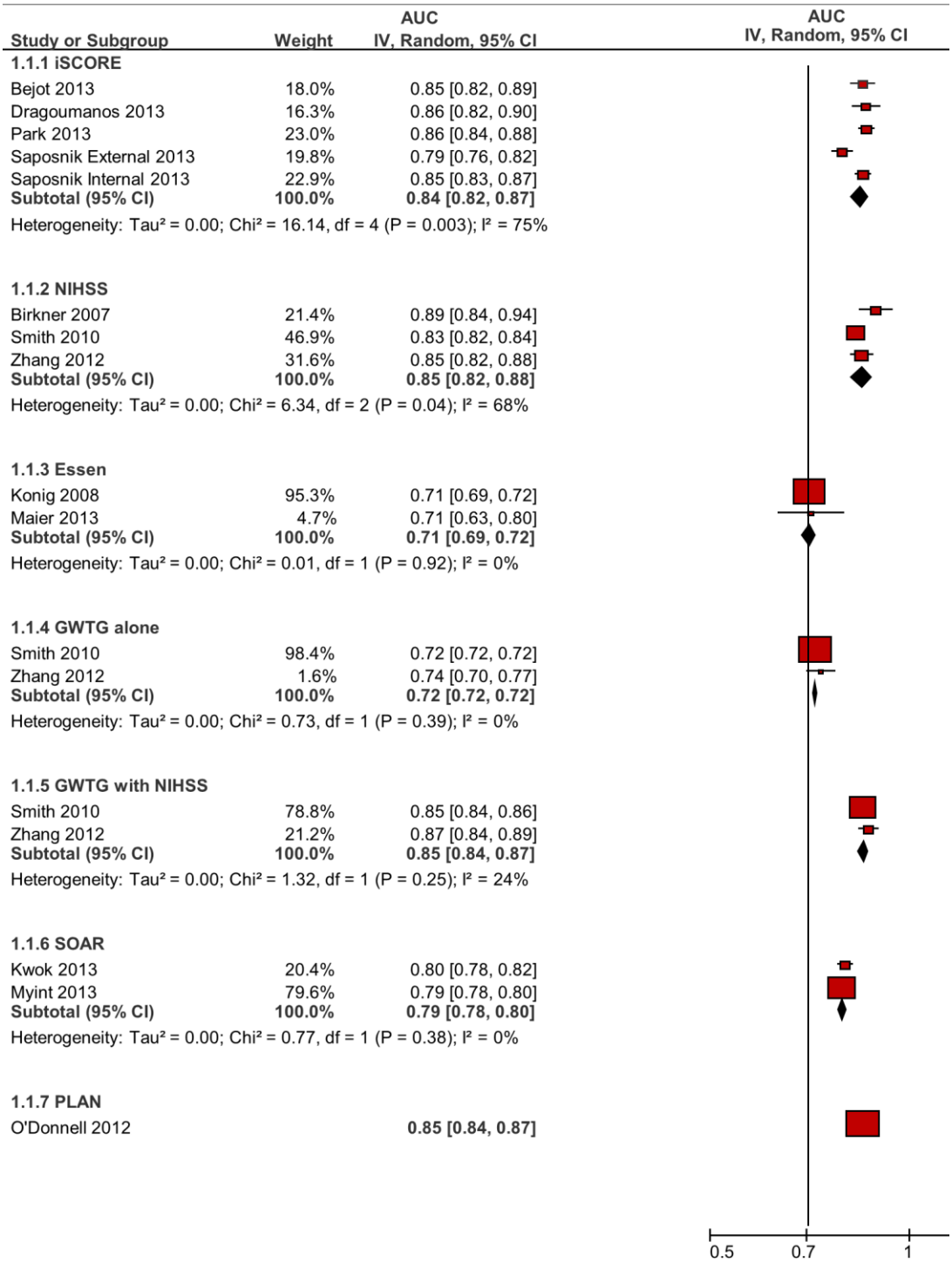


Figure 2. Meta-analysis of AUC for prognostic models



References

(1) WORLD-HEALTH-ORGANIZATION. Global burden of stroke. www.who.int/cardiovascular_diseases/en/cvd_atlas_15_burden_stroke.pdf. Accessed 8 April 2014.

(2) KWOK CS, SKINNER J, METCALF AK, POTTER JF, MYINT PK. Prior antiplatelet or anticoagulant therapy and mortality in stroke. Heart 2012; 98: 712-7.

(3) KWOK CS, POTTER JF, DALTON G, et al. The SOAR stroke score predicts inpatient and 7-day mortality in acute stroke. Stroke 2013; 44: 2010-2.

(4) SAPOSNIK G, KAPRAL MK, LIU Y, et al. IScore: a risk score to predict death early after hospitalization for an acute ischemic stroke. Circulation 2011; 123: 739-49.

(5) SMITH EE, SHOBHA N, DAI D, et al. Risk score for in-hospital ischemic stroke mortality derived and validated within the Get With the Guidelines-Stroke Program. Circulation 2010; 122: 1496-504.

(6) SAPOSNIK G, COTE R, MAMDANI M, et al. JURaSSiC: accuracy of clinician vs risk score prediction of ischemic stroke outcomes. Neurology 2013; 81: 448-55.

(7) COUNSELL C, DENNIS M. Systematic review of prognostic models in patients with acute stroke. Cerebrovasc Dis 2001; 12: 159-70.

(8) MYINT PK, CLARK AB, KWOK CS, et al. The SOAR (Stroke subtype, Oxford Community Stroke Project classification, Age, prestroke modified Rankin) score strongly predicts early outcomes in acute stroke. Int J Stroke 2014; 9: 278-83.

(9) O'DONNELL MJ, FANG J, D'UVA C, et al. The PLAN score: a bedside prediction rule for death and severe disability following acute ischemic stroke. Arch Intern Med 2012; 172: 1548-56.

(10) WEIMAR C, KONIG IR, KRAYWINKEL K, ZIEGLER A, DIENER HC. Age and National Institutes of Health Stroke Scale Score within 6 hours after onset are accurate predictors of outcome after cerebral ischemia: development and external validation of prognostic models. Stroke 2004; 35: 158-62.

(11) AUSTIN PC, STEYERBERG EW. Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable. BMC Med Res Methodol 2012; 12: 82.

(12) HANLEY JA, MCNEIL BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143: 29-36.

(13) MULLER MP, MCGEER AJ, HASSAN K, MARSHALL J, CHRISTIAN M. Evaluation of pneumonia severity and acute physiology scores to predict ICU admission and mortality in patients hospitalized for influenza. PLoS One 2010; 5: e9563.

(14) BEJOT Y, JACQUIN A, DAUBAIL B, DURIER J, GIROUD M. Population-based validation of the iscore for predicting mortality and early functional outcome in ischemic stroke patients. Neuroepidemiology 2013; 41: 169-73.

(15) BIRKNER MD, KALANTRI S, SOLAO V, et al. Creating diagnostic scores using data-adaptive regression: An application to prediction of 30-day mortality among stroke victims in a rural hospital in India. Ther Clin Risk Manag 2007; 3: 475-84.

(16) DRAGOUMANOS V, TZIROGIANNIS KN, PANOUTSOPOULOS GI, et al. Evaluation of IScore validity in a greek cohort of patients with type 2 diabetes. BMC Neurol 2013; 13.

(17) EL-SHEIKH WM. Stroke mortality: Predictive value of simple laboratory tests and APACHE III scoring system. Egyptian Journal of Neurology, Psychiatry and Neurosurgery 2010; 47: 489-95.

(18) KONIG IR, ZIEGLER A, BLUHMKI E, et al. Predicting long-term outcome after acute ischemic stroke: a simple index works in patients from controlled clinical trials. Stroke 2008; 39: 1821-6.

(19) LEWIS SC, SANDERCOCK PA, DENNIS MS. Predicting outcome in hyper-acute stroke: validation of a prognostic model in the Third International Stroke Trial (IST3). J Neurol Neurosurg Psychiatry 2008; 79: 397-400.
(20) MAIER IL, BAUERLE M, KERMER P, HELMS HJ, BUETTNER T. Risk prediction of very early recurrence, death and progression after acute ischaemic stroke. Eur J Neurol 2013; 20: 599-604.
(21) MARTINSSON L, EKSBORG S. Activity Index - a complementary ADL scale to the Barthel Index in the acute stage in patients with severe stroke. Cerebrovasc Dis 2006; 22: 231-9.
(22) PARK TH, SAPOSNIK G, BAE HJ, et al. The iScore predicts functional outcome in Korean patients with ischemic stroke. Stroke 2013; 44: 1440-2.
(23) ROQUER J, OIS A, RODRIGUEZ CAMPELLO A, et al. Clustering of vascular risk factors and in-hospital death after acute ischemic stroke. J Neurol 2007; 254: 1636-41.
(24) WEIR CJ, BRADFORD AP, LEES KR. The prognostic value of the components of the Glasgow Coma Scale following acute stroke. QJM 2003; 96: 67-74.
(25) ZHANG N, LIU G, ZHANG G, et al. A risk score based on Get With the Guidelines-Stroke program data works in patients with acute ischemic stroke in China. Stroke 2012; 43: 3108-9.
(26) BUSHNELL CD, JOHNSTON DC, GOLDSTEIN LB. Retrospective assessment of initial stroke severity: comparison of the NIH Stroke Scale and the Canadian Neurological Scale. Stroke 2001; 32: 656-60.
(27) ASDAGHI N, JEERAKATHIL T, HAMEED B, et al. Oxfordshire community stroke project classification poorly differentiates small cortical and subcortical infarcts. Stroke 2011; 42: 2143-8.
(28) QUINN TJ, DAWSON J, WALTERS MR, LEES KR. Exploring the reliability of the modified rankin scale. Stroke 2009; 40: 762-6.
(29) SERISIER DJ, WILLIAMS S, BOWLER SD. Australasian respiratory and emergency physicians do not use the pneumonia severity index in community-acquired pneumonia. Respirology 2013; 18: 291-6.
(30) THOMPSON DD, MURRAY GD, DENNIS M, SUDLOW CL, WHITELEY WN. Formal and informal prediction of recurrent stroke and myocardial infarction after stroke: a systematic review and evaluation of clinical prediction models in a new cohort. BMC Med 2014; 12: 58.

Online Supplementary Material: Search Strategy
Interface: OvidSP

Databases: MEDLINE, EMBASE
No language restrictions

Search terms used with .mp suffix: stroke AND score AND (prognostic OR prognosis OR predicti*) AND (mortality OR death OR survival)
For this search, .mp includes the fields of title, abstract, subject headings, heading words, original title, drug or device manufacturer, trade name, keyword, keyword heading word, unique identifier.

We checked bibliographies of included articles for any additional relevant studies.

We contacted authors for more information if there were any uncertainties when reviewing the articles.

We used online translation tools if there were any foreign language articles that we were unable to translate ourselves.

Online Supplementary Tables

Table A1: Characteristics of included studies

Study ID	Study setting and year, number of centres	Study design, and name of score	Patient population	Patients, n	Age, yr (mean or median)	Male, %	% Mortality at follow-up	AUC
Bejot 2013 (14)	Dijon Stroke Registry, two centres2006-2011	Retrospective validation, iScore	Acute ischaemic sroke	1199	76	46	30-day: not stated	30-day:0.85 (0.82-0.89)
Birkner 2007 (15)	Rural hospital, India, 1999-2001.	Prospective cohort, validation of NIHSS	Acute Stroke (66% ischaemic, 33% haemorrhagic)	175	59	62	30-day: 29%	30-day: sensitivity 0.92, specificity 0.65%, AUC 0.89 (0.84-0.94).
Dragoumanos 2013 (16)	Greece, tertiary hospital, 2008-2011	Prospective validation, iScore	Acute ischaemic stroke	534	75	49	30-day	30-day: 0.87 (0.80 – 0.93) and 0.85 (0.79 – 0.91)
El-Sheikh 2010 (17)	Hospital in Egypt, 2007-8	Prospective cohort validation, APACHE III	Acute stroke within 48 hrs	93	59	65	30-day: 18%	Score >40 Cerebral infarction: Sensitivity 0.89 Specificity 0.70
Konig 2008 (18)	VISTA Data Set patients extracted from 11 trials in	Retrospective cohort validation,	Ischaemic stroke patients in clinical trials	5843	69	56	90-day:18%	AUC = 0.706 (S.E. = 0.009) for prediction of survival after 3 months

	many countries	ESRS						
Kwok 2013 (3)	(8 hospitals in Anglia Stroke & Heart clinical network), Sept 2008 – Apr 2011	Retrospective, external validation SOAR	Ischaemic (92%) and hemorrhagic stroke	3547	Median around 76-80 years.	51	Inpatient: 17% 7-day: 6%	Inpatient mortality 0.80 (95% CI 0.78 – 0.82) 7-day mortality 0.82 (95% CI 0.79 – 0.84) Cutoff greater or equal to 3::Sensitivity: 0.73 (inpatient); Specificity: 0.76 (inpatient)
Lewis 2008 (19)	Participant in International Stroke Trial 3 before Feb 2007	Six simple variable (SSV) derivation model	Acute Ischaemic stroke patients in clinical trials	537	74	54	30-day: 21%	30-day AUC 0.73
Maier 2013 (20)	Hospital, Germany 2007-2011	Retrospective validation, RRE-90 (cut-off); ABCD; ESRS	Ischaemic stroke	1727	71	56	Inpatient mortality: not stated 7-day mortality: 2.3%	Inpatient ESRS 0.71 (0.63-0.79); ABCD 0.66 (0.59-0.73); RRE 0.64 (0.56-0.730 Early death (7-day): ESRS 0.58 (0.49 – 0.66); ABCD 0.65 (0.58 -0.72); RRE 0.72 (0.66-0.78)
Martinsson 2006 (21)	Trial patients in Stockholm, Sweden, from 1998 to 2001	Retrospective conort, validation, Barthel Index (BI), Activity Index (AI)	Ischaemic stroke in clnical trials	75	74	49	7-day or 90-day mortality rate not stated	7-day mortality: BI baseline score > 10 (sensitivity 0.76; specificity 0.80) and AI(ADL) baseline score > 15 (sensitivity 0.58; specificity 1.00) 90-day mortality: BI>10 (sensitivity 0.81; specificity

								0.53, AI(ADL) >15 (sensitivity 0.65; specificity 0.75)
Myint 2013 (8)	UK, 3 hospitals (1997 – 2010)	Retrospective derivation, SOAR	Ischaemic (91%) and haemorrhagic stroke	12355	Median around 76-80 years.	47	Inpatient 20% 7-day 10%	Inpatient mortality: 0·79 (95% CI 0·78–0·80) 7-day mortality 0·79 (95% CI 0·78–0·80)
O'Donell 2012 (9)	Canada. 11 centres, 2003-2008	Retrospective derivation and, validation PLAN	Acute ischaemic stroke	4904	73	52	30-day: 13.5%	30-day mortality: Not thrombolysed 0.87 (0.85 – 0.88); Thrombolysed: 0.72 (0.69-0.75)
Park 2013 (22)	12 centres Korea, 2011	Retrospective validation, iSCOre	Acute ischaemic stroke	4061	68	59	90-day 7.2%	0.861 (0.840-0.883)
Roquer 2007 (23)	Hospital, Spain 1997-2005	Retrospective derivation, VRS II	Acute ischaemic stroke	1527	73	51	Inpatient: 12.9%	Inpatient: AUC 0.711 (0.673–0.749)
Saposnik 2011 (4)	Multicentre Canada 2003-2008	Retrospective derivation and validation, iSCORE	Acute ischaemic stroke	Int val: 4039 Ext val: 3270	Int val: 72 Ext val: 74	Int val: 52 Ext 50	30-day: 12.6% (int validation) 11.6% (ext validation	AUC:Int validation 0.851 (SE 0.0109) Ext validation: 0.792 (SE 0.0142)
Smith 2010 (5)	Multicentre, US and Canada,	Retrospective derivation and	Acute ischaemic	109995 (validatio	74	47	Inpatient:	GWTG AUC overall 0.72 (SE 0.0038)

	2001-2007	validation, GWTG, NIHSS	stroke	n) NIHSS available: 43674			5.5% overall 5.2% NIHSS available	NIHSS available: 0.85 (SE 0.0051) NIHSS alone: 0.83 (SE 0.0054)
Weimar 2004 (10)	Multicentre, Germany, 2001-2002	Prospective derivation and validation, Age and NIHSS	Acute ischaemic stroke	1307 validation	68	57	100-day: rate not stated	Model II: sensitivity 0.579 specificity 0.915 based on 0.289 threshold
Weir 2003 (24)	Single centre, Scotland, 1990-1995	Retrospective validation, GCS	All strokes (87% ischaemic)	1217	71	49	14-day: 19%	AUC 0.78 (SE 0.0188) Score of E+V9 Sensitivity 0.74, Specificity 0.76
Zhang 2012 (25)	China registry multicenter, 2007-2008	Retrospective validation, NIHSS, GWTG	Acute ischaemic stroke	7015	68	61	In-hospital: 2.9%	GWTG alone: 0.735 (0.701–0.770) GWTG with NIHSS 0.867 (0.84-0.895) NIHSS alone 0.847 (0.816 – 0.879)

Abbreviations: BI = Barthel Index; ADL = Activities of Daily Living, GCS = Glasgow Coma Score

Table A2: Assessment of study validity

Study ID	Did the authors state that CT/MRI was done for all patients?	Was the index/score obtained early in course of presentation?	Did the authors give numbers or reasons on loss to follow-up or withdrawals?	Did the authors state whether the patients were treated on a standardized or similar care pathway?	Amount of missing data
Bejot 2013 (14)	☹	☹	☺	☺	107 (8.9% missing)
Birkner 2007 (15)	☺	☹	☺	☹	
Dragoumanos 2013 (16)	☺	☹	☹	☹	
El-Sheikh 2010 (17)	☺	☺	☹	☹	
Konig 2008 (18)	☹	☺	☹	☺	
Kwok 2013 (3)	☺	☹	☺	☹	59% of patients not eligible for SOAR incomplete data
Lewis 2008 (19)	☺	☺	☹	☺	
Maier 2013 (20)	☺	☹	☹	☹	
Martinsson 2006 (21)	☺	☹	☹	☹	
Myint 2013 (8)	☺	☹	☺	☹	

O'Donell 2012 (9)	☺	☹	☺	☹	225 incomplete data
Paark 2013 (22)	☹	☹	☺	☹	699 patients incomplete data
Roquer 2007 (23)	☺	☺	☺	☹	163 incomplete data
Saposnik 2011 (4)	☹	☺	☹	☹	
Smith 2010 (5)	☹	☺	☹	☹	
Weimar 2004 (10)	☺	☺	☺	☹	>200 patients incomplete data
Weir 2003 (24)	☺	☹	☺	☹	300 patients incomplete data
Zhang 2012 (25)	☹	☹	☺	☹	2623 + 265 NIHSS incomplete

Abbreviations: CT = Computerized tomography; MRI = magnetic resonance imaging

Table A3: Information needed to calculate prognostic scores in Stroke

Score	Patient characteristics and past history					Clinical examination			Stroke Classification	Laboratory measures				Other	Software required to calculate
	Age	Gender	Risk Factors	Comorbid conditions	Preadmission status	BP	Temp	Neurological	Stroke Subtype	Glucose	Hb	WBC	Creatinine		
iSCORE	Y	Y	AF, MI, CHF, Smoker	Cancer, Renal dialysis,	Disability	_	_	CNS or NIHSS	Lacunar, Non-lacunar, Unknown	Y	_	_	_	_	Yes (web or app available)
NIHSS	N	N	_	_	_	_	_	Level of consciousness, horizontal eye movement, visual field test, facial palsy, motor arm, motor leg, limb ataxia, sensory language, dysarthria, extinction and inattention	_	_	_	_	_	_	No
GWTG	Y	N	AF, Previous stroke /TIA, carotid stenosis (>50%), hypertension, dyslipidaemia,	Diabetes mellitus, CAD, PVD,	_	_	_	_	_	_	_	_	_	Mode of arrival, day & time of arrival	No

			smoker												
SOAR	Y	N	_	_	Prestroke Rankin Score	_	_	_	OSCP	_	_	_	_	_	No
PLAN	Y	Y	Hypertension, Hyperlipidemia, CHF, MI/angina, AF,	Diabetes, chronic liver disease, dementia, cancer	Dependence	Y	Y	Weakness of the face, arms, and legs, aphasia; dysphagia; neglect; visual field deficit; and side of the symptoms	_	_	Y	Y	Y	_	No

Abbreviations: AF = atrial fibrillation, MI = myocardial infarction, CHF = chronic heart failure, TIA = transient ischaemic attack, OCSP = Oxford Community Stroke Project