APPENDICES (online only)

APPENDIX 1

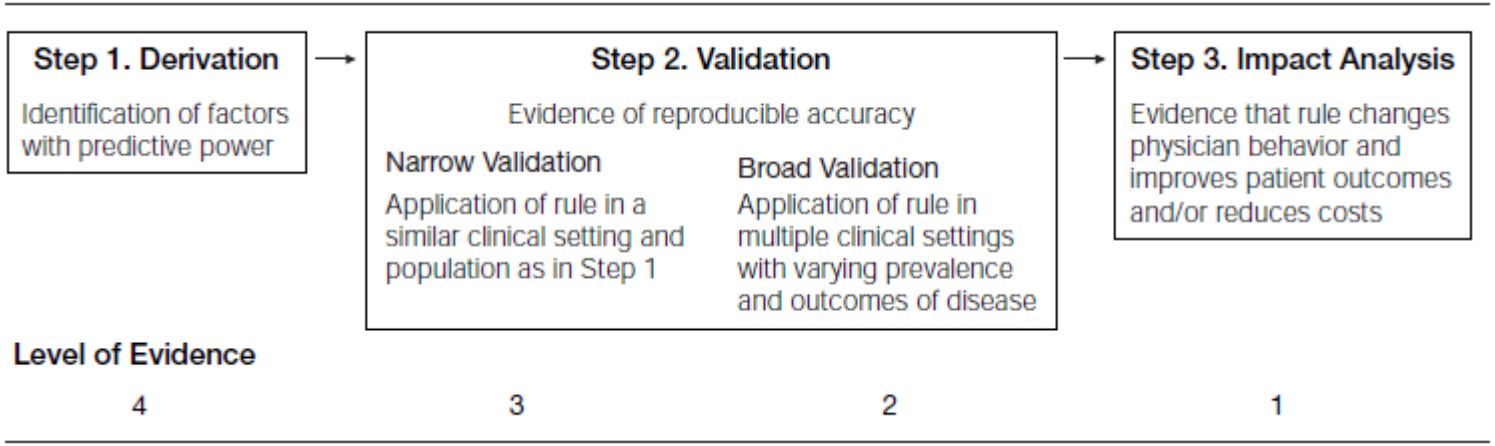


Fig. A1.1 Steps in the development of a clinical prediction rule
Reproduced with the written permission from the paper by McGinn et al [3] (JAMA, 2000)

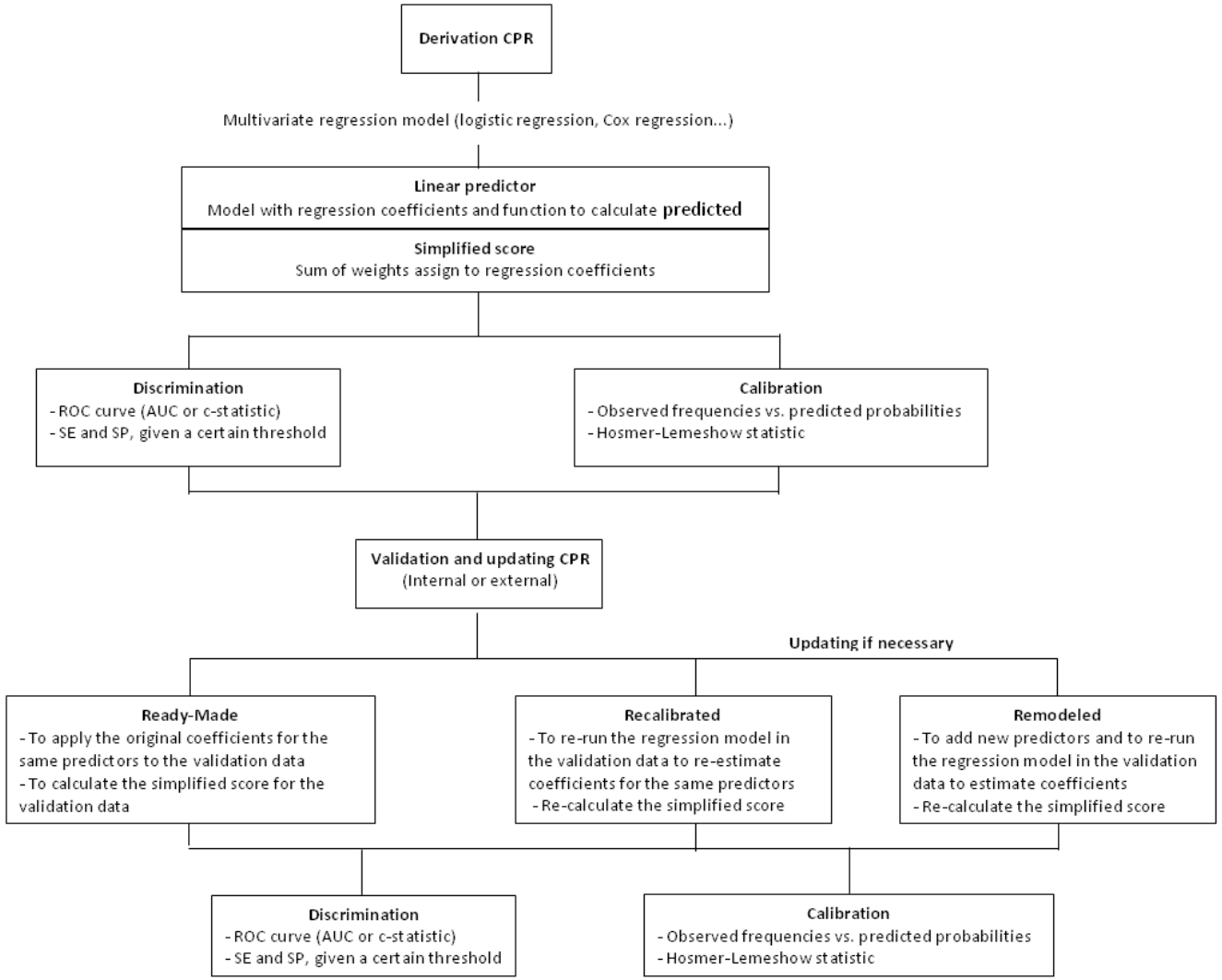


Fig. A1.2 Flow-chart of the process of derivation and validation of a clinical prediction rule (CPR)

Footnote: Recalibration may not always mean that the model is refitted; Remodelling may be also considered as a model revision or extension.

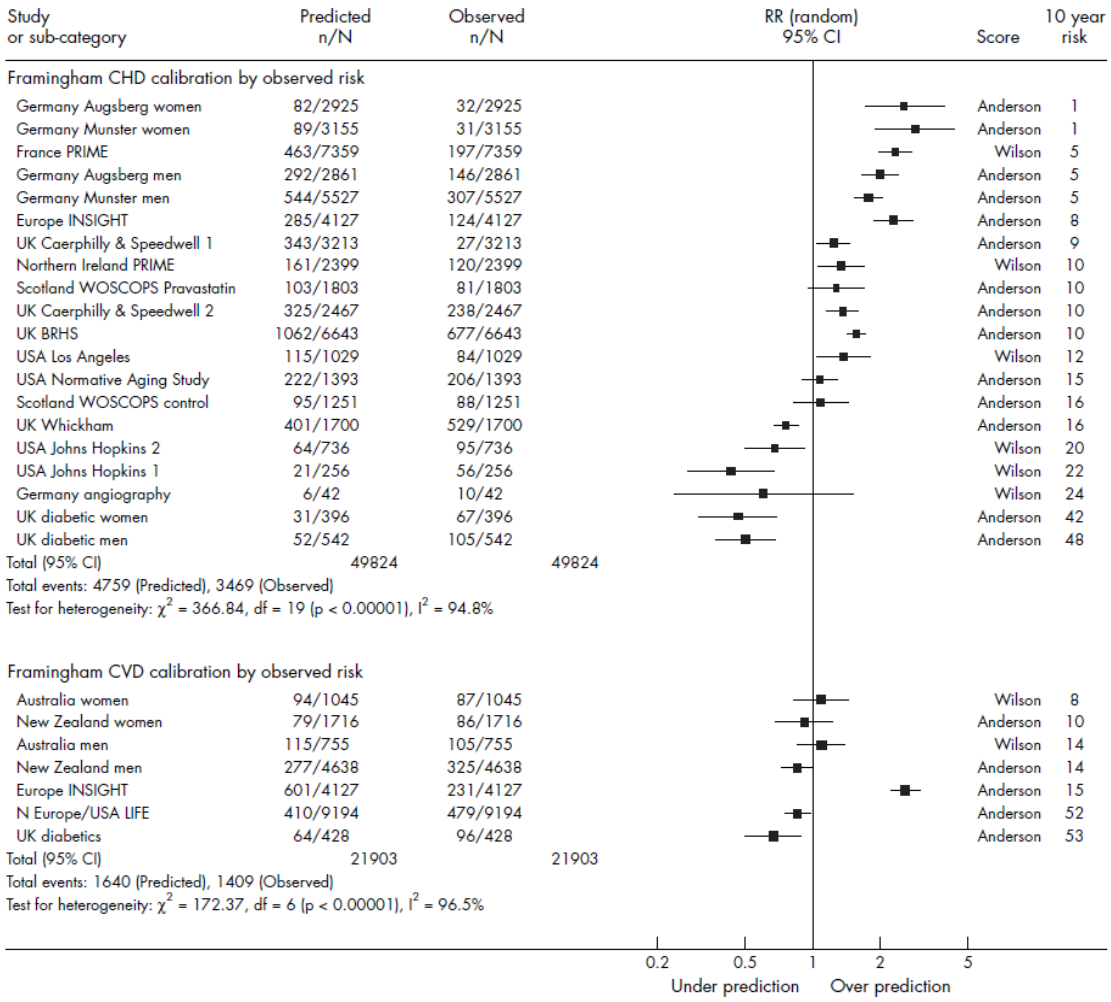


Fig. A1.3 Studies examining the predicted to observed ratio of Framingham Anderson and Wilson risk scores, ordered by the observed 10-year risk (%) in the test populations. *Reproduced with the written permission from the paper by Brindle et al (Heart, 2006*)*

**Reference: Brindle PM et al. Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review. Heart 2006; 92: 1752-59.*

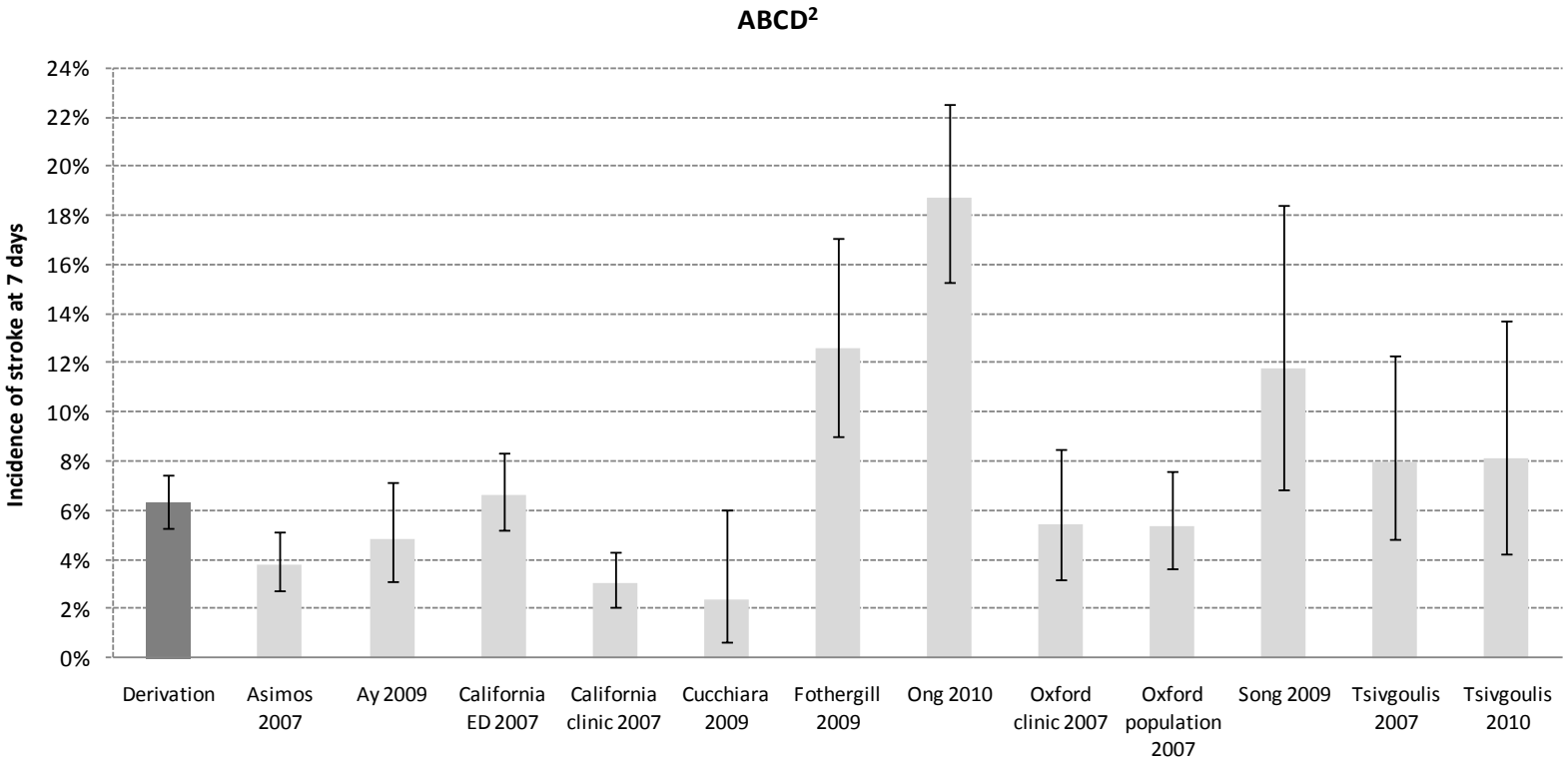


Fig. A1.4 Incidence rate of stroke at 7 days across the various studies of the ABCD^{2}rule

Table A1.1 Stroke incidence at 7 days across the derivation and validation studies of ABCD^2 rule

Study	Total % (N)		Low risk (0-3 score) % (n)		Intermediate risk (4-5 score) % (n)		High risk (6-7 score) % (n)	
CED & OPB (derivation)*	6.28	(1910)	1.35	(520)	6.51	(921)	11.3	(469)
Asimos 2007	3.80	(1054)	0	(231)	3.76	(559)	7.20	(264)
Ay 2009	4.82	(477)	1.23	(162)	6.17	(227)	7.95	(88)
California ED*	6.64	(1069)	3.09	(259)	5.93	(506)	10.86	(304)
California Clinic data*	3.01	(962)	0.47	(426)	4.28	(397)	7.19	(139)
Cucchiara 2009	2.40	(167)	1.64	(61)	1.18	(85)	9.52	(21)
Fothergill 2009	12.63	(285)	5.63	(71)	13.33	(150)	18.75	(64)
Ong 2010	18.72	(470)	8.28	(145)	19.47	(226)	32.32	(99)
Oxford Clinic data*	5.40	(315)	0.62	(162)	9.24	(119)	14.71	(34)
Oxford Population data*	5.34	(543)	0.77	(261)	4.63	(216)	25.76	(66)
Song 2009	11.76	(136)	2.78	(72)	20.69	(58)	33.33	(6)
Tsivgoulis 2007	7.96	(226)	1.19	(84)	9.73	(113)	20.69	(29)
Tsivgoulis 2010	8.11	(148)	2.82	(71)	8.93	(56)	23.81	(21)

*Note: Databases form the derivation/validation study by Johnston (2007)

APPENDIX 2

A2.1 Simplified approach to derive the predicted values of the target outcome or disease (30-day mortality of pneumonia) predicted by CRB-65 rule

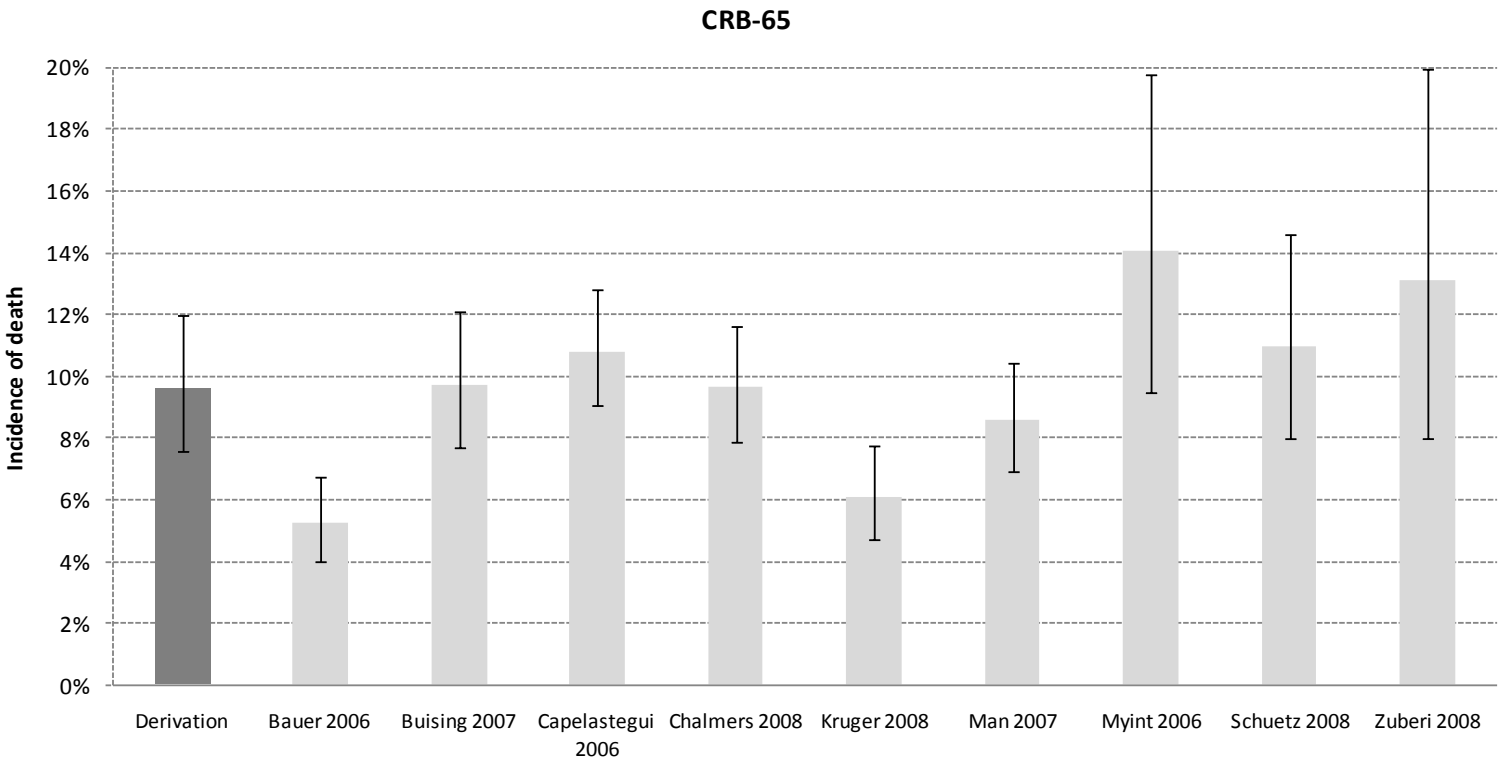


Fig. A2.1 Mortality rate across the various studies of CRB-65

Table A2.1A Mortality rate (at 30-days) across the derivation and validation studies of CRB65 rule

Study	Total % (N)		Low risk (0 score) % (n)		Intermediate risk (1-2 score) % (n)		High risk (3-4 score) % (n)	
Lim 2003	9.61	(718)	1.20	(167)	8.13	(455)	31.25	(96)
Bauer 2006	5.26	(1084)	0	(299)	6.26	(735)	22.00	(50)
Buising 2007	9.73	(740)	0	(100)	4.85	(433)	24.64	(207)
Capelastegui 2006	10.82	(1100)	0	(201)	10.13	(819)	45.00	(80)
Chalmers 2008	9.63	(1007)	1.84	(217)	7.47	(629)	28.57	(161)
Kruger 2008	6.10	(1032)	0.76	(262)	6.97	(703)	17.91	(67)
Man 2007	8.56	(1016)	2.34	(128)	7.41	(783)	24.76	(105)
Myint 2006	14.06	(192)	0	(3)	10.45	(134)	23.64	(55)
Schuetz 2008	10.99	(373)	3.67	(109)	13.51	(259)	40.00	(5)
Zuberi 2008	13.14	(137)	0	(34)	13.83	(94)	55.56	(9)

The distribution pattern in the initial derivation study of the CPR is used as a "predictive model" to which all validation studies are related. The number of deaths as target outcome as predicted by the **CRB-65 severity index** (rule) is compared to the observed number of deaths in each of the validation studies, across the three risk strata as defined by the CPR (low risk: score = 0; intermediate risk: score = 1 or 2; and high risk: score = 3 or 4). In order to calculate the predicted number of deaths, the proportionate mortality estimate (%) from the original derivation study was applied according to the three risk strata: low risk (mortality=1.2%), intermediate risk (mortality=8.2%) and high risk (mortality=31.3%). The calculation of the absolute risk (mortality or incidence of death) using the distribution in the derivation study [1] as a predictive model is shown on Table A2.2A below.

Table A2.1B Results on derivation data for CRB-65 rule

CRB-65 risk stratification	N	Deaths observed (n)	Deaths observed (%)
Low risk (0 score)	167	2	2/167 = 1.20%
Intermediate risk (1-2 score)	455	37	37/455 = 8.10%
High risk (3-4 score)	96	30	30/96 = 31.3%

We show an example on data from a validation study by Man et al (2007).[2] - the blue arrow indicates how the estimate of observed deaths (percentage) is used to be applied to the data from the validation study to produce the predicted number of deaths and to compare those to the observed deaths as in the last column of Table A2.2B.

Table A2.1C Results on validation data for CRB-65 rule

CRB-65 risk stratification	N	Deaths predicted (%)*	Deaths predicted (n)	Deaths observed (n)**
Low risk (0 score)	128	1.2%	1.5 (≈2)	3
Intermediate risk (1-2 score)	783	8.1%	63.4 (≈63)	58
High risk (3-4 score)	105	31.3%	32.9 (≈33)	26

*Notes: *Using original derivation study as a predictive model; ** actual number of deaths reported in each stratum of risk; values in the parentheses are rounded numbers.*

[1] *Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax 2003; 58(5): 377–382.*
[2] *Man SY, Lee N, IpM, et al. Prospective comparison of three predictive rules for assessing severity of community-acquired pneumonia in Hong Kong. Thorax 2007; 62(4): 348–353.*

A2.2 Predicted values from coefficients of the logistic regression model in the derivation study of the CRB-65 rule

In order to confirm our new simplified approach, we generated "individual-level data" by creating dummy variables of the CRB-65 rule (considering those as a *proxy* for the original data given the distribution in the derivation study) and applied a logistic regression model to predict the outcome. From the derivation study it was possible to reconstruct the data with the outcome "death" as a dependent, binary variable (yes=1; no=0) and the CRB-65 score as a single, independent semi-quantitative variable. In fact, the CRB-65 score is not a true continuous variable and thus we avoided an assumption of linearity by including CRB-65 score in the models as a discrete, categorical variable, with the lowest CRB-65 stratum (score 0) as the reference category. This was achieved by converting the CRB-65 score into two dummy dichotomous variables *(one for the intermediate risk at score 1 or 2; the other – for high risk at score of 3 or 4)* and include these into a multiple linear regression model as two separate covariates (predictors). An extract of the data transformation is shown below in Table A2.2A.

A2.2A Reconstruction of the individual-level dataset from the derivation study (dummy variables)

Table A2.2A Creation of dummy variables for the CRB-65 score by using derivation study data

CRB-65	Deaths	N
0	2	167
1	14	266
2	23	189
3	28	85
4	2	11

Patient	Death	CRB-65	Dummy Intermediate (X_{INT})	Dummy High (X_{HIGH})
1	1	0	0	0
2	1	0	0	0
…	…	…	…	…
167	0	0	0	0
168	1	1	1	0
169	1	1	1	0
…	…	…	…	…
433	0	1	1	0
434	1	2	1	0
435	1	2	1	0
…	…	…	…	…
622	0	2	1	0
623	1	3	0	1
624	1	3	0	1
…	…	…	…	…
707	0	3	0	1
708	1	4	0	1
709	1	4	0	1
…	…	…	…	…
718	0	4	0	1

The equation {A1} of the prediction logistic model is of the form:

$$\text{Log}\left(\frac{\text{risk of death}}{1-\text{risk of death}}\right) = \text{linear predictor Y} = \alpha + \beta_{INT} X_{INT} + \beta_{HIGH} X_{HIGH} \qquad \{A1\},$$

where α is the coefficient of the intercept and β_{INT} and β_{HIGH} are the regression coefficients of the dummy variable intermediate and high risk, respectively. The predicted probability of death for each individual patient can be calculated from equation {A2}:

$$P = \frac{e^{Y}}{(1+e^{Y})} \qquad \{A2\}.$$

In the derivation study, the linear model was described by equation [3]:

$$Y = -4.413 + 1.988\, X_{INT} + 3.624\, X_{HIGH} \qquad \{A3\}.$$

For instance, a patient with score 0 will have Y = -4.413, resulting in a probability of 0.012 (1.2%, 95% CI 0.3-4.7%). A patient with a score of 1-2 will produce Y = -4.413+1.988 = -2.425 that equals to a probability of 0.081 (8.1%, 95% CI 5.9-11.0%). Similarly, a patient with a score of 3-4 will have Y = -4.413+3.624 = -0.789, resulting in a probability of 0.313 (31.3%, 95% CI 22.8-41.2%).

Using the above logistic regression approach as shown on the data from the derivation study (Table A2.2A) it was possible to approximate the individual-level data in each of the validation studies of the CRB-65 rule. As an example only, we present below one such computation as applied to the validation study of Man et al (2007) as in Table A2.2B below.

A2.2B Reconstruction of the individual-level dataset from the validation study (dummy variables)

Table A2.2B Creation of dummy variables for the CRB-65 rule by using validation study data

CRB-65	Deaths	N
0	3	128
1	25	489
2	33	294
3	22	95
4	4	10

Patient	Death	CRB-65	Dummy Intermediate ($X_{INT,VAL}$)	Dummy High ($X_{HIGH,VAL}$)
1	1	0	0	0
2	1	0	0	0
...	...	...	...	...
128	0	0	0	0
129	1	1	1	0
130	1	1	1	0
...	...	...	...	...
617	0	1	1	0
618	1	2	1	0
619	1	2	1	0
...	...	...	...	...
911	0	2	1	0
912	1	3	0	1
913	1	3	0	1
...	...	...	...	...
1006	0	3	0	1
1007	1	4	0	1
1008	1	4	0	1
...	...	...	...	...
1016	0	4	0	1

The linear predictor Y_{VAL} on the data from the validation study was calculated after the equation [4]:

$$Y_{VAL} = \alpha_{DER} + \beta_{INT,DER} X_{INT,VAL} + \beta_{HIGH,DER} X_{HIGH,VAL} \quad \{A4\},$$

where α_{DER} (-4.413) is the intercept as well as $\beta_{INT,DER}$ (1.988) and $\beta_{HIGH,DER}$ (3.624) are the dummy variable coefficients from the derivation study (as shown above), while $X_{INT,VAL}$ and $X_{HIGH,VAL}$ are the CRB-65 values as dummy variables of intermediate and high risk, respectively, from the validation study. The predicted probability of death for each individual patient (Table A2.2C) was calculated as:

$$P = \frac{e^{Y_{VAL}}}{(1+e^{Y_{VAL}})} \quad \{A5\}.$$

Table A2.2C Reconstructed individual-level dataset from the validation study of the CRB-65 rule by Man et al (2007)

Patient	Death	CRB-65	Dummy Intermediate ($X_{INT,VAL}$)	Dummy High ($X_{HIGH,VAL}$)	Y_{VAL}	P
1	1	0	0	0	-4.413	0.0119
2	1	0	0	0	-4.413	0.0119
…	…	…	…	…	…	…
128	0	0	0	0	-4.413	0.0119
129	1	1	1	0	-2.425	0.0813
130	1	1	1	0	-2.425	0.0813
…	…	…	…	…	…	…
617	0	1	1	0	-2.425	0.0813
618	1	2	1	0	-2.425	0.0813
619	1	2	1	0	-2.425	0.0813
…	…	…	…	…	…	…
911	0	2	1	0	-2.425	0.0813
912	1	3	0	1	-0.789	0.313
913	1	3	0	1	-0.789	0.313
…	…	…	…	…	…	…
1006	0	3	0	1	-0.789	0.3125
1007	1	4	0	1	-0.789	0.3125
1008	1	4	0	1	-0.789	0.3125
…	…	…	…	…	…	…
1016	0	4	0	1	-0.789	0.3125

The predicted number of deaths was obtained by adding up the individual probabilities within each stratum of the CRB-65 rule and then compared with the observed number of deaths (Table A2.2D).

Table A2.2D. Results on the data from a validation study of the CRB-65 rule by Man et al (2007)

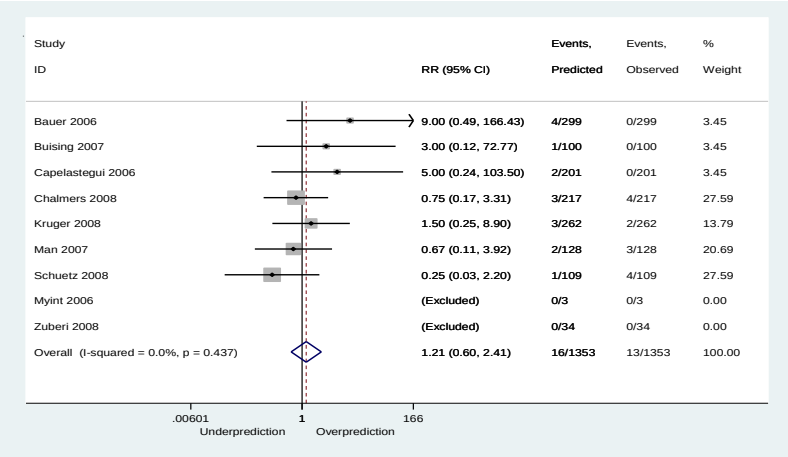
CRB-65 risk stratification	Observed deaths (n)*	N	Predicted deaths (n)**
Low risk (0 score)	3	128	1.5 (≈2)
Intermediate risk (1-2 score)	58	783	63.7 (≈64)
High risk (3-4 score)	26	105	32.8 (≈33)

*Notes: *Actual number of deaths reported in each stratum of risk; **values in the parentheses are rounded numbers.*

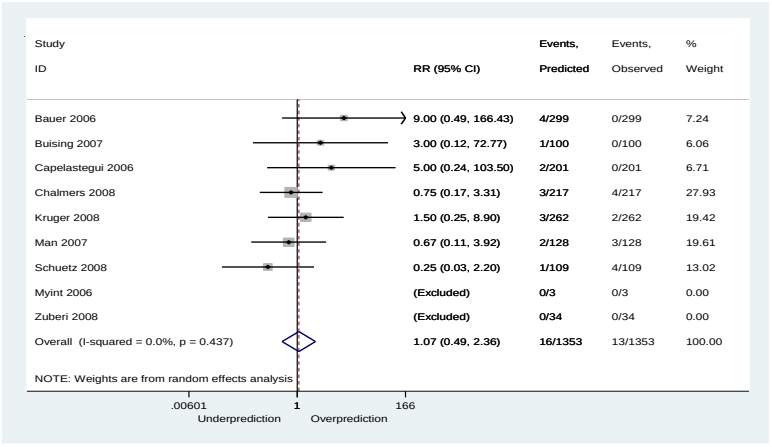
It is clear that the predicted numbers of deaths in Table A2.2D as obtained by our new simplified approach in each risk stratum of the CRB-65 rule (2, 63 and 33) in the validation study by Man et al (2007) do correspond exactly to the numbers of deaths in each stratum of Table A2.1C (2, 64 and 33) as predicted by the logistic regression model with the reconstructed dummy variables *(as considered to be our "gold standard" as a reference for our new approach).*

A2.3. Results of the meta analysis of the validation studies of CRB-65 rule (Forest plots)

A2.3.1A) No adjustment (the original prediction rule) – fixed effects



A2.3.1B) No adjustment (the original prediction rule) – random effects



A2.3.1C) Adjustment of the intercept – fixed effects

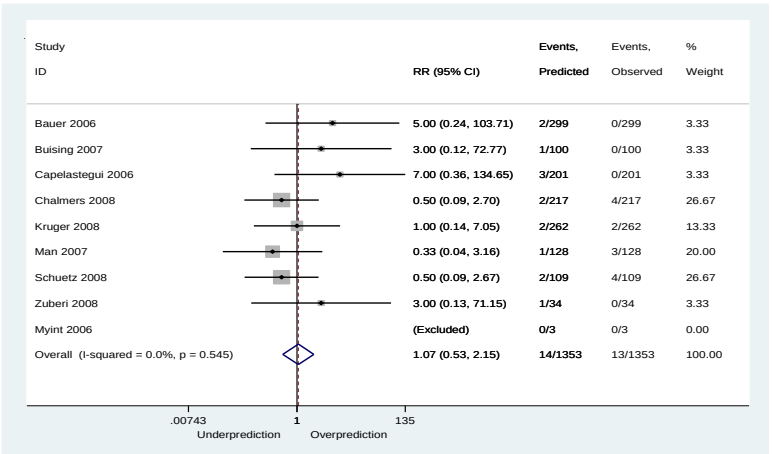
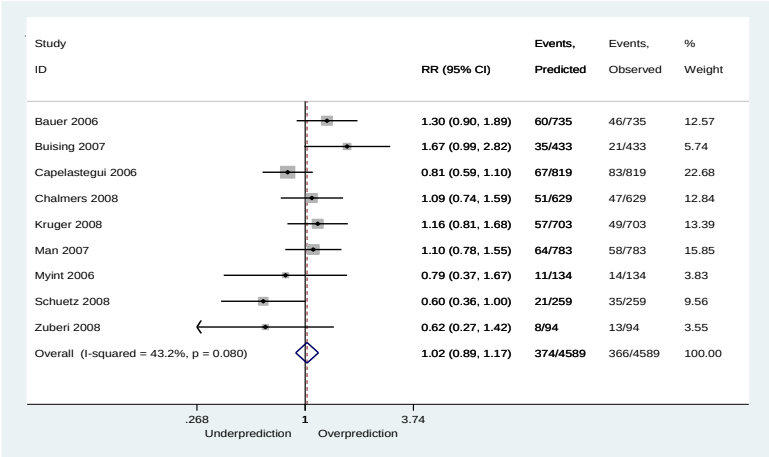
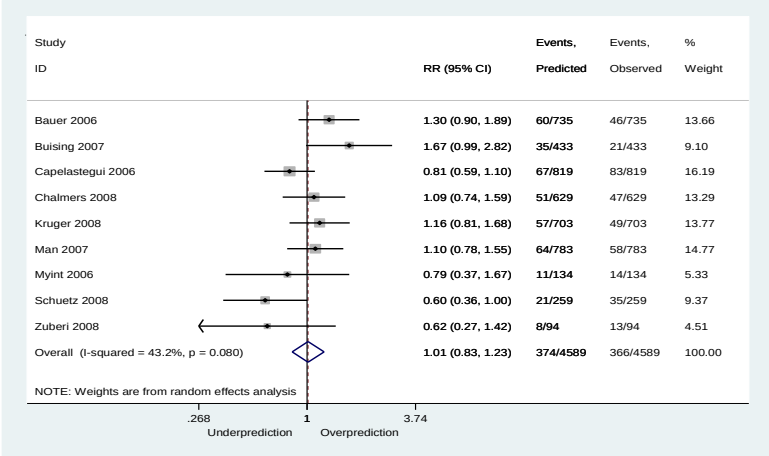


Fig. A2.3.1 CRB-65 (low risk)

A2.3.2A) No adjustment (the original prediction rule) – fixed effects



A2.3.2B) No adjustment (the original prediction rule) – random effects



A2.3.2C) Adjustment of the intercept – fixed effects

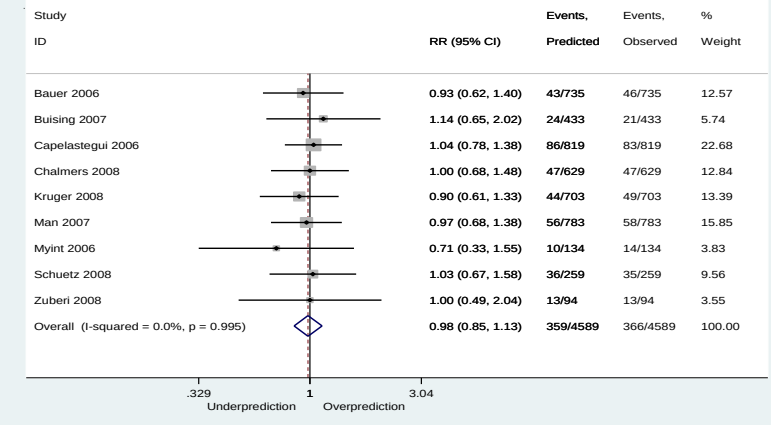
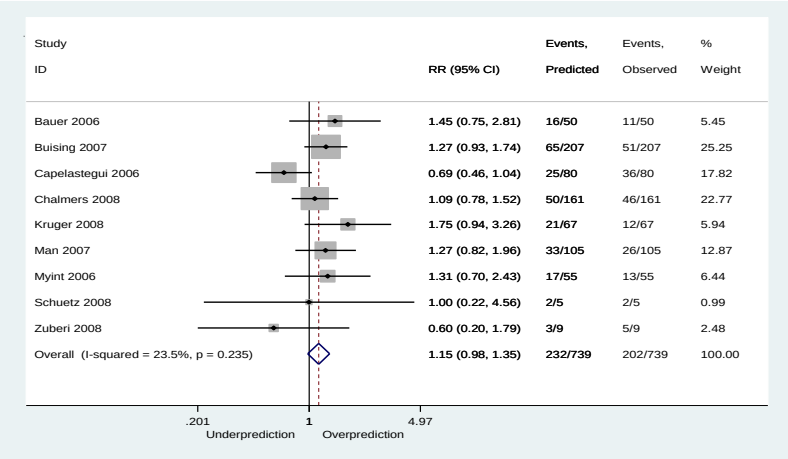
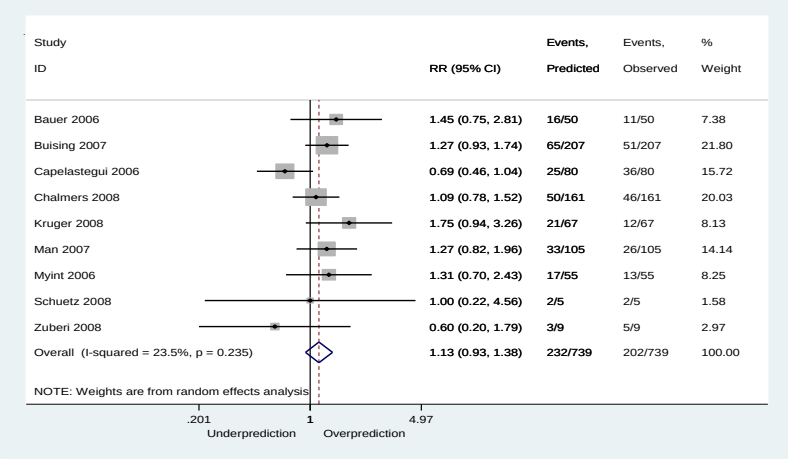


Fig. A2.3.2 CRB-65 (intermediate risk)

A2.3.3A) No adjustment (the original prediction rule) – fixed effects



A2.3.3B) No adjustment (the original prediction rule) – random effects



A2.3.3C) Adjustment of the intercept – fixed effects

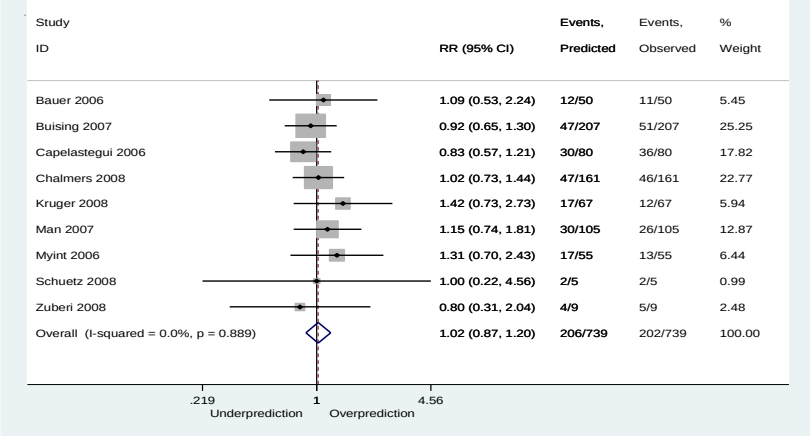


Figure A2.3.3 CRB-65 (high risk)

A2.4. Performance and pooled results of the meta analysis of the validation studies of CRB-65 rule

Table A2.4. Performance of the CRB65 rule in validation studies

Study	Discrimination (*c*-statistic)		Calibration (H-L p-value*)	
	No adjustment (original CPR)	Adjustment of intercept	No adjustment (original CPR)	Adjustment of intercept
Bauer 2006	0.695	0.695	0.028	0.404
Buising 2007	0.759	0.759	0.009	0.691
Capelastegui 2006	0.700	0.700	0.003	0.156
Chalmers 2008	0.717	0.717	0.644	0.762
Kruger 2008	0.661	0.661	0.065	0.437
Man 2007	0.638	0.638	0.258	0.417
Myint 2006	0.618	0.618	0.479	0.475
Schuetz 2008	0.622	0.622	0.001	0.743
Zuberi 2008	0.725	0.725	0.073	0.771

Note: *H-L, Hosmer-Lemeshow "goodness-of-fit" p-value *(a non-significant p-value means good fit – the higher the p-value, the better the fit)*

Table A2.5. Pooled RRs with 95% CIs from the meta-analysis of the validation studies of CRB65 rule*

		No adjustment (original CPR)			Adjustment of intercept
CRB-65 score (mortality risk)	I^2	Fixed effects	Random effects	I^2	Fixed effects
Low risk (0 score)	0.0%	1.21 (0.60-2.41)	1.08 (0.49-2.36)	0.0%	1.07 (0.53-2.15)
Intermediate risk (1-2 score)	43.2%	1.02 (0.89-1.17)	1.01 (0.83-1.23)	0.0%	0.98 (0.85-1.13)
High risk (3-4 score)	23.5%	1.15 (0.98-1.35)	1.13 (0.93-1.38)	0.0%	1.02 (0.87-1.20)

Note: *Abbreviations: RR, risk ratio, CI, confidence interval; I^2, coefficient of heterogeneity