

1 **External validation of prognostic models to predict stillbirth using the International**
2 **Prediction of Pregnancy Complications (IPPIC) Network database: An individual**
3 **participant data meta- analysis**
4

5 John Allotey*^{#1,2}, Rebecca Whittle*³, Kym IE Snell³, Melanie Smuk⁴, Rosemary Townsend⁵,
6 Peter von Dadelszen⁶, Alexander EP Heazell⁷, Laura Magee⁶, Gordon C.S. Smith⁸, Jane
7 Sandall^{6,9}, Basky Thilaganathan⁵, Javier Zamora^{1,10,11}, Richard D. Riley³, Asma Khalil⁵, Shakila
8 Thangaratinam^{1,12} for the IPPIC Collaborative Network⁺
9

10 ¹WHO Collaborating Centre for Global Women's Health, Institute of Metabolism and Systems
11 Research, University of Birmingham, Birmingham, UK

12 ²Institute of Applied Health Research, University of Birmingham, Birmingham, UK

13 ³Centre for Prognosis Research, School of Medicine, Keele University, Keele, UK

14 ⁴Medical Statistics Department, London School of Hygiene and Tropical Medicine, London, UK

15 ⁵ Fetal Medicine Unit, St George's University Hospitals NHS Foundation Trust and Molecular
16 and Clinical Sciences Research Institute, St George's University of London, London, UK

17 ⁶Department of Women and Children's Health, School of Life Course Sciences, King's College
18 London, London, UK

19 ⁷Maternal and Fetal Health Research Centre, School of Medical Sciences, Faculty of Biology,
20 Medicine and Health, University of Manchester, UK

21 ⁸ Department of Obstetrics and Gynaecology, NIHR Biomedical Research Centre, Cambridge
22 University, UK

23 ⁹Health Service and Population Research Department, Centre for Implementation Science,
24 Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

25 ¹⁰Clinical Biostatistics Unit, Hospital Universitario Ramón y Cajal (IRYCIS), Madrid, Spain

26 ¹¹CIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain

27 ¹²Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK

28

29 * Joint first authors

30 #Corresponding author

31

32 **ABSTRACT**

33 **Objective**

34 Stillbirth is a potentially preventable complication of pregnancy. Identifying women at risk can
35 guide decisions on closer surveillance or timing of birth to prevent fetal death. Prognostic
36 models have been developed to predict the risk of stillbirth, but none have yet been externally
37 validated. We externally validated published prediction models for stillbirth using individual
38 participant data (IPD) meta-analysis to assess their predictive performance.

39

40 **Methods**

41 We searched Medline, EMBASE, DH-DATA and AMED databases from inception to
42 December 2020 to identify stillbirth prediction models. We included studies that developed or
43 updated prediction models for stillbirth for use at any time during pregnancy. IPD from cohorts
44 within the International Prediction of Pregnancy Complication (IPPIC) Network were used to
45 externally validate the identified prediction models whose individual variables were available in
46 the IPD. We assessed the risk of bias of the models and IPD using PROBAST, and reported
47 discriminative performance using the C-statistic, and calibration performance using calibration
48 plots, calibration slope and calibration-in-the-large. We estimated performance measures
49 separately in each study, and then summarised across studies using random-effects meta-
50 analysis. Clinical utility was assessed using net benefit.

51

52 **Results**

53 We identified 17 studies reporting the development of 40 prognostic models for stillbirth. None
54 of the models were previously externally validated, and only a fifth (20%, 8/40) reported the full
55 model equation. We were able to validate three of these models using the IPD from 19 cohort

56 studies (491,201 pregnant women) within the IPPIC Network database. Based on evaluating
57 their development studies, all three models had an overall high risk of bias according to
58 PROBAST. In our IPD meta-analysis, the models had summary C-statistics ranging from 0.53
59 to 0.65; summary calibration slopes of 0.40 to 0.88, and generally with observed risks
60 predictions that were too extreme compared to observed risks; and little to no clinical utility as
61 assessed by net benefit. However, there remained uncertainty in performance for some models
62 due to small available sample sizes

63

64 **Conclusion**

65 The three validated models generally showed poor and uncertain predictive performance in new
66 data, with limited evidence to support their clinical application. Findings suggest
67 methodological shortcomings in their development including overfitting of models. Further
68 research is needed to further validate these and other models, identify stronger prognostic
69 factors, and to develop more robust prediction models.

70

71 **Study registration**

72 PROSPERO ID: CRD42018074788

73

74 **Keywords:** stillbirth, intra-uterine death, prediction model, individual participant data, external
75 validation

76

77 **Word count:** 376

78 **INTRODUCTION**

79 Stillbirth continues to be a major burden globally, accounting for almost two thirds of perinatal
80 mortality.^{1,2} In the UK, stillbirth rates were largely unchanged from 2000 – 2015, and at 4.2
81 stillbirths/1,000 births in 2017 had one of the highest rates in Europe.³⁻⁵ Prediction and
82 individualisation of risk remain key priorities for stillbirth research,^{6,7} because accurate
83 identification of women at risk of stillbirth can guide decisions on closer surveillance, or timing
84 of birth to prevent fetal death. A recent review that identified existing prediction models for
85 stillbirth reported that none had been externally validated.⁸ As a result, no prediction models are
86 routinely used in clinical practice and none have been recommended by any national or
87 international guidelines.

88

89 An independent, external validation and comparison of existing multivariable stillbirth
90 prediction models is important to help identify which prediction model (if any) performs best
91 and is potentially applicable in clinical practice. However, the relative rarity of this devastating
92 outcome limits rigorous investigation of existing stillbirth prediction models in single cohort
93 studies. An individual participant data (IPD) meta-analysis that combines the raw data from
94 multiple studies, has great potential for use in externally validating existing models, by
95 increasing the sample size beyond what is feasible in a single study, thereby increasing the
96 number of events observed.⁹⁻¹² It also allows us to evaluate the generalisability and
97 transportability of the predictive performance of the models across a range of clinical settings
98 being considered for their application.

99

100 We therefore set out to identify, critically appraise and externally validate existing multivariable
101 prognostic models for stillbirth prediction using IPD meta-analysis within the independent
102 International Prediction of Pregnancy Complication (IPPIC) Network database, and to assess the
103 clinical utility of the models using decision curve analysis.

104

105 **METHODS**

106 This study was based on a prospective protocol registered on PROSPERO (registration number
107 CRD42018074788), and reported in line with TRIPOD recommendations for reporting risk
108 prediction model validation studies.¹³

109

110 *Literature search and selection of prediction models for external validation using the IPPIC*
111 *network database*

112 We systematically searched Medline, EMBASE, DH-DATA and AMED databases from
113 inception to December 2020 to identify all studies that developed or updated prognostic models
114 for stillbirth for use at any time during pregnancy. We also hand searched reference lists of
115 relevant articles and systematic reviews to identify potentially eligible studies. Our search
116 included terms for stillbirth, intrauterine fetal death and perinatal mortality, and study selection
117 was done independently by two researchers. The complete search strategy is provided in
118 appendix 1.

119

120 *Stillbirth model eligibility criteria, data extraction and risk of bias assessment*

121 We included studies that reported the development or update of a multivariable model with at
122 least three variables to predict the risk of stillbirth in pregnant women and reported the model
123 equation in the publication. No attempts were made to contact authors of studies that did not
124 publish their model equation. Given the wide international variation in definitions of stillbirth,

125 we accepted the authors' definition of stillbirth (both antepartum and intrapartum), and included
126 models developed for use at any time in pregnancy. We excluded models that: predicted
127 stillbirth as part of a composite adverse outcome; contained predictors that were not measured in
128 any of the cohorts within the IPPIC IPD; or if there were too few outcomes (<10 stillbirths)
129 reported across the IPPIC IPD cohorts with the same predictors as the model, to allow for its
130 external validation.

131

132 We extracted data on the definition of stillbirth, number of participants and events, population
133 type, predictors in the final model, and the reported model performance. Based on information
134 in the original articles, we assessed the risk of bias of included models using the Prediction
135 study Risk of Bias Assessment tool (PROBAST),¹⁴ across the four domains of participant
136 selection, predictors, outcome and analysis, and this was done independently by two researchers.
137 Disagreement were resolved through discussions with a third researcher. We classified the risk
138 of bias to be low, high or unclear for each domain, as well as an overall risk of bias. Each
139 domain included signalling questions rated as "yes", "probably yes", "probably no", "no" or "no
140 information". Domains with any signalling question rated as "probably no" or "no" were
141 considered to have potential for bias and classed as high risk. The overall risk of bias was
142 considered to be low if it scored low in all domains, high if any one domain had a high risk of
143 bias, and unclear for any other classifications.

144

145 *International Prediction of Pregnancy Complications (IPPIC) Network*

146 We identified cohorts for the IPPIC Network by systematically reviewing evidence for risk of
147 pregnancy complications including pre-eclampsia, stillbirth and fetal growth restriction (FGR),
148 and inviting research groups that had undertaken the primary studies to join the IPPIC Network

149 and share their primary IPD. We also searched major databases and repositories and contacted
150 researchers within the IPPIC Network to identify relevant studies or datasets that may have been
151 missed, including unpublished research and birth cohorts. We formatted, cleaned and
152 harmonised datasets received and assessed the quality of each cohort using the participants,
153 predictors and outcome domains of the PROBAST tool.¹⁴ Study population could vary from low
154 to high risk of development of complications. The network includes nearly 150 collaborators
155 from 26 countries, contributing IPD of over 4 million pregnancies, and contains data on
156 maternal characteristics, obstetric history, clinical assessment and tests, as well as various
157 maternal and offspring outcomes. The database is a living repository and is regularly being
158 enriched with additional studies. We consider the predictor variables contained within the IPPIC
159 Network to represent measures which are easy to obtain in a clinical setting, reflecting their
160 availability in routine practice. Methods on how cohorts within the IPPIC Network database
161 were identified and harmonised have previously been published.¹⁵⁻¹⁷

162

163 **Statistical analysis for external validation using IPPIC network database**

164 *Data harmonisation and set-up*

165 Predictors or outcomes of existing prediction models that were partially missing for <95% of
166 individuals in any cohort were multiply imputed under the missing at random assumption using
167 multiple imputation by chained equations.^{18,19} We used linear regression to impute for
168 approximately normally distributed continuous variables, logistic regression for binary
169 variables, and multinomial logistic regression for categorical variables. We carried out multiple
170 imputation for each individual cohort separately and generated fifty imputed datasets for each.
171 We also included other predictors that were available within the cohort as auxiliary variables in
172 the imputation models. Imputation checks were completed by looking at histograms, summary

173 statistics and tables of values across imputations, as well as checking trace plots for convergence
174 issues.

175

176 *External validation of models*

177 Each model was validated by applying the model equation to each participant in the cohort to
178 calculate the linear predictor for that participant (LP_i , value of the linear combination of
179 predictors in the model equation for individual i), as well as the predicted probability of
180 stillbirth (inverse logit transformation of LP_i). For each prediction model, the distribution of LP_i
181 values were summarised for each cohort, and performance statistics were calculated in each
182 imputed dataset and then averaged across imputations using Rubin's rules to obtain one estimate
183 and standard error (SE) for each performance statistic in each cohort.²⁰

184

185 The discriminatory performance of models were assessed using the C-statistic (summarised as
186 the area under receiver operating characteristic curve, where 1 indicates perfect discrimination
187 and 0.5 indicates no discrimination beyond chance), and calibration statistics of the calibration
188 slope (slope of the regression line fitted between predicted and observed risk probabilities on the
189 logit scale, with 1 being the ideal value), and calibration-in-the-large (the extent that model
190 predictions are systematically too low or too high across the cohort, ideal value of 0).^{21 22} Model
191 calibration was also visually assessed using calibration plots representing the average predicted
192 probability for risk groups categorised using deciles of predicted probability against the
193 observed proportion in each group, in cohorts with at least 100 events. A lowess smoother curve
194 was applied to show calibration across the entire range of predicted probabilities at the
195 individual-level (i.e. without categorisation). For the calibration plots, average predicted

196 probabilities were obtained for individuals by pooling their linear predictor values across
197 imputed datasets using Rubin's rules, and then transforming to the probability scale.

198

199 Performance measures of prediction models that were validated in more than two independent
200 cohorts were summarised using a random effects meta-analysis to calculate a summary estimate
201 for the model's discrimination and calibration performance. Model performance was
202 summarised for each statistic as the average and 95% confidence interval (CI) calculated using
203 the Hartung-Knapp-Sidik-Jonkman approach.^{23,24} Between-study heterogeneity (τ^2) and the
204 proportion of variability due to between-study heterogeneity (I^2)²⁵ were summarised. We also
205 reported the approximate 95% prediction intervals, for potential predictive performance in a
206 new study, as calculated using the approach of Higgins et al.²⁶

207

208 *Decision curve analysis*

209 We performed decision curve analysis (DCA) to assess the clinical value of the models on
210 cohorts with at least 100 events. This analysis allowed us to determine the net benefit of the
211 models across a range of clinically plausible threshold probabilities (which included any values
212 up to 0.1, given the generally very low risk of stillbirth), compared to either simply classifying
213 all women as having the outcome or no women as having the outcome.²⁷ The strategy with the
214 highest net benefit at a particular threshold has the highest clinical value.²⁸ The net benefit is
215 represented as a function of the decision threshold in decision curve plots.

216 All statistical analyses were performed using Stata software version 15.

217

218

219

220 **RESULTS**

221 From 5055 citations we identified 17 articles describing the development of 40 stillbirth
222 prediction models published between 2007 and 2020 (Appendix 2). Three studies reporting
223 three prediction models - Smith 2007,²⁹ Yerlikaya 2016,³⁰ and Trudell 2017³¹ met our inclusion
224 criteria for external validation in the IPPIC IPD datasets (Figure 1).

225

226 **Characteristics of included models**

227 All three models were developed using binary logistic regression in unselected populations of
228 pregnant women,²⁹⁻³¹ and the definition of stillbirth varied between the studies. Two models
229 included only maternal clinical characteristics as predictors,^{30,31} while one model additionally
230 included ultrasound markers.²⁹ Only one study had at least 10 events per predictor for model
231 development,³⁰ the others did not justify whether their sample size was sufficient. Using the
232 PROBAST tool, the overall risk of bias for all three models was high, with all models assessed
233 as being at high risk of bias in the analysis domain. The characteristics of included studies and
234 models are described in Table 1.

235

236 **Characteristics of the IPPIC validation cohorts**

237 Of the 78 cohorts in the IPPIC data repository, 19 cohorts (24%) contained relevant data that
238 could be used to externally validate at least one of the three prediction models identified. Only
239 women with singleton pregnancies in the cohorts were used for external validation. The
240 prevalence of stillbirth ≥ 24 weeks gestation in the cohorts ranged from 0.1% - 1.6%. A quarter
241 of the studies used for external validation included only low risk (26%, 5/19) women, while a
242 fifth (21%, 4/19) included only high-risk women in the cohorts. Seventy-five percent (14/19) of
243 the cohorts used for external validation had an overall low risk of bias as assessed by
244 PROBAST, 21% (4/19) were assessed as high risk and one cohort as unclear (appendix 3).

245 Summary maternal characteristics and outcomes of women in the validation cohort are provided
246 in table 2, and a summary of missing data for each predictor and outcome is provided in
247 appendix 4.

248

249 **External validation and meta-analysis of predictive performance**

250 The Smith 2007 model²⁹ was validated in 3 cohorts, Yerlikaya 2016 model³⁰ in 4 cohorts and
251 the Trudell 2017 model³¹ in 17 cohorts. Two of the cohorts used to validate the Smith 2007
252 model and all four of the cohorts used to validate the Yerlikaya 2016 model were also used to
253 validate the Trudell 2017 model. A direct comparison of performance of the prediction models
254 was not possible due to differences in outcomes of each model. The distribution of the linear
255 predictor and predicted probability for each model and validation cohort are shown in appendix
256 5.

257

258 *Model predictive performance*

259 The C-statistics of models in the different validation cohorts ranged from 0.56-0.82 in the Smith
260 2007 model, 0.54-0.73 in the Yerlikaya 2016 model and 0.34-0.69 in the Trudell 2017 model
261 (Table 3). The Trudell 2017 model had the lowest overall discrimination across the validation
262 cohorts. Summary C-statistics of the models were 0.65 (95% CI 0.53 to 0.75) for the Smith
263 2007 model, 0.61 (95% CI 0.43 to 0.77) for the Yerlikaya 2016 model, and 0.53 (95% CI 0.51
264 to 0.55) for the Trudell 2017 model (Table 4). Confidence intervals for the Smith 2007 and
265 Yerlikaya 2016 models were wide, due to the fewer number of cohorts available for their
266 validation.

267

268 Calibration statistics for each model in the different validation cohorts are shown in Table 3.

269 Summary calibration slopes were < 1 for all models, indicative of overfitting during model

270 development; in particular, the 95% confidence intervals for the calibration slope were all below
271 1 for the Yerlikaya 2016 and Trudell 2017 models, indicating extreme predictions compared to
272 what was observed (Table 4).

273 Each of the three models were validated in one cohort with at least 100 events. The average
274 calibration plots showed miscalibration of the predicted risk of stillbirth in all three models
275 (Figure 2). However, predicted probabilities were all less than 0.02, therefore absolute risk
276 differences remain small. The 95% CI was wide for the calibration slope of the Smith 2007
277 model, due to less data on stillbirth outcome in the validation cohorts available for this model,
278 and so further research is needed for this model.

279

280 **Net benefit of model use**

281 The DCA for all three models in cohorts with at least 100 events, showed little or no
282 improvement in the net benefit at any probability threshold compared to a treat all or treat none
283 strategy (Figure 3).

284

285 **DISCUSSION**

286 **Summary of findings**

287 Only a fifth of published stillbirth prognostic models reported the model equation required for
288 independent external validation. Three models developed in high-income countries could be
289 externally validated using cohorts from the IPPIC data repository. The models were mostly
290 developed using maternal clinical characteristics, but one model additionally included
291 ultrasound markers. PROBAST of the original model development articles suggested risk of
292 bias concerns, and our IPD meta-analysis of model performance showed low discriminatory
293 ability and poor calibration, with calibration slopes mostly <1 , indicative of overfitting during
294 model development. The models had no clinical utility as assessed by DCA. Although each of

295 the three models could be validated in at least one cohort with >100 events, confidence intervals
296 of predictive performance were wide for the Smith 2007 model, suggesting further validation is
297 needed for this model.

298

299 **Strengths and limitations**

300 To our knowledge, this is the first systematic review and external validation study of stillbirth
301 prediction models.^{8,32} Our study with its large sample size, allowed for the evaluation of the
302 predictive performance of each model across multiple cohorts, as well as the overall
303 performance through an IPD meta-analysis. We used multiple imputation of predictors and
304 outcomes for each cohort separately, to avoid loss of useful information, and ensure we did not
305 mask any heterogeneity across cohorts.^{20,33} Although the definition of stillbirth in the validation
306 cohorts were standardised, stillbirth was defined differently in each model, which prevented a
307 head-to-head comparison of model performance.

308

309 Our study has some limitations. We were only able to validate three of the 40 identified models ,
310 mainly due to the failure of studies to adhere to reporting standards of publishing the model
311 equation.^{34,35} Only two models were published before release of TRIPOD. Some cohorts used in
312 the external validation had few observed cases of stillbirths, and only two had more than 100
313 events. Predicted probabilities in the cohorts only went up to 3%, which makes it difficult for
314 the models to discriminate between women who had and did not have the outcome. This further
315 highlights the primary limitation of stillbirth research, which is the comparative rarity of the
316 outcome.

317

318 **Comparison to existing studies**

319 External validation of prediction models are needed to confirm generalisability and
320 transportability of a model in populations with different characteristics.³⁶ However, independent
321 data with sufficiently large sample sizes of stillbirth and relevant predictors for external
322 validation of models are not readily available. This is a factor on why none of the published
323 models have been recommended for use in clinical practice.³⁵ Our meta-analysis obtained lower
324 summary estimates for discrimination to that reported in the development datasets, although this
325 might be due to chance as some confidence intervals were wide (e.g. Smith 2007), further
326 research is recommended.²⁹⁻³¹ Some published stillbirth models report discrimination of >
327 0.8,^{37,38} but these studies either did not report the model equation needed for independent
328 external validation,³⁸ or did not provide enough information on predictors.³⁷ In most cases, the
329 performance of a prediction model is often overestimated when only estimated in the dataset
330 used to develop the model, especially when there are few outcomes relative to the number of
331 predictors considered.^{39,40} Our study highlighted several methodological shortcomings in the
332 development of stillbirth prediction models, which is further reflected in the risk of bias
333 assessment of the models.

334

335 **Relevance to clinical care**

336 The UK Government and NHS launched a care initiative in a bid to halve stillbirth rates by
337 2025, which includes risk assessment as part of a wider care-bundle.⁴¹ The bundle does not
338 include tools to help determine if a woman is at increased risk of stillbirth, instead individual
339 factors have been identified to categorise women as low, moderate or high risk of FGR, the most
340 frequent cause of stillbirth in the UK. An accurate tool to predict which woman is at increased
341 risk of stillbirth would allow for personalised risk stratification in pregnancy, and enable
342 clinicians to make decisions on closer surveillance, or timing of birth to prevent fetal death. It

343 would also empower mothers to make informed decisions on their risk of stillbirth. This would
344 be a more targeted approach than the currently used system of a generalised population level
345 risk factor to identify women at risk of stillbirth. However, none of the models validated in this
346 study had sufficient performance or clinical utility to be recommended for use in practice.

347

348 **Recommendations for further research**

349 Stillbirth prediction models that can be used in routine care would be especially valuable in low-
350 and-middle-income countries, where stillbirth burden is disproportionately high. Models we
351 were unable to externally validate will need to be independently validated before they can be
352 recommended for use. Apart from improvement in the model development process to reduce
353 overfitting by using larger sample sizes and adjusting for optimism of the predictor effects (e.g.
354 by post-estimation shrinkage or penalising the model coefficients), additional work is needed to
355 identify novel prognostic factors for use in model development, to improve the discriminatory
356 performance of prediction models.⁴² A closer examination of existing stillbirth risk factors could
357 potentially enable us to abandon inaccurate risk predictors and focus clinical care and research
358 on the highest value predictors.

359 Systematic reviews using aggregate data meta-analysis, currently represent the best available
360 evidence on predictors of stillbirth, and have proposed several risk factors to categorise women
361 as high-risk.⁴³ However, these studies are limited by heterogeneity in the data reported within
362 the primary studies, such as in the definition of stillbirth.⁴³ Existing primary studies are often
363 small with imprecise estimates, and inconsistencies in confounding factors adjusted for in their
364 analysis, which sometimes leads to contradictory factor-outcome associations. Large cohorts are
365 needed to collect richer data on risk factors to enable development and validation of prediction
366 models.

367

368 Whilst this study has explored validation of different stillbirth prediction models, stillbirth is the
369 final endpoint of several heterogeneous antecedent pathways, with varying biological
370 mechanisms involved (for example, those involving FGR, and those secondary to diabetes,
371 typically with a large for gestational age infant). It is possible that more than one model will be
372 needed, either for prediction at different gestational ages, or for stillbirths with similar
373 phenotypes.

374

375

376 **CONCLUSION**

377 This is a comprehensive assessment and independent external validation of published stillbirth
378 prognostic models across multiple cohorts. Findings suggest methodological shortcomings
379 including overfitting of models during development. None of the three previously published
380 stillbirth models that were validated in this study showed sufficient performance or clinical
381 utility to be recommended for use in practice. Although there were differences in predictor and
382 outcome definitions used for the different models, all three models considered similar candidate
383 predictors for model development, which may suggest additional and better predictors
384 (prognostic factors) of stillbirth still need to be identified.

385 **Abbreviations**

386	IPD	Individual participant data
387	IPPIC	International Prediction of Pregnancy Complications
388	PROBAST	Prediction study Risk of Bias Assessment
389	SE	Standard error
390	CI	Confidence interval
391	LP	Linear predictor

392

393 **Declarations**

394 **Ethics approval and consent to participate**

395 Not applicable. The study involved secondary analysis of existing anonymised data.

396

397 **Consent for publication**

398 Not applicable

399

400 **Availability of data and materials**

401 The data that support the findings of this study are available from the IPPIC data sharing
402 committee, but restrictions apply to the availability of these data, which were used under license
403 for the current study, and so are not publicly available. Data are however available from the
404 authors upon reasonable request and with permission of contributing collaborators.

405

406 **Competing interests**

407 None to declare

408

409

410 **Funding**

411 The IPPIC data repository was set up by funding from the National Institute for Health Research
412 Health Technology Assessment Programme (Ref no: 14/158/02). This project was funded by
413 Sands charity. Kym Snell is funded by the National Institute for Health Research School for
414 Primary Care Research (NIHR SPCR Launching Fellowship).

415 The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University
416 of Bristol provide core support for ALSPAC. This publication is the work of the authors and JA,
417 ST, RR, and RW will serve as guarantors for the contents of this paper.

418
419 **Authors' contributions**

420 ST, AK developed the protocol. RW wrote the statistical analysis plan and performed the
421 analysis, JA produced the first draft of the article and revised the article. RR and KS oversaw
422 the statistical analyses and analysis plan. MS and JA formatted, harmonised and cleaned IPPIC
423 datasets, in preparation for analysis. JA, MS mapped the variables in the datasets, and cleaned
424 and quality checked the data. JA, ST, MS and RT undertook the literature searches, study
425 selection, acquired Individual Participant Data, contributed to the development of all versions of
426 the manuscript and led the project. All authors provided input at all stages of the project and
427 helped revise the article.

428

429 **Acknowledgements**

430 The following are members of the IPPIC Collaborative Network⁺

431

432 Arri Coomarasamy - University of Birmingham; Alex Kwong - University of Bristol; Ary I.
433 Savitri - University Medical Center Utrecht; Kjell Åsmund Salvesen - Norwegian University of
434 Science and Technology; Sohinee Bhattacharya - University of Aberdeen; Cuno S.P.M.
435 Uiterwaal - University Medical Center Utrecht; Annetine C. Staff - University of Oslo; Louise
436 Bjoerkholt Andersen - University of Southern Denmark; Elisa Llurba Olive - Hospital
437 Universitari Vall d'Hebron; Christopher Redman - University of Oxford; Line Sletner -
438 University of Oslo; George Daskalakis - University of Athens; Maureen Macleod - University
439 of Dundee; Baskaran Thilaganathan - St George's University of London; Mali Abdollahain -
440 RMIT University; Javier Arenas Ramírez - University Hospital de Cabueñes; Jacques Massé -

441 Laval University; Asma Khalil - St George's University of London; Francois Audibert -
442 Université de Montréal; Per Minor Magnus - Norwegian Institute of Public Health; Anne Karen
443 Jenum - University of Oslo; Ahmet Baschat - Johns Hopkins University School of Medicine;
444 Akihide Ohkuchi - University School of Medicine, Shimotsuke-shi; Fionnuala M. McAuliffe -
445 University College Dublin; Jane West - University of Bristol; Lisa M. Askie - University of
446 Sydney; Fionnuala Mone - University College Dublin; Diane Farrar - Bradford Teaching
447 Hospitals; Peter A. Zimmerman - Päijät-Häme Central Hospital; Luc J.M. Smits -
448 Maastricht University Medical Centre; Catherine Riddell - Better Outcomes Registry &
449 Network (BORN); John C. Kingdom - University of Toronto; Joris van de Post - Academisch
450 Medisch Centrum; Sebastián E. Illanes - University of the Andes; Claudia Holzman - Michigan
451 State University; Sander M.J. van Kuijk - Maastricht University Medical Centre; Lionel
452 Carbillon - Assistance Publique-Hôpitaux de Paris Université; Pia M. Villa - University of
453 Helsinki and Helsinki University Hospital; Anne Eskild - University of Oslo; Lucy Chappell -
454 King's College London; Federico Prefumo - University of Brescia; Luxmi Velauthar - Queen
455 Mary University of London; Paul Seed - King's College London; Miriam van Oostwaard -
456 IJsselland Hospital; Stefan Verlohren - Charité University Medicine; Lucilla Poston - King's
457 College London; Enrico Ferrazzi - University of Milan; Christina A. Vinter - University of
458 Southern Denmark; Chie Nagata - National Center for Child Health and Development, Tokyo,
459 Japan; Mark Brown - University of New South Wales; Karlijn C. Vollebregt - Academisch
460 Medisch Centrum; Satoru Takeda - Juntendo University, Tokyo, Japan; Josje Langenveld -
461 Atrium Medisch Centrum Parkstad; Mariana Widmer - World Health Organization; Shigeru
462 Saito - University of Toyama, Toyama, Japan; Camilla Haavaldsen - Akershus University
463 Hospital; Guillermo Carroli - Centro Rosarino De Estudios Perinatales; Jørn Olsen - Aarhus
464 University; Hans Wolf - Academisch Medisch Centrum; Nelly Zavaleta - Instituto Nacional De
465 Salud; Inge Eisensee - Aarhus University; Patrizia Vergani - University of Milano-Bicocca;
466 Pisake Lumbiganon - Khon Kaen University; Maria Makrides - South Australian Health and
467 Medical Research Institute; Fabio Facchinetti - Università degli Studi di Modena e Reggio
468 Emilia; Evan Sequeira - ga Khan University; Robert Gibson - University of Adelaide; Sergio
469 Ferrazzani - Università Cattolica del Sacro Cuore; Tiziana Frusca - Università degli Studi di
470 Parma; Jane E. Norman - University of Bristol; Ernesto A. Figueiró-Filho - Mount Sinai
471 Hospital; Olav Lapaire - Universitätsspital Basel; Hannele Laivuori - University of Helsinki and
472 Helsinki University Hospital; Jacob A. Lykke - Rigshospitalet; Agustin Conde-Agudelo -
473 Eunice Kennedy Shriver National Institute of Child Health and Human Development; Alberto
474 Galindo - Universidad Complutense de Madrid; Alfred Mbah - University of South Florida; Ana

475 Pilar Betran - World Health Organisation; Ignacio Herraiz - Universidad Complutense de
476 Madrid; Lill Trogstad - Norwegian Institute of Public Health; Gordon G.S. Smith - Cambridge
477 University; Eric A.P. Steegers - University Hospital Nijmegen; Read Salim - HaEmek Medical
478 Center; Tianhua Huang - North York General Hospital; Annemarijne Adank - Erasmus Medical
479 Centre; Jun Zhang - National Institute of Child Health and Human Development; Wendy S.
480 Meschino - North York General Hospital; Joyce L Browne - University Medical Centre Utrecht;
481 Rebecca E. Allen - Queen Mary University of London; Fabricio Da Silva Costa - University of
482 São Paulo; Kerstin Klipstein-Grobusch Browne - University Medical Centre Utrecht; Caroline
483 A. Crowther - University of Adelaide; Jan Stener Jørgensen - Syddansk Universitet; Jean-
484 Claude Forest - Centre hospitalier universitaire de Québec; Alice R. Rumbold - University of
485 Adelaide; Ben W. Mol - Monash University; Yves Giguère - Laval University; Louise C. Kenny
486 - University of Liverpool; Wessel Ganzevoort - Academisch Medisch Centrum; Anthony O.
487 Odibo - University of South Florida; Jenny Myers - University of Manchester; SeonAe Yeo -
488 University of North Carolina at Chapel Hill; Francois Goffinet - Assistance publique – Hôpitaux
489 de Paris; Lesley McCowan - University of Auckland; Eva Pajkrt - Academisch Medisch
490 Centrum; Helena J. Teede - Monash University and Monash Health; Bassam G. Haddad -
491 Portland State University; Gustaaf Dekker - University of Adelaide; Emily C. Kleinrouweler -
492 Academisch Medisch Centrum; Édouard LeCarpentier - Centre Hospitalier Intercommunal
493 Creteil; Claire T. Roberts - University of Adelaide; Henk Groen - University Medical Center
494 Groningen; Ragnhild Bergene Skråstad - St Olavs Hospital; Seppo Heinonen - University of
495 Helsinki and Helsinki University Hospital; Kajantie Eero - University of Helsinki and Helsinki
496 University Hospital; Dewi Anggraini - University of Lambung Mangkurat; Athena Souka -
497 University of Athens Medical School; Jose Guilherme Cecatti - University of Campinas; Ilza
498 Monterio - University of Campinas; Athanasios Pillalis - University of Athens; Renato Souza -
499 University of Campinas; Lee Ann Hawkins - University of Calgary; Rinat Gabbay- Benziv -
500 Hillel Yaffe Medical Center; Francesca Crovetto - University of Barcelona; Francesc Figuera -
501 University of Barcelona, Laura Jorgensen - Queen Mary University of London, Julie Dodds -
502 Queen Mary University of London, Mehali Patel - Sands, stillbirth and neonatal death charity,
503 London, Amir Aviram - University of Toronto, Toronto, Ontario, Canada, Aris Papageorghiou -
504 St George's University of London, London, UK, Khalid Khan - University of Granada,
505 Granada, Spain

506
507 We would like to acknowledge all researchers who contributed data to this IPD meta-analysis,
508 including the original teams involved in the collection of the data, and participants who took

509 part in the research studies. We are extremely grateful to all the families who took part in this
510 study, the midwives for their help in recruiting them, and the whole ALSPAC team, which
511 includes interviewers, computer and laboratory technicians, clerical workers, research scientists,
512 volunteers, managers, receptionists and nurses.

513

514 **References**

- 515 1. Flenady V, Wojcieszek AM, Middleton P, Ellwood D, Erwich JJ, Coory M, Khong TY, Silver
516 RM, Smith GCS, Boyle FM, Lawn JE, Blencowe H, Leisher SH, Gross MM, Horey D, Farrales L,
517 Bloomfield F, McCowan L, Brown SJ, Joseph KS, Zeitlin J, Reinebrant HE, Ravaldi C, Vannacci A,
518 Cassidy J, Cassidy P, Farquhar C, Wallace E, Siassakos D, Heazell AEP, Storey C, Sadler L,
519 Petersen S, Frøen JF, Goldenberg RL. Stillbirths: recall to action in high-income countries. *The*
520 *Lancet* 2016; 387(10019): 691-702.
- 521 2. Flenady V, Koopmans L, Middleton P, Frøen JF, Smith GC, Gibbons K, Coory M, Gordon
522 A, Ellwood D, McIntyre HD, Fretts R, Ezzati M. Major risk factors for stillbirth in high-income
523 countries: a systematic review and meta-analysis. *The Lancet* 2011; 377(9774): 1331-40.
- 524 3. Draper ES, Gallimore ID, Kurinczuk JJ, Smith PW, Boby T, Smith LK, Manktelow BN, on
525 behalf of the MBRRACE-UK Collaboration. MBRRACE-UK Perinatal Mortality Surveillance
526 Report, UK Perinatal Deaths for Births from January to December 2016. Leicester: The Infant
527 Mortality and Morbidity Studies, Department of Health Sciences, University of Leicester. 2018.
528 .
- 529 4. Euro-Peristat Project. European Perinatal Health Report. Core indicators of the health
530 and care of pregnant women and babies in Europe in 2015. November 2018. Available
531 www.europeristat.com.
- 532 5. ONS (2018) Vital statistics in the UK: births, deaths and marriages - 2018 update, Office
533 of National Statistics, London, England
534 <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigrati..>
- 535 6. Heazell AE, Whitworth MK, Whitcombe J, Glover SW, Bevan C, Brewin J, Calderwood C,
536 Canter A, Jessop F, Johnson G, Martin I, Metcalf L. Research priorities for stillbirth: process
537 overview and results from UK Stillbirth Priority Setting Partnership. *Ultrasound in obstetrics &*
538 *gynecology : the official journal of the International Society of Ultrasound in Obstetrics and*
539 *Gynecology* 2015; 46(6): 641-7.
- 540 7. Sexton J, Coory M, Kumar S, Smith G, Gordon A, Chambers G, Pereira G, Raynes-
541 Greenow C, Hilder L, Middleton P, Bowman A, Lieske S, Warrilow K, Ellwood D, Flenady V.
542 Protocol for the development and validation of a risk prediction model for stillbirths from 35
543 weeks gestation in Australia, 10 March 2020, PREPRINT (Version 1) available at Research
544 Square [+<https://doi.org/10.21203/rs.3.rs-16494/v1>]. 2020.
- 545 8. Townsend R, Manji A, Allotey J, Heazell A, Jorgensen L, Magee LA, Mol BW, Snell K,
546 Riley RD, Sandall J, Smith G, Patel M, Thilaganathan B, von Dadelszen P, Thangaratinam S, Khalil
547 A. Can risk prediction models help us individualise stillbirth prevention? A systematic review
548 and critical appraisal of published risk models. *BJOG : an international journal of obstetrics and*
549 *gynaecology* 2020.
- 550 9. Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, Collins GS. External
551 validation of clinical prediction models using big datasets from e-health records or IPD meta-
552 analysis: opportunities and challenges. *Bmj* 2016; 353: i3140.
- 553 10. Debray TP, Riley RD, Rovers MM, Reitsma JB, Moons KG, Cochrane IPDM-aMg.
554 Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies:
555 guidance on their use. *PLoS medicine* 2015; 12(10): e1001886.
- 556 11. Debray TPA, Moons KGM, Ahmed I, Koffijberg H, Riley RD. A framework for developing,
557 implementing, and evaluating clinical prediction models in an individual participant data meta-
558 analysis. *Statistics in Medicine* 2013; 32(18): 3158-80.

- 559 12. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new
560 framework to enhance the interpretation of external validation studies of clinical prediction
561 models. *Journal of clinical epidemiology* 2015; 68(3): 279-89.
- 562 13. Collins GS, Reitsma JB, Altman DG, Moons KG, for the members of the Tg. Transparent
563 Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD):
564 The TRIPOD Statement. *European urology* 2014.
- 565 14. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB,
566 Kleijnen J, Mallett S, Groupdagger P. PROBAST: A Tool to Assess the Risk of Bias and
567 Applicability of Prediction Model Studies. *Annals of internal medicine* 2019; 170(1): 51-8.
- 568 15. Allotey J, Snell KIE, Chan C, Hooper R, Dodds J, Rogozinska E, Khan KS, Poston L, Kenny
569 L, Myers J, Thilaganathan B, Chappell L, Mol BW, Von Dadelszen P, Ahmed A, Green M, Poon L,
570 Khalil A, Moons KGM, Riley RD, Thangaratnam S, Network IC. External validation, update and
571 development of prediction models for pre-eclampsia using an Individual Participant Data (IPD)
572 meta-analysis: the International Prediction of Pregnancy Complication Network (IPPIC pre-
573 eclampsia) protocol. *Diagn Progn Res* 2017; 1: 16.
- 574 16. Snell KIE, Allotey J, Smuk M, et al, for the IPPIC Collaborative Network. External
575 validation of prognostic models predicting pre-eclampsia: Individual participant data meta-
576 analysis. *BMC Medicine* 2020 (in press)
- 577 17. Allotey J, Snell KIE, Smuk M, et al, for the IPPIC Collaborative Network. Accuracy of
578 clinical characteristics, biochemical and ultrasound markers in predicting pre-eclampsia:
579 External validation and development of prediction models using an Individual Participant Data
580 (IPD) meta-analysis *Health Technol Assess* 2020 (in press).
- 581 18. Resche-Rigon M, White IR. Multiple imputation by chained equations for systematically
582 and sporadically missing multilevel data. *Stat Methods Med Res* 2016.
- 583 19. Jolani S, Debray TP, Koffijberg H, van Buuren S, Moons KG. Imputation of systematically
584 missing predictors in an individual participant data meta-analysis: a generalized approach using
585 MICE. *Statistics in medicine* 2015; 34(11): 1841-63.
- 586 20. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.
- 587 21. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research:
588 validating a prognostic model. *BMJ* 2009; 338: b605.
- 589 22. Hosmer DW, Lemeshow, S. Assessing the Fit of the Model. Applied Logistic Regression.
590 2nd ed. New York: Wiley; 2000: 143-202.
- 591 23. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials
592 with binary outcome. *Statistics in medicine* 2001; 20(24): 3875-89.
- 593 24. Langan D, Higgins JPT, Jackson D, Bowden J, Veroniki AA, Kontopantelis E, Viechtbauer
594 W, Simmonds M. A comparison of heterogeneity variance estimators in simulated random-
595 effects meta-analyses. *Res Synth Methods* 2019; 10(1): 83-98.
- 596 25. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-
597 analyses. *BMJ* 2003; 327(7414): 557-60.
- 598 26. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-
599 analysis. *J R Stat Soc Ser A Stat Soc* 2009; 172(1): 137-59.
- 600 27. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction
601 models. *Med Decis Making* 2006; 26(6): 565-74.
- 602 28. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of
603 prediction models, molecular markers, and diagnostic tests. *BMJ* 2016; 352: i6.

- 604 29. Smith GC, Yu CK, Papageorghiou AT, Cacho AM, Nicolaides KH, Fetal Medicine
605 Foundation Second Trimester Screening G. Maternal uterine artery Doppler flow velocimetry
606 and the risk of stillbirth. *Obstet Gynecol* 2007; 109(1): 144-51.
- 607 30. Yerlikaya G, Akolekar R, McPherson K, Syngelaki A, Nicolaides KH. Prediction of stillbirth
608 from maternal demographic and pregnancy characteristics. *Ultrasound in obstetrics &
609 gynecology : the official journal of the International Society of Ultrasound in Obstetrics and
610 Gynecology* 2016; 48(5): 607-12.
- 611 31. Trudell AS, Tuuli MG, Colditz GA, Macones GA, Odibo AO. A stillbirth calculator:
612 Development and internal validation of a clinical prediction model to quantify stillbirth risk.
613 *PloS one* 2017; 12(3): e0173461.
- 614 32. Kleinrouweler CE, Cheong-See Mrcog FM, Collins GS, Kwee A, Thangaratinam S, Khan
615 KS, Mol BW, Pajkrt E, Moons KG, Schuit E. Prognostic models in obstetrics: available, but far
616 from applicable. *American journal of obstetrics and gynecology* 2015.
- 617 33. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues
618 and guidance for practice. *Statistics in medicine* 2011; 30(4): 377-99.
- 619 34. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ,
620 Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for
621 Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal
622 medicine* 2015; 162(1): W1-73.
- 623 35. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable
624 prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Annals
625 of internal medicine* 2015; 162(1): 55-63.
- 626 36. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M.
627 Risk prediction models: II. External validation, model updating, and impact assessment. *Heart
628* 2012; 98(9): 691-8.
- 629 37. Kayode GA, Grobbee DE, Amoakoh-Coleman M, Adeleke IT, Ansah E, de Groot JA,
630 Klipstein-Grobusch K. Predicting stillbirth in a low resource setting. *BMC pregnancy and
631 childbirth* 2016; 16: 274.
- 632 38. Aupont JE, Akolekar R, Illian A, Neonakis S, Nicolaides KH. Prediction of stillbirth from
633 placental growth factor at 19-24 weeks. *Ultrasound in obstetrics & gynecology : the official
634 journal of the International Society of Ultrasound in Obstetrics and Gynecology* 2016; 48(5):
635 631-5.
- 636 39. Riley RD, Ensor J, Snell KIE, Harrell FE, Jr., Martin GP, Reitsma JB, Moons KGM, Collins G,
637 van Smeden M. Calculating the sample size required for developing a clinical prediction model.
638 *BMJ* 2020; 368: m441.
- 639 40. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Jr., Moons KG, Collins GS. Minimum
640 sample size for developing a multivariable prediction model: PART II - binary and time-to-event
641 outcomes. *Statistics in medicine* 2019; 38(7): 1276-96.
- 642 41. Saving Babies' Lives Version Two: A care bundle for reducing perinatal mortality
643 [https://www.england.nhs.uk/wp-content/uploads/2019/03/Saving-Babies-Lives-Care-Bundle-
644 Version-Two-Updated-Final-Version.pdf](https://www.england.nhs.uk/wp-content/uploads/2019/03/Saving-Babies-Lives-Care-Bundle-Version-Two-Updated-Final-Version.pdf) Accessed 15th October 2020.
- 645 42. Riley RD, van der Windt D, Croft P, Moons KGM. Prognosis Research in Healthcare:
646 Concepts, Methods and Impact. Oxford, UK: Oxford University Press; 2019.
- 647 43. Townsend R, Sileo FG, Allotey J, Dodds J, Heazell A, Jorgensen L, Kim VB, Magee L, Mol
648 B, Sandall J, Smith G, Thilaganathan B, von Dadelszen P, Thangaratinam S, Khalil A. Prediction
649 of stillbirth: an umbrella review of evaluation of prognostic variables. *BJOG : an international
650 journal of obstetrics and gynaecology* 2020.

