

1 **TITLE**

2 External validation and updating of clinical severity scores to guide referral of young children
3 with acute respiratory infections in resource-limited primary care settings

4

5

6 **AUTHORS**

7 Arjun Chandna,^{1,2} Lazaro Mwandigha,³ Constantinos Koshiaris,³ Direk

8 Limmathurotsakul,^{2,4,5} Francois Nosten,^{2,6} Yoel Lubell,^{2,4} Rafael Perera-Salazar,³ Claudia

9 Turner,^{1,2} and Paul Turner^{1,2}

10

11

12 **AFFILIATIONS**

13 1. Cambodia Oxford Medical Research Unit, Angkor Hospital for Children, Siem Reap,
14 Cambodia

15 2. Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, United
16 Kingdom

17 3. Department of Primary Care Health Sciences, University of Oxford, Oxford, United
18 Kingdom

19 4. Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine,
20 Mahidol University, Bangkok, Thailand

21 5. Department of Tropical Hygiene, Faculty of Tropical Medicine, Mahidol University,
22 Bangkok, Thailand

23 6. Shoklo Malaria Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae
24 Sot, Thailand

25

26 **CORRESPONDING AUTHOR**

27 Dr. Arjun Chandna

28 Cambodia Oxford Medical Research Unit, Angkor Hospital for Children, Siem Reap,

29 Cambodia

30 Email: arjun@tropmedres.ac

31 Telephone: +855-85-712-586

32

33

34 **SHORT TITLE**

35 Risk stratification of paediatric acute respiratory infections

36

37

38 **KEY WORDS**

39 Acute respiratory infection; paediatrics; risk stratification; triage; referral; primary care; low-

40 and middle-income country

41

42

43 **ABBREVIATIONS**

44 **ARI** = acute respiratory infection

45 **AUC** = area under the receiver operating characteristic curve

46 **AVPU** = Alert Voice Pain Unresponsive

47 **CI** = confidence interval

48 **EPP** = events per parameter

49 **GCS** = Glasgow Coma Scale

50 **iCCM** = integrated Community Case Management

- 51 **IMCI** = Integrated Management of Childhood Illnesses
- 52 **IQR** = interquartile range
- 53 **LAZ** = length-for-age z-score
- 54 **LMIC** = low- and middle-income country
- 55 **LqSOFA** = Liverpool quick Sequential Organ Failure Assessment
- 56 **LRT** = likelihood ratio test
- 57 **MAZ** = MUAC-for-age z-score
- 58 **mSIRS** = modified Systematic Inflammatory Response Syndrome
- 59 **MICE** = multiple imputation with chained equations
- 60 **MUAC** = mid-upper arm circumference
- 61 **NNR** = number needed to refer
- 62 **OxTREC** = Oxford Tropical Research Ethics Committee
- 63 **PICU** = Paediatric Intensive Care Unit
- 64 **qPELOD-2** = quick Pediatric Logistic Organ Dysfunction-2
- 65 **qSOFA** = quick Sequential Organ Failure Assessment
- 66 **SBP** = systolic blood pressure
- 67 **SpO₂** = peripheral oxygen saturation
- 68 **TMEC** = Tropical Medicine Ethics Committee
- 69 **TRIPOD** = Transparent Reporting of a multivariable prediction model for Individual
70 Prognosis Or Diagnosis
- 71 **WAZ** = weight-for-age z-score
- 72 **WLZ** = weight-for-length z-score
- 73
- 74
- 75

76 **ABSTRACT**

77

78 *Background*

79 Accurate and reliable guidelines for referral of children from resource-limited primary care
80 settings are lacking. We identified three practicable paediatric severity scores (Liverpool
81 quick Sequential Organ Failure Assessment [LqSOFA], quick Pediatric Logistic Organ
82 Dysfunction-2 [qPELOD-2], and the modified Systemic Inflammatory Response Syndrome
83 [mSIRS]) and externally validated their performance in young children presenting with acute
84 respiratory infections to a primary care clinic located within a refugee camp on the Thailand-
85 Myanmar border.

86

87 *Methods*

88 This secondary analysis of data from a longitudinal birth cohort study consisted of 3,010
89 acute respiratory infections in children aged ≤ 24 months. The primary outcome was receipt
90 of supplemental oxygen. We externally validated the discrimination, calibration, and net-
91 benefit of the scores, and quantified gains in performance that might be expected if they were
92 deployed as simple clinical prediction models, and updated to include nutritional status and
93 respiratory distress.

94

95 *Results*

96 104/3,010 (3.5%) presentations met the primary outcome. The LqSOFA score demonstrated
97 the best discrimination (AUC 0.84; 95% CI 0.79-0.89) and achieved a sensitivity and
98 specificity > 0.80 . Converting the scores into clinical prediction models improved
99 performance, resulting in ~20% fewer unnecessary referrals and ~30-60% fewer children
100 incorrectly managed in the community.

101 *Conclusions*

102 The LqSOFA score is a promising triage tool for young children presenting with acute
103 respiratory infections in resource-limited primary care settings. Where feasible, deploying the
104 score as a simple clinical prediction model might enable more accurate and nuanced risk
105 stratification, increasing applicability across a wider range of contexts.

106

107

108

109

110 INTRODUCTION

111

112 Acute respiratory infections (ARIs) are the leading reason for unscheduled childhood medical
113 consultations worldwide.^{1,2} Primary care workers function as gatekeepers to the formal health
114 system, aiming to distinguish the minority of ARIs requiring onward referral from those
115 suitable for community-based care.³

116

117 In rural regions of many low- and middle-income countries (LMICs) poorly functioning
118 infrastructure, as well as geographic, climatic, socioeconomic, and cultural factors, can
119 complicate referral mechanisms. Particularly in humanitarian and conflict settings referral
120 can entail risks for both patients and providers.⁴ Consequently, there can be substantial inter-
121 and intra-health system variation in referral thresholds.

122

123 Existing tools to support community healthcare providers in their assessment of unwell
124 children, such as the World Health Organization's Integrated Management of Childhood
125 Illnesses (IMCI) and Integrated Community Case Management (iCCM) guidelines,
126 recommend certain 'Danger Signs' to guide referrals.^{5,6} However, these lack sensitivity and
127 specificity, and suffer from considerable interobserver variability.^{7,8} A systematic review of
128 paediatric triage tools concluded that none would be reliable in resource-constrained settings
129 and that lack of follow-up data on children managed in the community rendered the validity
130 of existing tools questionable.⁹

131

132 In this study we identified paediatric severity scores suitable for use in resource-limited
133 primary care settings and externally validated their ability to guide referral of young children
134 presenting with ARIs.¹⁰ We characterised the improvement in performance that might be

135 expected if the scores were deployed as simple clinical prediction models and updated to

136 include variables relevant to children presenting with ARIs in rural LMIC settings.

137

138

139

140

141 **METHODS**

142

143 *Study population*

144 Data were collected during a prospective birth cohort study at a medical clinic for refugees
145 and internally displaced people on the Thailand-Myanmar border.¹⁰ Between September 2007
146 and September 2008 pregnant women receiving antenatal care at the clinic were invited to
147 participate. Children of consenting women were reviewed at birth and followed-up each
148 month (routine visit) and during any intercurrent illness (illness visit) until 24 months of age.
149 The local circumstances (inability of the population to move freely out of the camp and lack
150 of other medical providers) contributed to low attrition rates and capture of the majority of
151 acute illnesses for which care was sought.

152

153 All ARI illness visits were included in this secondary analysis. An ARI was defined as (A) a
154 presentation with rhinorrhoea, nasal congestion, cough, respiratory distress (chest indrawing,
155 nasal flaring, grunting, tracheal tug, and/or head bobbing), stridor, and/or abnormal lung
156 auscultation (crepitations and/or wheeze), and (B) a compatible contemporaneous syndromic
157 diagnosis (rhinitis, croup, bronchiolitis, influenza-like illness, pneumonia, viral infection
158 and/or wheeze) for children sent home directly from the clinic.

159

160

161 *Identification and shortlisting of scores*

162 Drawing on the results of two recent systematic reviews, we longlisted 16 severity scores that
163 might risk stratify young children presenting from the community with acute respiratory
164 infections (Supplementary Table 1).^{11,12} After considering reliability, validity, and feasibility
165 for implementation we excluded eight scores that required specialist equipment and/or

166 laboratory tests unlikely to be practical for the assessment of young children in busy LMIC
167 primary care settings.¹³⁻²⁰ Four others were excluded as $\geq 25\%$ of the constituent variables
168 were unavailable in the primary dataset.²¹⁻²⁴ Two of the remaining scores (quick Sequential
169 Organ Failure Assessment [qSOFA] and quick Pediatric Logistic Organ Dysfunction-2
170 [qPELOD-2]) contained blood pressure.^{25,26} Hypotension is a late sign in paediatric sepsis
171 and not suitable for early recognition of impending serious illness at the community level.²⁷
172 Furthermore, accurate use and maintenance of sphygmomanometers and stethoscopes may
173 not be feasible in resource-limited settings.²⁸ Recently, Romaine et al. replaced systolic blood
174 pressure (SBP) with alternate signs of circulatory compromise (heart rate and capillary refill
175 time) to develop the Liverpool-qSOFA (LqSOFA) score, and demonstrated superior
176 performance compared to qSOFA in febrile children presenting from the community.²⁹
177 Hence, we elected to evaluate the LqSOFA score in preference to qSOFA and to evaluate an
178 adapted qPELOD-2 score (replacing SBP with capillary refill time and assessing mental
179 status using the simpler Alert Voice Pain Unresponsive [AVPU] scale rather than the
180 Glasgow Coma Scale [GCS]). The three scores shortlisted for evaluation were the LqSOFA,
181 qPELOD-2, and modified Systemic Inflammatory Response Syndrome (mSIRS) scores
182 (Table 1).^{26,29,30}

183

184

185 *Selection of variables for model updating*

186 To update and improve model performance, additional predictors relevant for children
187 presenting with ARIs in LMIC primary care settings were considered for inclusion.
188 Nutritional status (weight-for-age z-score [WAZ]) and presence of respiratory distress were
189 selected a priori, after considering resource constraints, reliability, validity, biological

190 plausibility, availability of data in the primary dataset, and sample size (Supplementary Table
191 2).²⁸

192

193

194 *Data collection*

195 All data were measured by study staff and entered on to structured case report forms. With
196 the exception of anthropometric data, all clinical data were collected at the time of
197 presentation. Core (rectal) temperature was measured for neonates and infants and adjusted to
198 axillary temperature by subtracting 0.5°C.⁶ Mental status was assessed using the AVPU scale.
199 Capillary refill time was measured centrally. For children admitted to the clinic, weight was
200 measured at the time of presentation (seca scale; precision ± 5 g for neonates or ± 50 g after
201 birth). In addition, all children had their mid-upper arm circumference (MUAC), weight, and
202 height measured at each monthly routine visit. For the purposes of these analyses, age-
203 adjusted z-scores (R package: *z scorer*)³¹ were calculated using the closest anthropometric
204 data to the illness visit within the following window periods: height ≤ 28 days; MUAC ≤ 28
205 days without intervening admission; weight ≤ 14 days without intervening admission. Median
206 time between the index illness visit and each anthropometric measurement is reported.

207

208

209 *Primary outcome*

210 The primary outcome was receipt of supplemental oxygen during the illness visit. Study staff
211 were unaware which baseline variables were to be used as candidate predictors at the time of
212 ascertaining outcome status. Clinic treatment protocols specified that peripheral oxygen
213 saturation (SpO₂) must be checked prior to initiation of supplemental oxygen, with therapy

214 only indicated if SpO₂ was < 90%. All staff were trained on the treatment protocols prior to
215 study commencement.

216

217

218 *Missing data*

219 616 presentations were missing data on one or more candidate predictors (616/3,010; 20.5%)
220 with capillary refill time containing the highest proportion of missingness (442/3,010; 14.7%;
221 Supplementary Table 3). Under a missing-at-random assumption (Supplementary Figure 1),
222 we used multiple imputation with chained equations (MICE) to deal with missing data (R
223 package: *mice*).³² Analyses were done in each of 100 imputed datasets and results pooled.
224 Variables included in the imputation model are reported in Supplementary Table 4.

225

226

227 *Statistical methods*

228 We assessed discrimination and calibration of each score by quantifying the area under the
229 receiver operating characteristic curve (AUC) and plotting model scores against observed
230 outcome proportions. We examined predicted classifications at each of the scores' cut-offs.

231

232 Prior to model building we explored the relationship between continuous predictors and the
233 primary outcome using locally-weighted smoothing to identify non-linear patterns.

234 Accordingly, temperature was modelled using restricted cubic splines (R package: *rms*)³³

235 with three knots placed at locations based on percentiles (5th and 95th) and recognised

236 physiological thresholds (36°C).^{34,35} We used logistic regression to derive the models and

237 tested for important interactions using likelihood ratio tests (LRT). Random-effects were not

238 modelled as 22% (169/756) of children presented only once. All predictors were prespecified

239 and no predictor selection was performed during model development. Internal validation was
240 performed using 100 bootstrap samples with replacement and optimism-adjusted
241 discrimination and calibration reported (R package: *rms*).³³

242

243 Finally, the models were updated by including respiratory distress and WAZ as additional
244 candidate predictors. Penalised (lasso) logistic regression was used for model updating,
245 variable selection, and shrinkage to minimise overfitting (R package: *glmnet*).³⁶ A sensitivity
246 analysis confirmed that median imputation grouped by outcome status produced similar
247 results to MICE and hence to avoid conflicts in variable selection across multiply imputed
248 datasets we used this approach to address missing data for model updating (Supplementary
249 Table 5). We assessed discrimination and calibration of the updated models, examined
250 predicted classifications at clinically-relevant referral thresholds, and compared their clinical
251 utility (net-benefit) to the best-performing points-based severity score using decision curve
252 analysis (R package: *dcurves*).³⁷ A sensitivity analysis was performed excluding children who
253 were hypoxic at the time of presentation.

254

255 All analyses were done in R, version 4.0.2.³⁸

256

257

258 *Sample size*

259 No formal sample size calculation for external validation of the existing severity scores was
260 performed. All available data were used to maximise power and generalisability. Of the 3,010
261 eligible ARI presentations, 104 met the primary outcome, ensuring sufficient outcome events
262 for a robust external validation.³⁹ For derivation and updating of the clinical prediction
263 models we followed the methods of Riley et al. and assumed a conservative R^2 Nagelkerke of

264 0.15.⁴⁰ At an outcome prevalence of 3.5% (104/3,010) we estimated that up to 13 candidate
265 predictors (events per parameter [EPP] = 8) could be used to build the prediction models
266 whilst minimising the risk of overfitting (R package: *pmsampsize*).⁴¹

267

268

269 *Ethics and reporting*

270 Ethical approvals were provided by the Mahidol University Ethics Committee (TMEC 21-
271 023) and Oxford Tropical Research Ethics Committee (OxTREC 511-21). The study is
272 reported in accordance with the Transparent Reporting of a multivariable prediction model
273 for Individual Prognosis Or Diagnosis (TRIPOD) guidelines (Supplementary Table 6).⁴²

274

275

276

277

278 **RESULTS**

279

280 From September 2007 to September 2008, 999 pregnant women were enrolled, with 965
281 children born into the cohort. Amongst 4,061 acute illness presentations, 3,064 were for
282 ARIs. Fifty-four ARI presentations were excluded as information on oxygen therapy was not
283 available in the study database, leaving 3,010 presentations from 756 individual children for
284 the primary analysis (Supplementary Figure 2).

285

286 Baseline characteristics of the cohort are summarised (Table 2; Supplementary Table 7). The
287 majority of children were managed in the community (72.3%; 2,175/3,010). Median length of
288 stay for the 835 admissions was 3 days (IQR 2 to 4 days). One hundred and four (3.5%;
289 104/3,010) presentations met the primary outcome, with those with signs of respiratory
290 distress, age-adjusted tachycardia and/or tachypnoea, lower baseline SpO₂, prolonged
291 capillary refill times, altered mental status, and lower WAZ more likely to require
292 supplemental oxygen ($p < 0.001$ to 0.014; Table 2).

293

294

295 *LqSOFA and qPELOD-2 scores outperform the mSIRS score for risk stratification of ARIs*

296 Discrimination and calibration of the LqSOFA (AUC = 0.84; 95% confidence interval [CI] =
297 0.79 to 0.89) and qPELOD-2 (AUC = 0.79; 95% CI = 0.74 to 0.84) scores were considerably
298 better than the mSIRS score (AUC = 0.57; 95% CI = 0.51 to 0.63; Figure 1; Supplementary
299 Table 8; Supplementary Figure 3). At a cut-off of ≥ 1 the LqSOFA score demonstrated a
300 sensitivity of 0.80 (95% CI = 0.72 to 0.89) and specificity of 0.86 (95% CI = 0.85 to 0.88);
301 neither the mSIRS nor qPELOD-2 scores achieved a sensitivity and specificity > 0.70 at any
302 cut-off (Table 3).

303 *Improved performance of clinical severity scores when deployed as clinical prediction*
304 *models*

305 Relationships between continuous predictors and the primary outcome are illustrated
306 (Supplementary Figure 4). There was no evidence of interaction between heart rate (LRT =
307 2.09; $p = 0.35$) or respiratory rate (LRT = 0.77; $p = 0.68$) and age. Optimism-adjusted
308 discrimination of the three models ranged from 0.81 to 0.90, with the LqSOFA model
309 appearing most promising (AUC = 0.90; 95% CI = 0.86 to 0.94; Figure 2; Supplementary
310 Figure 5). Calibration of the qPELOD-2 model was good. The LqSOFA and mSIRS models
311 overestimated risk at higher predicted probabilities.

312

313 Discrimination of all three updated models containing respiratory distress and WAZ
314 improved (AUCs = 0.93 to 0.95). Calibration of the updated LqSOFA and qPELOD-2
315 models was good, whereas the updated mSIRS model underestimated risk at higher predicted
316 probabilities (Figure 3). The full models are reported in Supplementary Table 9.

317

318

319 *Promising clinical utility of the LqSOFA and qPELOD-2 models to guide referrals from*
320 *primary care*

321 We recognised that the relative value of correct and incorrect referrals is highly context-
322 dependent, reflecting resource availability, practicalities of referral, and capacity for follow-
323 up. Decision curve analyses accounting for differing circumstances suggest that the updated
324 models could provide greater utility (net-benefit) compared to the best points-based score (the
325 LqSOFA score), with the LqSOFA and qPELOD-2 models appearing most promising over a
326 wide range of plausible referral thresholds (Figure 4).

327

328 The ability of each updated model to guide referrals at thresholds ranging from 1% to 40% is
329 shown (Table 4). A referral threshold of 5% reflects a strategy whereby any child with a
330 predicted probability of requiring oxygen $\geq 5\%$ is referred. At this cut off, the models would
331 suggest referral in ~15% of all presentations, correctly identifying ~86-87% of children
332 requiring referral, at a cost of also recommending referral in ~12-13% of children not
333 requiring referral; i.e., a number needed to refer (NNR; the number of children referred to
334 identify one child who would require oxygen) of five. In contrast, at a similar threshold the
335 LqSOFA score using a cut-off ≥ 1 would suggest referral in a similar proportion of
336 presentations but result in a ~25% increase in incorrect referrals (a NNR of six) and a ~30-
337 60% increase in the number of children incorrectly identified for community-based
338 management.

339

340

341 *Sensitivity analysis*

342 The WHO recommend that pulse oximetry should be universally available at first-level health
343 facilities.^{6,43} Although many barriers exist to realising this laudable goal, to account for the
344 fact that in such contexts a severity score would not be required to guide referral for children
345 who are already hypoxic at the time of presentation, we performed a sensitivity analysis
346 excluding presentations with $\text{SpO}_2 < 90\%$ at enrolment. Discrimination remained comparable
347 but clinical utility of the models reduced slightly, with higher NNRs at the lowest referral
348 thresholds (Supplementary Tables 11 & 12).

349

350

351

352

353 **DISCUSSION**

354

355 We report the external validation of three pre-existing severity scores amongst young
356 children presenting with ARIs to a medical clinic on the Thailand-Myanmar border. Unlike
357 other studies which investigated the scores' prognostic accuracy in hospital settings,^{17,25} we
358 evaluated their performance at the community level and demonstrate that the LqSOFA and
359 qPELOD-2 scores could support early recognition of children requiring referral or closer
360 follow-up in settings with limited resources. In keeping with previous literature, we found
361 that the mSIRS score was poorly discriminative, not well calibrated, and led to substantial
362 misclassification.¹⁷

363

364 An LqSOFA score ≥ 1 yielded a sensitivity and specificity $> 80\%$. Encouragingly, this is
365 remarkably consistent with the performance reported in the original LqSOFA development
366 study and may reflect similarities in the use-case (febrile children presenting from the
367 community) and severity of the cohorts (outcome prevalence 1.1% vs. 3.5%; admission rate
368 12.1% vs. 27.7%), albeit despite obvious demographic differences.²⁹ In contrast to qPELOD-
369 2, LqSOFA contains age-adjusted tachypnoea, which may have improved performance in
370 children with respiratory illnesses. Furthermore, the performance of LqSOFA (or qSOFA)
371 has been shown to improve outside of the PICU, when used to predict more proximal
372 outcomes (e.g. critical care admission rather than mortality), and if the AVPU scale (vs.
373 GCS) is employed to assess mental status.⁴⁴ These all apply to our cohort.

374

375 We demonstrated improvement in performance when the severity scores were deployed as
376 clinical prediction models and when nutritional status and respiratory distress were included
377 as additional predictors. Whilst discrimination of all three updated models was good, the

378 AUC is a summary measure of model performance and does not necessarily reflect clinical
379 utility.⁴⁵⁻⁴⁷ Decision curve analyses illustrate the superiority of the LqSOFA and qPELOD-2
380 models compared with the mSIRS model across a range of clinically-relevant referral
381 thresholds.

382

383 With growing access to smartphones there may be contexts where the increased accuracy
384 afforded by a clinical prediction model outweighs the simplicity and practicality of points-
385 based scoring systems. At a 5% referral threshold, the updated LqSOFA model identified a
386 similar proportion of presentations for referral as the LqSOFA score at a cut-off of ≥ 1
387 (14.1% vs. 16.1%), however use of the model would have resulted in ~25% fewer incorrect
388 referrals and a ~30% decrease in the number of presentations incorrectly recommended for
389 community-based management. In addition to greater accuracy, prediction models permit
390 more nuanced evaluation of risk; referral thresholds can be adjusted to the needs of an
391 individual patient and/or health system and this flexibility may be particularly impactful in
392 the heterogeneous environments commonplace in many LMIC primary care contexts. For
393 example, in locations where community follow-up is feasible (e.g. via a telephone call or
394 return clinic visit) and/or referral carries great cost (to the patient or system), a higher referral
395 threshold (lower NNR) may be acceptable, compared with settings where safety-netting is
396 impractical and/or access to secondary care is less challenging.

397

398 We followed the latest guidelines in prediction model building and used bootstrap internal
399 validation, penalised regression, placed knots at predefined locations, and limited the number
400 of candidate predictors to avoid overfitting the models.^{40,42,48,49} Nevertheless, they require
401 validation on new data to assess generalisability and provide a fairer comparison with the pre-

402 existing points-based scores. We have published our full models to encourage independent
403 validation.

404

405 As others have highlighted, a limitation of many studies evaluating community-based triage
406 tools in low-resource settings is the lack of follow-up data for patients categorised as low
407 risk;⁹ 72.3% (2,175/3,010) of our cohort were sent away from the clinic without admission.

408 As acute illness visits were nested within the longitudinal birth cohort, we were able to
409 confirm that 1.4% (30/2,083) of presentations sent away from the clinic without admission
410 received supplemental oxygen within the next 28 days, although it is unknown whether this
411 related to the index ARI or a new illness. A sensitivity analysis conservatively classifying
412 these 30 presentations as meeting the primary outcome (i.e. assuming the oxygen therapy
413 related to the index ARI) resulted in a decrease in the sensitivity of all three models
414 (Supplementary Tables 12 & 13). Prospective research with dedicated outpatient follow-up is
415 ongoing to investigate this issue further.⁵⁰

416

417 We selected supplemental oxygen therapy as the primary outcome as this reflects a clinically-
418 meaningful endpoint for ARIs and a pragmatic referral threshold for many resource-limited
419 primary care settings. Oxygen was a scarce resource during the study (cylinders were
420 transported in each week from ~60km away) and oxygen therapy was protocolised; hence
421 outcome misclassification is less likely.

422

423 For those who met the primary outcome, the time of oxygen initiation was not available in
424 the primary dataset. Although no patient had met the outcome when baseline predictors were
425 measured, some may have done so shortly after. Nevertheless, the sensitivity analysis
426 excluding presentations with baseline SpO₂ < 90% (the qualifying criterion for supplemental

427 oxygen) produced similar results. Furthermore, median length of stay was three days and
428 hence the time horizon for all those who met the primary outcome is likely to have been
429 relatively comparable.

430

431 We externally validated three severity scores that could guide assessment of young children
432 presenting with ARIs in resource-limited primary care settings to identify those in need of
433 referral or closer follow-up. Performance of the LqSOFA score was encouraging and
434 comparable to that in the original derivation setting.²⁹ Converting the LqSOFA score into a
435 clinical prediction model and including additional variables relevant to resource-constrained
436 LMIC settings improved accuracy and might permit application across a wider range of
437 contexts with differing referral thresholds.

438

439

440

441 **CONFLICT OF INTEREST DISCLOSURES**

442 The authors have no conflicts of interest relevant to this article to disclose.

443

444

445 **FUNDING/SUPPORT**

446 This research was funded by the UK Wellcome Trust [219644/Z/19/Z]. RPS acknowledges
447 part support from the NIHR Applied Research Collaboration Oxford & Thames Valley, the
448 NIHR Oxford Medtech and In-Vitro Diagnostics Co-operative and the Oxford Martin School.

449 CK is supported by a Wellcome Trust/Royal Society Sir Henry Dale Fellowship
450 [211182/Z/18/Z]. For the purpose of open access, the author has applied a CC BY public
451 copyright license to any Author Accepted Manuscript version arising from this submission.

452

453

454 **DATA SHARING**

455 De-identified, individual participant data from this study will be available to researchers
456 whose proposed purpose of use is approved by the data access committees at the Mahidol-
457 Oxford Tropical Medicine Research Unit. Inquiries or requests for the data may be sent to
458 datasharing@tropmedres.ac.

459 REFERENCES

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

1. Bigio J, MacLean E, Vasquez NA, et al. Most common reasons for primary care visits in low- and middle-income countries: A systematic review. *PLOS Global Public Health* 2022; **2**(5).
2. Finley CR, Chan DS, Garrison S, et al. What are the most common conditions in primary care? *Can Fam Phys* 2018; **64**(11): 832-40.
3. Buntinx F, Mant D, Van den Bruel A, Donner-Banzhof N, Dinant GJ. Dealing with low-incidence serious diseases in general practice. *Br J Gen Pract* 2011; **61**(582): 43-6.
4. Debarre A. Hard to Reach: Providing Healthcare in Armed Conflict: International Peace Institute, 2018.
5. World Health Organization. Integrated Community Case Management. Geneva, Switzerland; 2012.
6. World Health Organization. Integrated Management of Childhood Illnesses. Geneva, Switzerland; 2014.
7. Keitel K, Kilowoko M, Kyungu E, Genton B, D'Acremont V. Performance of prediction rules and guidelines in detecting serious bacterial infections among Tanzanian febrile children. *BMC Infect Dis* 2019; **19**(1): 769.
8. Izudi J, Anyigu S, Ndungutse D. Adherence to Integrated Management of Childhood Illnesses Guideline in Treating South Sudanese Children with Cough or Difficulty in Breathing. *Int J Pediatr* 2017; **2017**: 5173416.
9. Hansoti B, Jenson A, Keefe D, et al. Reliability and validity of pediatric triage tools evaluated in Low resource settings: a systematic review. *BMC Pediatr* 2017; **17**(1): 37.
10. Turner C, Turner P, Carrara V, et al. High rates of pneumonia in children under two years of age in a South East Asian refugee population. *PLoS One* 2013; **8**(1): e54026.
11. Chandna A, Tan R, Carter M, et al. Predictors of disease severity in children presenting from the community with febrile illnesses: a systematic review of prognostic studies. *BMJ Glob Health* 2021; **6**(1).
12. Deardorff KV, McCollum ED, Ginsburg AS. Pneumonia Risk Stratification Scores for Children in Low-Resource Settings: A Systematic Literature Review. *Pediatr Infect Dis J* 2018; **37**(8): 743-8.
13. Olson D, Davis NL, Milazi R, et al. Development of a severity of illness scoring system (inpatient triage, assessment and treatment) for resource-constrained hospitals in developing countries. *Trop Med Int Health* 2013; **18**(7): 871-8.
14. Hooli S, Colbourn T, Lufesi N, et al. Predicting Hospitalised Paediatric Pneumonia Mortality Risk: An External Validation of RISC and mRISC, and Local Tool Development (RISC-Malawi) from Malawi. *PLoS One* 2016; **11**(12): e0168126.
15. Reed C, Madhi SA, Klugman KP, et al. Development of the Respiratory Index of Severity in Children (RISC) score among young children with respiratory infections in South Africa. *PLoS One* 2012; **7**(1): e27793.
16. Pollack MM, Holubkov R, Funai T, et al. The Pediatric Risk of Mortality Score: Update 2015. *Pediatr Crit Care Med* 2016; **17**(1): 2-9.
17. van Nassau SC, van Beek RH, Driessen GJ, Hazelzet JA, van Wering HM, Boeddha NP. Translating Sepsis-3 Criteria in Children: Prognostic Accuracy of Age-Adjusted Quick SOFA Score in Children Visiting the Emergency Department With Suspected Bacterial Infection. *Front Pediatr* 2018; **6**: 266.
18. Goldstein B, Giroir B, Randolph A, International Consensus Conference on Pediatric S. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med* 2005; **6**(1): 2-8.
19. Egdell P, Finlay L, Pedley DK. The PAWS score: validation of an early warning scoring system for the initial assessment of children in the emergency department. *Emerg Med J* 2008; **25**(11): 745-9.
20. Parshuram CS, Hutchison J, Middaugh K. Development and initial validation of the Bedside Paediatric Early Warning System score. *Crit Care* 2009; **13**(4): R135.
21. George EC, Walker AS, Kiguli S, et al. Predicting mortality in sick African children: the FEAST Paediatric Emergency Triage (PET) Score. *BMC Med* 2015; **13**: 174.
22. Emukule GO, McMorrow M, Ulloa C, et al. Predicting mortality among hospitalized children with respiratory illness in Western Kenya, 2009-2012. *PLoS One* 2014; **9**(3): e92968.
23. Berkley JA, Ross A, Wangi I, et al. Prognostic indicators of early and late death in children admitted to district hospital in Kenya: cohort study. *BMJ* 2003; **326**(361).
24. Helbok R, Kendjo E, Issifou S, et al. The Lambarene Organ Dysfunction Score (LODS) is a simple clinical predictor of fatal malaria in African children. *J Infect Dis* 2009; **200**(12): 1834-41.
25. Schlapbach LJ, Straney L, Bellomo R, MacLaren G, Pilcher D. Prognostic accuracy of age-adapted SOFA, SIRS, PELOD-2, and qSOFA for in-hospital mortality among children with suspected infection admitted to the intensive care unit. *Intensive Care Med* 2018; **44**(2): 179-88.

- 513 26. Leclerc F, Duhamel A, Deken V, Grandbastien B, Leteurtre S, Groupe Francophone de Reanimation et
514 Urgences P. Can the Pediatric Logistic Organ Dysfunction-2 Score on Day 1 Be Used in Clinical Criteria for
515 Sepsis in Children? *Pediatr Crit Care Med* 2017; **18**(8): 758-63.
- 516 27. Brierley J, Carcillo JA, Choong K, et al. Clinical practice parameters for hemodynamic support of pediatric
517 and neonatal septic shock: 2007 update from the American College of Critical Care Medicine. *Crit Care Med*
518 2009; **37**(2): 666-88.
- 519 28. Fung JST, Akech S, Kissoon N, Wiens MO, English M, Ansermino JM. Determining predictors of sepsis at
520 triage among children under 5 years of age in resource-limited settings: A modified Delphi process. *PLoS*
521 *One* 2019; **14**(1): e0211274.
- 522 29. Romaine S.T, Potter J, Khanijau A, et al. Accuracy of a Modified qSOFA Score for Predicting Critical Care
523 Admission in Febrile Children. *Pediatrics* 2020; **146**(4): e20200782.
- 524 30. Beane A, Silva AP, Munasinghe S, et al. Comparison of Quick Sequential Organ Failure Assessment and
525 Modified Systemic Inflammatory Response Syndrome Criteria in a Lower Middle Income Setting. *J Acute*
526 *Med* 2017; **7**(4): 141-8.
- 527 31. Myatt M, Guevarra E. zscorer: Child Anthropometry z-Score Calculator. R package version 0.3.1. 2019.
- 528 32. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of*
529 *Statistical Software* 2011; **45**(3): 1-67.
- 530 33. Harrell FE, Jr. rms: Regression Modeling Strategies. R package version 6.2-0. 2021.
- 531 34. UK National Institute for Health and Care Excellence. Algorithm for managing suspected sepsis in children
532 aged under 5 years outside an acute hospital setting. United Kingdom; 2017.
- 533 35. World Health Organization. Pocket book of hospital care for children: guidelines for the management of
534 common childhood illnesses. Geneva, Switzerland; 2013.
- 535 36. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate
536 Descent. *Journal of Statistical Software* 2010; **33**(1): 1-22.
- 537 37. Sjoberg DD. dcurves: Decision Curve Analysis for Model Evaluation. R package version 0.3.0. 2022.
- 538 38. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for
539 Statistical Computing; 2020.
- 540 39. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required
541 for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005; **58**(5): 475-83.
- 542 40. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction
543 model. *BMJ* 2020; **368**: m441.
- 544 41. Ensor J, Martin EC, Riley RD. pmsampsize: Calculates the Minimum Sample Size Required for Developing a
545 Multivariable Prediction Model. R package version 1.1.1. R; 2021.
- 546 42. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model
547 for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015; **162**(1): 55-63.
- 548 43. World Health Organization. Interagency List of Priority Medical Devices for Essential Interventions for
549 Reproductive, Maternal, Newborn and Child Health. Geneva, Switzerland, 2016.
- 550 44. Eun S, Kim H, Kim HY, et al. Age-adjusted quick Sequential Organ Failure Assessment score for predicting
551 mortality and disease severity in children with infection: a systematic review and meta-analysis. *Sci Rep*
552 2021; **11**(1): 21699.
- 553 45. Fackler JC, Rehman M, Winslow RL. Please Welcome the New Team Member: The Algorithm. *Pediatr Crit*
554 *Care Med* 2019; **20**(12): 1200-1.
- 555 46. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models,
556 molecular markers, and diagnostic tests. *BMJ* 2016; **352**: i6.
- 557 47. de Hond AAH, Steyerberg EW, van Calster B. Interpreting area under the receiver operating characteristic
558 curve. *Lancet Digit Health* 2022; **0**(0).
- 559 48. Steyerberg EW. Clinical Prediction Models: Springer International Publishing; 2019.
- 560 49. Harrell FE, Jr. Regression Modeling Strategies: Springer International Publishing; 2006.
- 561 50. Chandna A, Aderie EM, Ahmad R, et al. Prediction of disease severity in young children presenting with
562 acute febrile illness in resource-limited settings: a protocol for a prospective observational study. *BMJ Open*
563 2021; **11**(1): e045826.

564

565 **TABLE 1. Shortlisted paediatric severity scores and comparison between original and study populations.** bpm = beats or breaths per
 566 minute; ED = emergency department; ICU = intensive care unit; PICU = paediatric intensive care unit.

Score	Constituent variables	Population	Outcome
LqSOFA ²⁹	1. Capillary refill time > 2 seconds 2. Mental status < alert on AVPU scale 3. Heart rate > age-adjusted threshold 4. Respiratory rate > age-adjusted threshold <i>Each variable allocated one point to give score of 0-4</i>	<u>Derivation:</u> 1,121 febrile children < 16y attending the ED and requiring a blood test at a specialist paediatric hospital in the United Kingdom. <u>Validation:</u> 12,241 febrile children < 16y attending the ED at a specialist paediatric hospital in the United Kingdom.	Critical care admission within 48h of ED attendance. <u>Prevalence:</u> 4.2% (derivation) and 1.1% (validation).
mSIRS ³⁰	1. Core temperature > 38.5°C or < 36°C 2. Heart rate > or < age-adjusted threshold 3. Respiratory rate > age-adjusted threshold <i>Each variable allocated one point to give score of 0-3</i>	<u>Derivation:</u> expert consensus (original SIRS score). ¹⁸ <u>Validation:</u> 1,184 adults > 18y admitted to a hospital in Sri Lanka with suspected infection.	In-hospital mortality, cardiac arrest or ICU admission (validation). <u>Prevalence:</u> 3.6% (validation).
qPELOD-2 ²⁶	1. Mental status < 11 on GCS 2. Heart rate > age-adjusted threshold 3. Blood pressure < age-adjusted threshold <i>Each variable allocated one point to give score of 0-3</i>	<u>Derivation:</u> 862 children < 18y admitted to nine European PICUs with suspected infection. <u>Validation:</u> 545 children < 18y admitted to a hospital in the Netherlands with suspected bacterial infection. ¹⁷	In-PICU mortality (derivation) or PICU admission and/or mortality (validation). <u>Prevalence:</u> 7.0% (derivation) and 3.3% (validation).
This study	1. Capillary refill time > 2 seconds 2. Mental status < alert on AVPU scale 3. Heart rate > age-adjusted threshold 4. Respiratory rate > age-adjusted threshold 5. Axillary temperature > 38°C or < 35.5°C	3,010 ARI presentations from 756 children < 2y presenting to a primary care clinic on the Thai-Myanmar border.	Supplemental oxygen therapy. <u>Prevalence:</u> 3.5%.

567 **TABLE 2. Baseline characteristics of the cohort stratified by primary outcome status.** ^aRespiratory distress defined as head bobbing,
568 tracheal tug, grunting and/or chest indrawing; ^babnormal chest auscultation defined as crepitations and/or wheeze; ^crectal temperature converted
569 to axillary temperature for neonates and infants. [†]Median interval between anthropometric measurement and index illness presentation: length =
570 8 days (IQR 4-12 days); MUAC = 9 days (IQR 4-13 days); weight = 4 days (IQR 0-10 days). *Missing data: gestation = 5; birthweight = 14;
571 comorbidity = 10; symptom duration = 21; unwell family member = 10; fever = 5; runny nose = 2; noisy breathing = 6; stridor = 1; respiratory
572 distress = 1; head bobbing = 1; tracheal tug = 1; grunting = 1; chest indrawing = 1; abnormal lung auscultation = 59; lung crepitations = 69;
573 wheeze = 79; dehydration = 7; colour = 50; heart rate = 9; respiratory rate = 8; temperature = 3; oxygen saturation = 1,645; capillary refill time =
574 442; mental status = 37; WLZ = 158; WAZ = 147; MAZ = 682; LAZ = 14.
575

Characteristic	Overall N = 3,010 ¹	Supplemental oxygen		p-value ²
		No N = 2,906 ¹	Yes N = 104 ¹	
Demographics				
Age (months)	8.1 (3.7, 13.7)	8.2 (3.8, 13.8)	7.3 (3.4, 12.7)	0.40
Male sex	1,592 / 3,010 (53%)	1,541 / 2,906 (53%)	51 / 104 (49%)	0.40
Birth history				
Gestation (weeks)*	39.1 (38.1, 40.0)	39.2 (38.2, 40.0)	38.4 (37.3, 39.7)	0.001
Birthweight (kg)*	2.9 (2.6, 3.2)	2.9 (2.6, 3.2)	2.6 (2.0, 3.0)	<0.001
Previous medical history				
Number of previous illness visits	3.0 (2.0, 6.0)	3.0 (2.0, 6.0)	4.0 (2.0, 9.0)	0.043

Characteristic	Overall N = 3,010 ¹	Supplemental oxygen		p-value ²
		No N = 2,906 ¹	Yes N = 104 ¹	
Time since last illness visit (days)	29.0 (3.0, 81.0)	31.0 (3.0, 82.0)	11.0 (2.0, 36.5)	<0.001
Number of previous ARI visits	3.0 (2.0, 5.0)	3.0 (2.0, 5.0)	3.5 (2.0, 8.0)	0.006
Known comorbidity*	53 / 3,000 (1.8%)	39 / 2,898 (1.3%)	14 / 102 (14%)	<0.001
History of current illness				
Duration of symptoms (days)*	3.0 (2.0, 5.0)	3.0 (2.0, 5.0)	3.0 (2.0, 5.0)	0.30
Antibiotics prior to presentation	145 / 3,010 (4.8%)	125 / 2,906 (4.3%)	20 / 104 (19%)	<0.001
Family member unwell*	287 / 3,000 (9.6%)	276 / 2,898 (9.5%)	11 / 102 (11%)	0.70
Presenting symptoms and signs				
Fever*	1,958 / 3,005 (65%)	1,885 / 2,901 (65%)	73 / 104 (70%)	0.30
Cough	2,767 / 3,010 (92%)	2,667 / 2,906 (92%)	100 / 104 (96%)	0.11
Runny nose*	2,565 / 3,008 (85%)	2,491 / 2,904 (86%)	74 / 104 (71%)	<0.001
Noisy breathing*	447 / 3,004 (15%)	430 / 2,901 (15%)	17 / 103 (17%)	0.60
Stridor*	6 / 3,009 (0.2%)	6 / 2,905 (0.2%)	0 / 104 (0%)	>0.90
Respiratory distress ^{a*}	508 / 3,009 (17%)	416 / 2,905 (14%)	92 / 104 (88%)	<0.001
Head bobbing*	52 / 3,009 (1.7%)	27 / 2,905 (0.9%)	25 / 104 (24%)	<0.001

Characteristic	Overall N = 3,010 ¹	Supplemental oxygen		p-value ²
		No N = 2,906 ¹	Yes N = 104 ¹	
Tracheal tug *	134 / 3,009 (4.5%)	96 / 2,905 (3.3%)	38 / 104 (37%)	<0.001
Grunting *	26 / 3,009 (0.9%)	11 / 2,905 (0.4%)	15 / 104 (14%)	<0.001
Chest indrawing *	493 / 3,009 (16%)	402 / 2,905 (14%)	91 / 104 (88%)	<0.001
Abnormal lung auscultation ^{b*}	1,455 / 2,951 (49%)	1,372 / 2,852 (48%)	83 / 99 (84%)	<0.001
Crepitations *	1,158 / 2,941 (39%)	1,085 / 2,844 (38%)	73 / 97 (75%)	<0.001
Wheeze *	794 / 2,931 (27%)	751 / 2,833 (27%)	43 / 98 (44%)	<0.001
Dehydration *	127 / 3,003 (4.2%)	121 / 2,899 (4.2%)	6 / 104 (5.8%)	0.40
Pale, mottled or cyanosed *	107 / 2,960 (3.6%)	91 / 2,862 (3.2%)	16 / 98 (16%)	<0.001
Vital signs				
Heart rate (bpm) *				
Neonate	140.0 (132.0, 150.0)	140.0 (132.0, 148.0)	150.0 (140.0, 165.0)	0.014
Infant	138.0 (128.0, 144.0)	136.0 (128.0, 144.0)	147.0 (136.5, 154.0)	<0.001
Child	128.0 (120.0, 140.0)	128.0 (120.0, 140.0)	140.0 (127.5, 149.0)	0.002
Respiratory rate (bpm) *				
Neonate	48.0 (45.0, 56.0)	48.0 (44.2, 54.0)	64.5 (54.0, 77.0)	0.008

Characteristic	Overall N = 3,010 ¹	Supplemental oxygen		p-value ²
		No N = 2,906 ¹	Yes N = 104 ¹	
Infant	48.0 (42.0, 56.0)	48.0 (42.0, 56.0)	58.0 (54.0, 66.0)	<0.001
Child	45.0 (38.0, 52.0)	44.0 (38.0, 52.0)	57.0 (46.5, 62.0)	<0.001
Axillary temperature (°C) ^{c*}	36.6 (36.0, 37.5)	36.6 (36.0, 37.4)	36.8 (36.2, 37.8)	0.040
Oxygen saturation (%) [*]	95.0 (93.0, 96.0)	95.0 (93.0, 96.0)	88.0 (85.0, 93.0)	<0.001
Capillary refill time > 2 secs [*]	36 / 2,568 (1.4%)	27 / 2,476 (1.1%)	9 / 92 (9.8%)	<0.001
Not alert [*]	372 / 2,973 (13%)	306 / 2,875 (11%)	66 / 98 (67%)	<0.001
Anthropometrics				
Weight-for-length z-score (WLZ) ^{*†}	0.0 (-0.8, 0.8)	0.0 (-0.8, 0.8)	-0.5 (-1.8, 0.7)	<0.001
Weight-for-age z-score (WAZ) ^{*†}	-0.9 (-1.6, -0.2)	-0.9 (-1.6, -0.2)	-1.9 (-3.4, -0.8)	<0.001
MUAC-for-age z-score (MAZ) ^{*†}	0.2 (-0.4, 0.8)	0.2 (-0.4, 0.8)	-0.7 (-1.9, 0.6)	<0.001
Length-for-age z-score (LAZ) ^{*†}	-1.5 (-2.3, -0.7)	-1.4 (-2.2, -0.7)	-2.4 (-3.4, -1.4)	<0.001

¹Median (IQR); n / N (%)

²Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

TABLE 3. Predicted classifications of the severity scores. Classifications calculated using full-case analysis: LqSOFA = 2,525 presentations (81 met primary outcome); mSIRS = 2,992 presentations (99 met primary outcome); qPELOD-2 = 2,531 presentations (83 met primary outcome).

Cut Off	Sensitivity (95% CI)	Specificity (95% CI)	Negative Predictive Value (95% CI)	Positive Predictive Value (95% CI)	Negative Likelihood Ratio (95% CI)	Positive Likelihood Ratio (95% CI)	Cases referred (%)	Cases managed in community (%)	Ratio of Incorrect to Correct referrals	Ratio of Correct to Incorrect cases managed in community
LqSOFA										
≥ 1	0.80 (0.72 to 0.89)	0.86 (0.85 to 0.88)	0.99 (0.99 to 1.00)	0.16 (0.13 to 0.20)	0.23 (0.15 to 0.36)	5.89 (5.08 to 6.82)	407 (16.1%)	2118 (83.9%)	5 to 1	131 to 1
≥ 2	0.23 (0.14 to 0.33)	0.98 (0.98 to 0.99)	0.98 (0.97 to 0.98)	0.34 (0.22 to 0.46)	0.78 (0.69 to 0.88)	15.49 (9.33 to 25.72)	68 (2.7%)	2457 (97.3%)	3 to 1	39 to 1
≥ 3	0.01 (-0.01 to 0.04)	1.00 (1.00 to 1.00)	0.97 (0.96 to 0.98)	0.33 (-0.20 to 0.87)	0.99 (0.96 to 1.01)	15.09 (1.38 to 164.69)	1 (< 0.01%)	2524 (> 99.9%)	0 to 1	31 to 1
mSIRS										
≥ 1	0.99 (0.97 to 1.00)	0.05 (0.04 to 0.05)	0.99 (0.98 to 1.01)	0.03 (0.03 to 0.04)	0.22 (0.03 to 1.54)	1.04 (1.02 to 1.06)	2846 (95.1%)	146 (4.9%)	28 to 1	145 to 1
≥ 2	0.22 (0.14 to 0.30)	0.88 (0.86 to 0.89)	0.97 (0.96 to 0.98)	0.06 (0.03 to 0.08)	0.89 (0.80 to 0.99)	1.79 (1.22 to 2.61)	369 (12.3%)	2623 (87.7%)	16 to 1	33 to 1
≥ 3	0.01 (-0.01 to 0.03)	1.00 (1.00 to 1.00)	0.97 (0.96 to 0.97)	0.50 (-0.19 to 1.19)	0.99 (0.97 to 1.01)	29.22 (1.84 to 463.84)	1 (< 0.1%)	2991 (> 99.9%)	0 to 1	30 to 1
qPELOD-2										
≥ 1	0.68 (0.57 to 0.78)	0.90 (0.88 to 0.91)	0.99 (0.98 to 0.99)	0.18 (0.14 to 0.22)	0.36 (0.27 to 0.50)	6.40 (5.30 to 7.73)	301 (11.9%)	2230 (88.1%)	4 to 1	82 to 1
≥ 2	0.08 (0.03 to 0.14)	0.99 (0.99 to 1.00)	0.97 (0.96 to 0.98)	0.32 (0.12 to 0.51)	0.92 (0.86 to 0.98)	13.76 (5.77 to 32.86)	31 (1.2%)	2500 (98.8%)	3 to 1	32 to 1

- 1 **TABLE 4. Predicted classifications at different referral thresholds using the updated LqSOFA, qPELOD-2, and mSIRS models.** A
- 2 referral threshold of 5% reflects a management strategy whereby any child with a predicted probability of requiring oxygen $\geq 5\%$ is referred.

Model	Sensitivity (95% CI)	Specificity (95% CI)	Negative Predictive Value (95% CI)	Positive Predictive Value (95% CI)	Negative Likelihood Ratio (95% CI)	Positive Likelihood Ratio (95% CI)	Cases referred (%)	Cases managed in community (%)	Ratio of Incorrect to Correct referrals	Ratio of Correct to Incorrect case managed in community
Referral threshold = 1%										
LqSOFA	0.97 (0.93 to 1.00)	0.78 (0.73 to 0.82)	1.00 (1.00 to 1.00)	0.14 (0.11 to 0.17)	0.04 (0.00 to 0.09)	4.48 (3.65 to 5.57)	722 (24.0%)	2288 (76.0%)	6 to 1	762 to 1
qPELOD-2	0.96 (0.93 to 0.99)	0.79 (0.75 to 0.83)	1.00 (1.00 to 1.00)	0.14 (0.12 to 0.17)	0.05 (0.01 to 0.10)	4.55 (3.97 to 5.87)	715 (23.8%)	2295 (76.2%)	6 to 1	573 to 1
mSIRS	0.94 (0.90 to 0.98)	0.78 (0.65 to 0.84)	1.00 (1.00 to 1.00)	0.13 (0.10 to 0.18)	0.08 (0.04 to 0.14)	4.36 (3.00 to 6.21)	737 (24.5%)	2273 (75.5%)	7 to 1	378 to 1
Referral threshold = 5%										
LqSOFA	0.87 (0.78 to 0.93)	0.88 (0.86 to 0.91)	0.99 (0.99 to 1.00)	0.21 (0.18 to 0.25)	0.15 (0.09 to 0.25)	7.40 (6.22 to 9.50)	423 (14.1%)	2587 (85.9%)	4 to 1	171 to 1
qPELOD-2	0.87 (0.78 to 0.93)	0.87 (0.85 to 0.91)	0.99 (0.99 to 1.00)	0.20 (0.16 to 0.23)	0.15 (0.08 to 0.25)	6.79 (5.96 to 8.98)	468 (15.5%)	2542 (84.5%)	4 to 1	181 to 1
mSIRS	0.86 (0.77 to 0.93)	0.87 (0.85 to 0.89)	0.99 (0.99 to 1.00)	0.19 (0.16 to 0.22)	0.16 (0.08 to 0.26)	6.55 (5.80 to 7.74)	470 (15.6%)	2540 (84.4%)	4 to 1	211 to 1
Referral threshold = 10%										
LqSOFA	0.75 (0.66 to 0.83)	0.93 (0.91 to 0.95)	0.99 (0.99 to 0.99)	0.29 (0.24 to 0.36)	0.26 (0.18 to 0.37)	11.76 (9.04 to 16.80)	270 (9.0%)	2740 (91.0%)	3 to 1	100 to 1

qPELOD-2	0.73 (0.61 to 0.82)	0.93 (0.90 to 0.95)	0.99 (0.99 to 0.99)	0.29 (0.23 to 0.37)	0.29 (0.20 to 0.41)	11.57 (8.21 to 17.02)	264 (8.8%)	2764 (91.2%)	3 to 1	94 to 1
mSIRS	0.76 (0.63 to 0.86)	0.91 (0.88 to 0.93)	0.99 (0.99 to 0.99)	0.23 (0.19 to 0.27)	0.27 (0.16 to 0.41)	8.22 (6.83 to 10.18)	344 (11.4%)	2666 (88.6%)	3 to 1	120 to 1

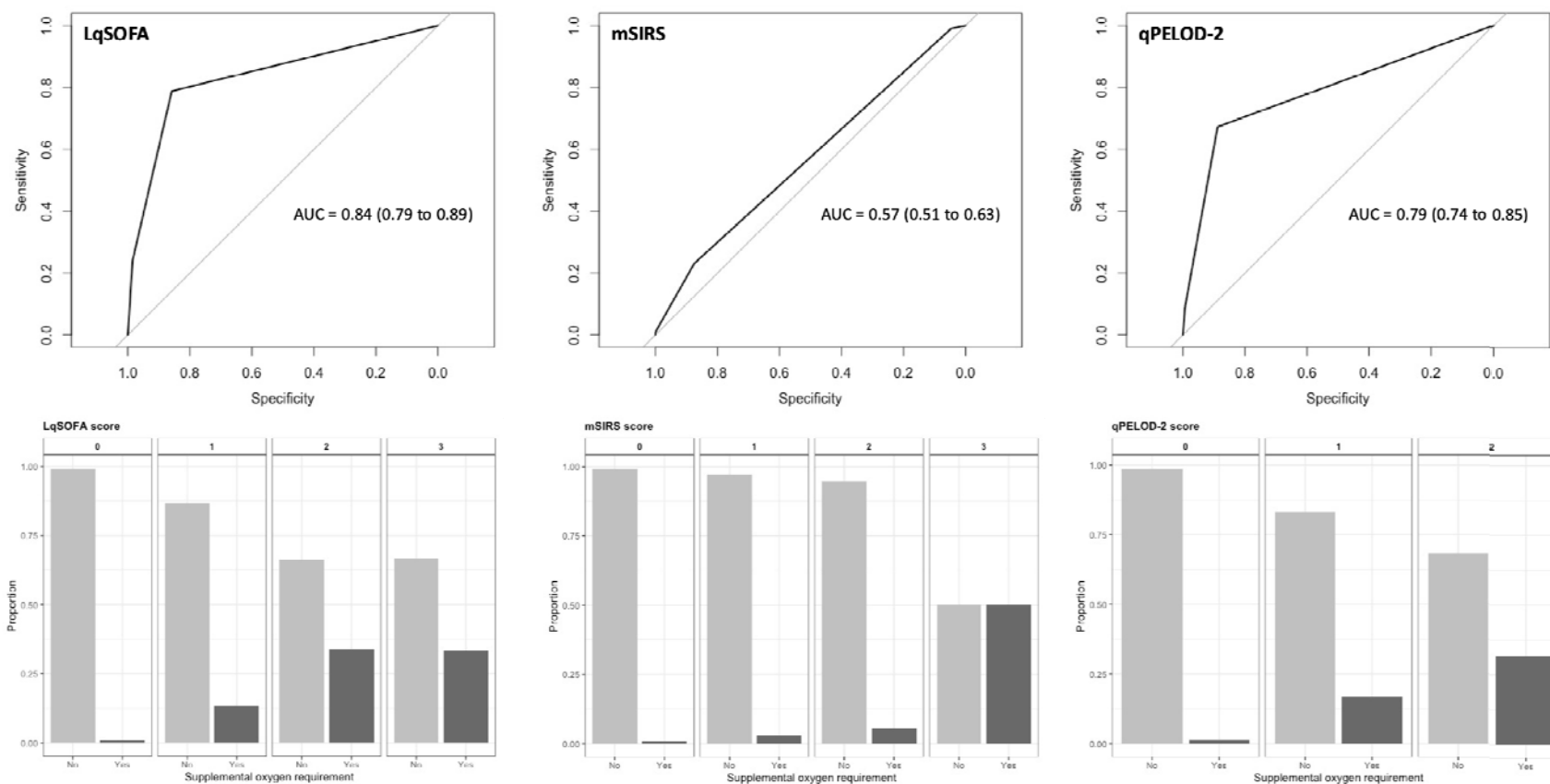
Referral threshold = 20%

LqSOFA	0.59 (0.45 to 0.69)	0.97 (0.96 to 0.97)	0.99 (0.98 to 0.99)	0.39 (0.32 to 0.45)	0.42 (0.32 to 0.56)	17.82 (13.83 to 23.17)	161 (5.3%)	2849 (94.7%)	2 to 1	68 to 1
qPELOD-2	0.56 (0.41 to 0.65)	0.97 (0.96 to 0.97)	0.98 (0.98 to 0.99)	0.37 (0.30 to 0.44)	0.46 (0.36 to 0.60)	16.81 (12.98 to 22.87)	153 (5.1%)	2857 (94.9%)	2 to 1	59 to 1
mSIRS	0.49 (0.37 to 0.61)	0.96 (0.95 to 0.97)	0.98 (0.98 to 0.99)	0.31 (0.26 to 0.38)	0.53 (0.40 to 0.65)	12.97 (9.85 to 19.44)	165 (5.5%)	2845 (94.5%)	2 to 1	51 to 1

Referral threshold = 40%

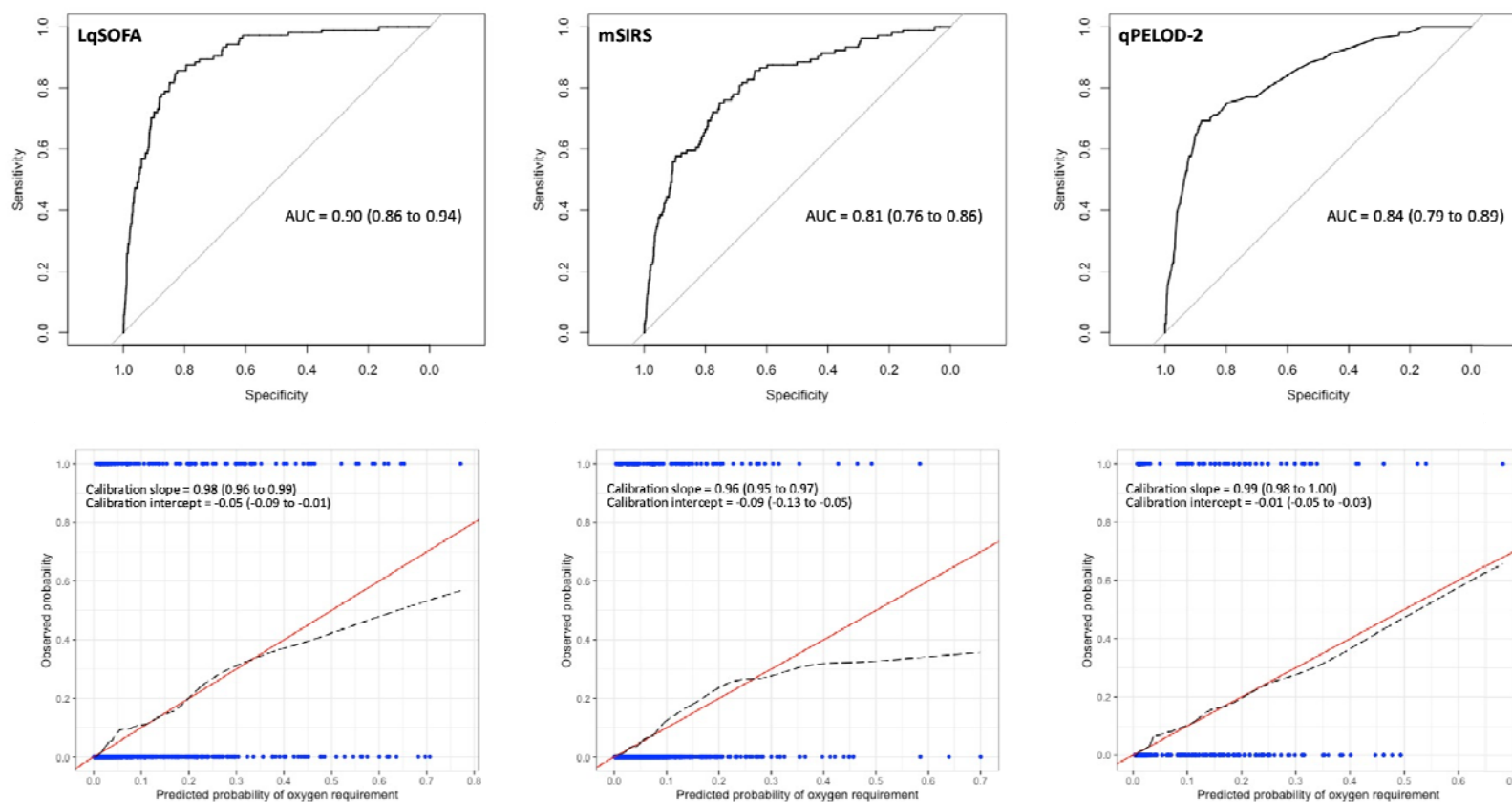
LqSOFA	0.28 (0.16 to 0.41)	0.99 (0.98 to 1.00)	0.97 (0.97 to 0.98)	0.49 (0.36 to 0.62)	0.73 (0.60 to 0.85)	27.50 (17.38 to 56.63)	62 (2.1%)	2948 (97.9%)	1 to 1	38 to 1
qPELOD-2	0.28 (0.13 to 0.41)	0.99 (0.98 to 0.99)	0.97 (0.97 to 0.98)	0.49 (0.35 to 0.59)	0.73 (0.59 to 0.87)	27.90 (16.46 to 47.56)	62 (2.1%)	2948 (97.9%)	1 to 1	39 to 1
mSIRS	0.21 (0.09 to 0.35)	1.00 (0.99 to 1.00)	0.97 (0.97 to 0.98)	0.61 (0.47 to 0.90)	0.80 (0.66 to 0.91)	Inf	20 (0.7%)	2990 (99.3%)	0 to 1	35 to 1

4 **FIGURE 1. Discrimination of the LqSOFA, mSIRS, and qPELOD-2 severity scores.** Receiver operating characteristic curve (ROC) for one
 5 imputed dataset shown. Variability in ROCs across multiply imputed datasets shown in Supplementary Figure 2. Pooled AUC reported. Bar
 6 plots showing risk scores against observed proportion of oxygen requirement using full case analysis: LqSOFA = 2,525 presentations (81 met
 7 primary outcome); mSIRS = 2,992 presentations (99 met primary outcome); qPELOD-2 = 2,531 presentations (83 met primary outcome).

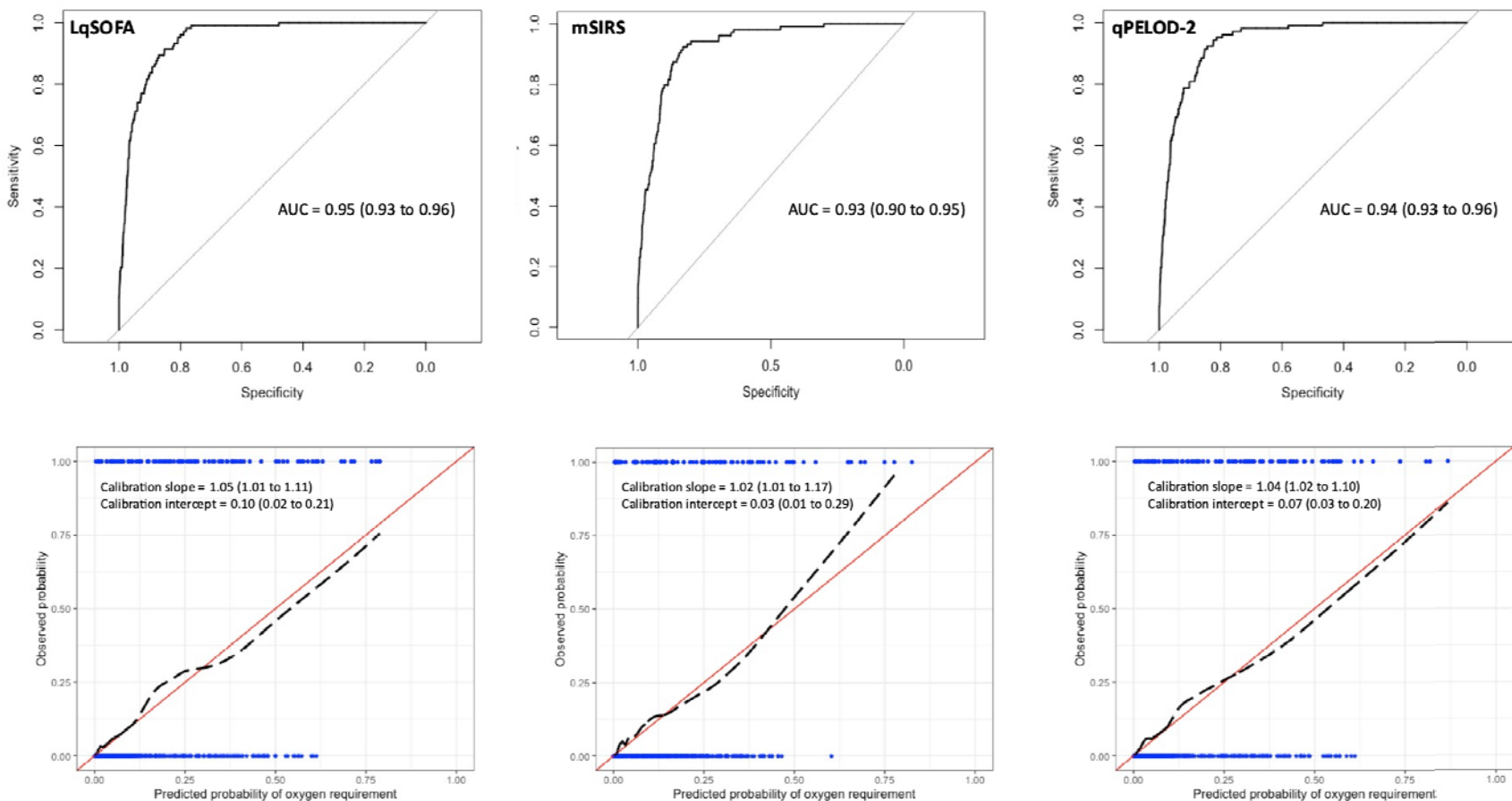


8

9 **FIGURE 2. Discrimination and calibration of the LqSOFA, mSIRS, and qPELOD-2 models.** Receiver operating characteristic curve
 10 (ROC) and calibration slope for one imputed dataset shown. Variability in ROCs and calibration slopes across multiply imputed datasets shown
 11 in Supplementary Figure 5. Pooled optimism-adjusted AUCs and calibration slopes reported (100 bootstrap samples). On calibration plots, red
 12 line indicates perfect calibration; black dashed line indicates calibration slope for that particular model; blue rug plots indicate distribution of
 13 predicted risks for participants who did (top) and did not (bottom) meet the primary outcome.

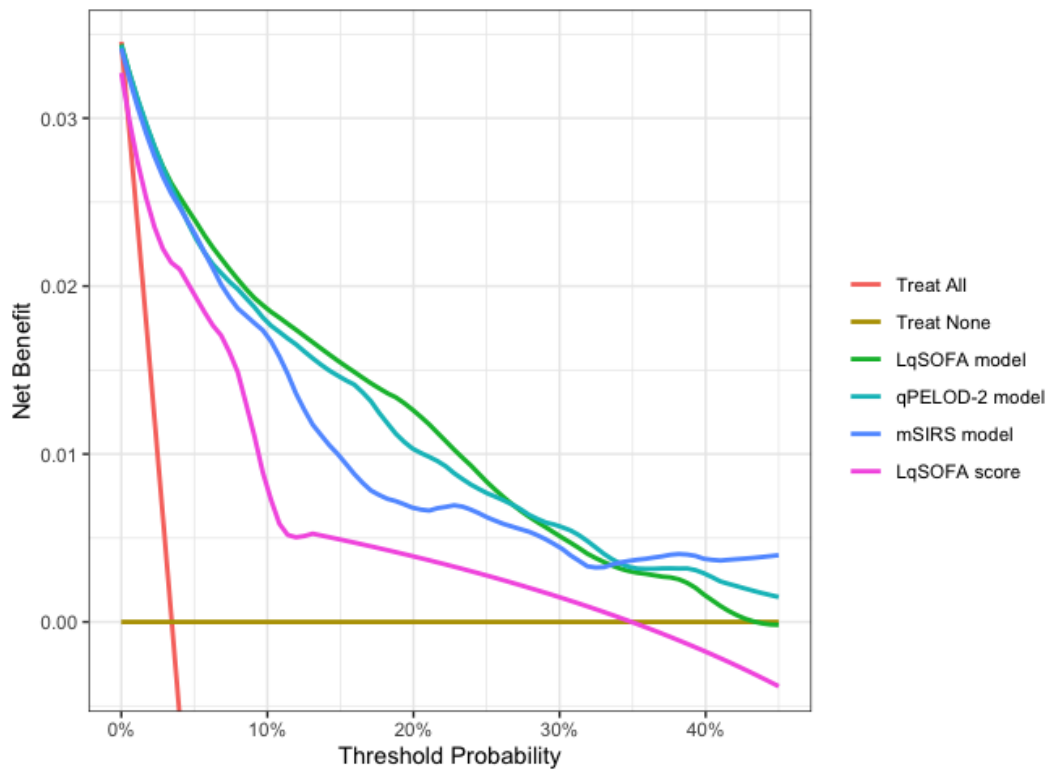


15 **FIGURE 3. Discrimination and calibration of updated LqSOFA, mSIRS, and qPELOD-2 models.** On calibration plots, red line indicates
 16 perfect calibration; black dashed line indicates calibration slope for that particular model; blue rug plots indicate distribution of predicted risks
 17 for participants who did (top) and did not (bottom) meet the primary outcome.



18
 19

20 **FIGURE 4. Decision curve analysis of the updated LqSOFA, mSIRS, and qPELOD-2 models.** The net benefit of the updated models (green
21 [LqSOFA], turquoise [qPELOD-2], and blue [mSIRS] lines) and original LqSOFA score (pink line), are compared to a “refer-all” (red line) and
22 “refer-none” (brown line) approach. A threshold probability of 5% indicates a management strategy whereby any child with a $\geq 5\%$ probability
23 of requiring oxygen is referred (i.e., a scenario where the value of one correct referral is equivalent to 19 incorrect referrals or a NNR of 20).
24



25