



**HAL**  
open science

## Multivariate evidence-based pediatric dengue severity prediction at hospital arrival

Kevin Bleakley, Camille Fortas, Veasna Duong, Sowath Ly, Tineke Cantaert, Heng Sothy, Denis Laurent, Philippe Buchy, Philippe Dussart, Anavaj Sakuntabhai

► **To cite this version:**

Kevin Bleakley, Camille Fortas, Veasna Duong, Sowath Ly, Tineke Cantaert, et al.. Multivariate evidence-based pediatric dengue severity prediction at hospital arrival. 2022. hal-03881145

**HAL Id: hal-03881145**

**<https://inria.hal.science/hal-03881145>**

Preprint submitted on 1 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multivariate evidence-based pediatric dengue severity prediction at hospital arrival

K. Bleakley<sup>†,\*1,2</sup>, C. Fortas<sup>3</sup>, V. Duong<sup>4</sup>, S. Ly<sup>3</sup>, T. Cantaert<sup>5</sup>, H. Sothy<sup>6</sup>, D. Laurent<sup>6</sup>, P. Buchy<sup>7</sup>, P. Dussart<sup>†,4,¶</sup>, A. Sakuntabhai<sup>†,8,9</sup>

<sup>1</sup>INRIA Saclay, France; <sup>2</sup>Laboratoire de Mathématiques d'Orsay, France; <sup>3</sup>Epidemiology and Public Health Unit, Institut Pasteur du Cambodge, Pasteur Network, 12201 Phnom Penh, Cambodia; <sup>4</sup>Virology Unit, Institut Pasteur du Cambodge, Pasteur Network, 12201 Phnom Penh, Cambodia; <sup>5</sup>Immunology Group, Institut Pasteur du Cambodge, Pasteur Network, 12201 Phnom Penh, Cambodia; <sup>6</sup>Kantha Bopha Hospital, Phnom Penh, Cambodia; <sup>7</sup>GlaxoSmithKline Vaccines R&D, Singapore; <sup>8</sup>Functional Genetics of Infectious Diseases Unit, Department of Genomes and Genetics, Institut Pasteur, 75015 Paris, France; <sup>9</sup>Centre National de la Recherche Scientifique, UMR 2000, 75015 Paris, France;

\* Corresponding author:

E-mail: [kevin.bleakley@inria.fr](mailto:kevin.bleakley@inria.fr)

† these authors contributed equally to this work.

¶ Present address: Institut Pasteur de Madagascar, Pasteur Network, 101 Antananarivo, Madagascar

26 **Abstract**

27 **Background.** For individuals with dengue-like symptoms arriving at hospitals, early  
28 detection of those likely to progress to—or not progress to—severe dengue can be of great  
29 use.

30

31 **Methods.** We studied 237 Cambodian children hospitalised in Kampong Cham hospital with  
32 dengue-like symptoms. Using dengue severity as primary endpoint, we ran univariate  
33 analyses and built multivariate random forest classifiers to predict this endpoint using early  
34 clinical and laboratory data.

35

36 **Findings.** In a random forest analysis using 56 available variables we obtained  $AUC = 0.94$ ,  
37 and for a sensitivity of 90%: specificity = 89%, positive predictive value (PPV) = 74%, and  
38 negative predictive value (NPV) = 96%. Platelet count, HDL cholesterol, and ultrasound  
39 pleural effusion and ascites were the four variables most associated with severe dengue  
40 outcomes. A random forest on only these four variables gave  $AUC = 0.88$ , and for a  
41 sensitivity of 90%: specificity = 82%, PPV = 64%, and NPV = 96%.

42

43 **Interpretation.** Future severe dengue with significant vascular leakage can be correctly  
44 predicted at hospital arrival in a large majority of cases using multivariate random forests. In  
45 addition to platelet count and ultrasound pleural effusion and ascites, HDL cholesterol level  
46 on the day of admission is also a strong predictor of severe dengue.

47

48 **Keywords:** dengue fever, dengue severity, random forest algorithm, Cambodia

49

50

51

## 52 **Introduction**

53 Dengue is an arboviral disease carried by mosquitos, leading to an estimated 390 million  
54 infections per year, including around 20,000 deaths.<sup>1</sup> Dengue is one disease entity with  
55 different clinical presentations and often unpredictable clinical evolution and outcome.<sup>2</sup> No  
56 specific therapies exist, and treatment essentially involves giving intravenous fluids and  
57 acetaminophen. There is one licensed dengue vaccine, though with limited use and efficacy.<sup>3-</sup>  
58 <sup>—5</sup> Dengue control relies on mosquito control strategies and social mobilisation campaigns,  
59 which are not necessarily efficient or sustainable, and thus the dengue virus can re-emerge or  
60 spread to uninfected areas.<sup>6,7</sup>

61

62 Dengue is associated with a large range of symptoms, from mild to life-threatening. Mild  
63 symptoms include headaches, retro-orbital pain, myalgia, arthralgia, rashes, and possibly  
64 minor haemorrhagic events. Extreme symptoms include vascular and plasma leakage, which  
65 often come with changes in haemostasis and thrombocytopenia.<sup>8</sup> If leakage is significant, it  
66 can lead to severe dengue (SD), including life-endangering dengue shock syndrome (DSS).

67

68 When individuals arrive at a hospital or clinic with a range of dengue-like symptoms, these  
69 clinical features do not always easily distinguish between severe and non-severe dengue  
70 cases.<sup>2</sup> In addition, it is not possible to know at hospital arrival if individuals have been  
71 previously exposed to the dengue virus, which would mean they possibly have a higher  
72 chance of developing a severe form of dengue. The question then is: who should be  
73 hospitalised and treated, and who can safely be sent home? The available literature suggests  
74 that it is difficult to find clinical, laboratory, or transcriptomic variables that are sufficiently  
75 predictive not to “miss” many individuals who later develop SD. Discrepancies between two

76 different ways to define dengue severity: the WHO 1997 <sup>9</sup> and WHO 2009 <sup>2</sup> classification  
77 criteria, have not helped.<sup>10,11</sup>

78

79 The WHO 1997 dengue case classification defines dengue fever (DF), dengue hemorrhagic  
80 fever (DHF), and dengue shock syndrome (DSS). It was developed in 1975 based on studies  
81 on Thai children in the 1950s and 1960s, then revised in 1997.<sup>9</sup> It looks for changes in  
82 haematocrit of more than 20% (percentage of cellular compartment in whole blood), a platelet  
83 count below  $100 \times 10^9/L$ , and signs of plasma leakage, e.g., ascites and pleural effusion, to  
84 define DHF and DSS. However, it remains difficult to apply prospectively in terms of DHF  
85 and DSS. Moreover, DHF criteria vary widely in frequency of occurrence. In reality, the  
86 existence of a continuous spectrum of dengue disease rather than two separate disease entities  
87 suggests that the WHO 1997 scheme acts as a longitudinal dengue-case classifier, using  
88 prerequisite clinical and/or biological signs to move from one clinical phase (e.g., DF) to the  
89 next (e.g., DHF).<sup>12,13</sup> Also, it turns out to be overly stringent: some patients showing  
90 symptoms of severe disease can be misclassified as non-severe cases if the rules are followed  
91 exactly.<sup>14,15</sup>

92

93 The more recent WHO 2009 dengue classification scheme includes appropriate triage and  
94 clinical management of patients, and considers that dengue is one disease entity with different  
95 clinical presentations that evolves with unpredictable outcomes.<sup>2,16</sup> It encompasses dengue  
96 without warning signs (DWoWS), dengue with warning signs (DWWS) requiring medical  
97 care, and severe dengue (SD). Although WHO 2009 is well-adapted to clinical management  
98 of dengue-hospitalised patients, the classification criteria are less strictly defined in it,  
99 opening the way for arbitrary interpretation. For instance, the list of warning signs is rather  
100 broad, ranging from subjective symptoms (such as abdominal pain or tenderness, persistent

101 vomiting, lethargy, restlessness) and objective signs (mucosal bleeding, liver enlargement,  
102 clinical fluid accumulation) to laboratory results (increase in haematocrit concurrent with  
103 rapid decrease in platelet count), and due to the unpredictable possibility of evolution towards  
104 severe dengue, this broadness can result in unnecessary admission of mild dengue cases. In  
105 addition, even dengue patients without warning signs may develop severe dengue.<sup>9</sup>

106

107 Many groups have attempted to build models to predict future dengue severity, with mixed  
108 results.<sup>17–19</sup> Among the more successful, Potts *et al.*<sup>20</sup> used classification and regression trees  
109 to predict DSS in Thai children, with high sensitivity (97%) but low specificity (48%) using  
110 5-fold cross-validation under the WHO 1997 criteria. In contrast, Pang *et al.*<sup>21</sup> used the WHO  
111 2009 criteria, with data involving both clinical variables and RNA microarray and protein  
112 assays. However, instead of trying to predict future SD, they aimed to predict whether  
113 patients would develop DWWS or not. They obtained sensitivity and specificity respectively  
114 of up to 83% and 84% on the original dataset, and up to 96% and 55% on a validation set.

115

116 Two recent articles provide more focused attempts to predict future SD cases using early  
117 clinical data. First, Lee *et al.*<sup>22</sup> used 85% of a retrospective cohort of 69 SD and 1184 non-SD  
118 adult patients from Taiwan to build logistic regression models, which were then tested on the  
119 remaining 15% of the cohort. The most significant variables, on which their models were  
120 based, were: age>65, minor gastrointestinal bleeding, leukocytosis, and platelet count below  
121  $100 \times 10^9/L$ . On the training set, they obtained an AUC (area under ROC curve) of 0.85 and  
122 sensitivity and specificity of 70% and 91% respectively. On the test set, they had an AUC of  
123 0.90 and sensitivity and specificity of 89% and 76% respectively. The large difference  
124 between the training and testing results indicates possible problems in generalising the model  
125 to new patients. Logistic models also can miss interesting interactions between variables

126 which can be picked up by tree-based classifiers such as CART and random forest. As child  
127 and adult clinical manifestations of dengue fever are known to differ,<sup>23–26</sup> one should also be  
128 careful when comparing results between studies involving different age groups.

129

130 Second, Nguyen *et al.*<sup>17</sup> performed a prospective study on 7,563 Vietnamese children  
131 presenting at outpatient clinics, of which 2,060 had lab-confirmed dengue and 117 reached  
132 the WHO 2009 criterion for SD. Using multivariate logistic modeling on up to 16 pre-defined  
133 variables, they discriminated between the 117 SD and the other 7,446 patients, obtaining an  
134 AUC of 0.95, sensitivity and specificity of 87% and 88% respectively, and positive (PPV)  
135 and negative (NPV) predictive values of 10% and 99.8%. Their final model included history  
136 of vomiting, platelet count, aspartate aminotransferase (AST) level, and NS1 rapid test status.

137

138 In this study, we have investigated a cohort of Cambodian children hospitalised at Kampong  
139 Cham City Provincial Hospital with clinical presentation of dengue-like symptoms during  
140 2011-2012. Our goals were: (i) detect variables individually associated with SD including  
141 significant vascular leakage, (ii) build a quality multivariate classifier to predict dengue  
142 severity (as defined by the WHO 2009 criteria) using early clinical and laboratory data of  
143 hospitalised dengue-suspected patients, and (iii) assess the performance of this predictive  
144 model.

145

## 146 **Methods**

147 **Data collection.** Patients presenting at Kampong Cham City Provincial Hospital with dengue-  
148 like illness during the two consecutive dengue epidemic seasons of June to October during  
149 2011 and 2012 were enrolled. Inclusion criteria were children between 2 and 15 years old  
150 who had fever or history of fever at presentation, and onset of at least two of the following

151 symptoms within the previous 72 hours: headache, retro-orbital pain, muscle pain, joint pain,  
152 rash, or any bleeding signs. Exclusion criteria were symptoms inconsistent with dengue,  
153 obvious non-dengue acute infections (e.g., otitis media, pneumonia, meningitis), or a known  
154 chronic illness. Briefly, for all included patients, the first visit was conducted shortly after  
155 admission and whole blood was collected for confirming the dengue diagnosis and obtaining  
156 biochemical and haematological parameters. The day of fever onset was defined as day 0 of  
157 illness and clinical signs related to dengue disease were recorded according to the WHO 2009  
158 classification. A clinical and biological follow-up was conducted during their hospitalisation  
159 at defervescence and at discharge as described in Dussart *et al.*<sup>15</sup> Ultrasonographic  
160 examination of the abdomen was also systematically performed at the three time points to  
161 identify any evidence of plasma leakage, severe dengue, or recovery. Finally, all dengue  
162 patients were classified using an adapted WHO 2009 classification for SD case definition as  
163 described in Dussart *et al.*<sup>15</sup>

164

165 **Dengue diagnosis and classification of dengue cases.** The laboratory dengue diagnosis was  
166 performed as previously described in Dussart *et al.*<sup>15</sup>; see also Supplementary Methods S4  
167 and further references.<sup>27–31</sup> We then placed patients into one of two groups according to these  
168 diagnoses: dengue-infected patient or non-dengue patient, as shown in Supplementary Table  
169 S1. Next, classification as SD or non-SD patient for acute or recent dengue cases was  
170 performed according to an adapted WHO 2009 classification for SD criteria using clinical,  
171 biological and ultrasound data recorded at admission, defervescence, and discharge as  
172 previously described in Dussart *et al.*<sup>15</sup> (Supplementary Methods S4). For the purposes of this  
173 study, and as was done in a previous study,<sup>17</sup> non-dengue patients and infected non-SD  
174 patients were then combined to form the non-SD class.

175



176 **Variable filtering.** We focused on variables of physiological interest including a large  
177 number of clinical and laboratory variables from the original dataset at admission and basic  
178 demographic variables; this gave 56 variables (Supplementary Table S2). Variables were  
179 either continuous or binary categorical (mostly presence/absence of a given symptom).  
180 Although the immune status (primary versus secondary dengue) may have an impact on the  
181 course or outcome of the disease, as we were only working from clinical and biological data  
182 available at hospital admission, we chose not to include this variable. Extremely rare cases of  
183 missing data were imputed using mean imputation or linear regression.

184

185 **Univariate statistical analysis.** Univariate analyses were performed on each of the dataset's  
186 original 56 variables using Student's *t*-test for continuous variables and Fisher exact tests for  
187 categorical variables (N=237).

188

189 **Multivariate analysis: random forest.** A random forest is a collection of standard  
190 hierarchical decision trees.<sup>31</sup> A forest's prediction for a new patient is simply the average  
191 prediction across all trees. Each tree in the forest is constructed exactly like a standard  
192 decision tree is, on a training set, except that a *random* subset of the patients is used as the  
193 training set to construct each tree, and furthermore the "best" variable at each split is not  
194 chosen from all available variables, but from a *random* subset of them. This tends to create a  
195 forest that outperforms a single decision tree in terms of prediction accuracy.<sup>31</sup> Since for each  
196 tree a random subset of around one third of the patients is left-out when constructing it, these  
197 patients can be used as a validation set for that specific tree. No supplementary validation set  
198 is therefore required.

199

200 In this article, we ran the random forest<sup>31</sup> algorithm using the randomForest package in R<sup>32</sup>  
201 with 20,000 trees and otherwise default parameters on the full set of 56 variables. Unlike  
202 logistic regression, no preselection of a small number of variables is required for stable  
203 modeling, meaning that potentially important interactions between variables are not thrown  
204 away before modeling begins. The trade-off is that detection of important variables remains  
205 more intuitive when using logistic regression, though attempts have been made to improve  
206 variable selection for random forests.<sup>33</sup>

207

208 **Prediction.** We built a random forest on the original dataset of 62 SD patients as previously  
209 defined (Ref. 15) (labelled 1) and 175 febrile non-SD, of which 128 were eventually found to  
210 be dengue and 47 non-dengue patients, both labelled 0. As dengue status at admission was  
211 only suspected, it would be artificial to remove the *a posteriori* non-dengue patients when  
212 building a SD prediction model at hospital admission. This same choice follows Nguyen *et*  
213 *al.*<sup>17</sup> In the same way, we did not excluded children at admission who had moderate or  
214 abundant fluid on their first scan, but considered all clinical and biological signs observed at  
215 the time of admission, regardless the final dengue status. The regression version of random  
216 forest outputs a prediction between 0 and 1 for each of the randomly left-out patients of each  
217 tree. The final prediction for each individual is then the *average* of these predictions over all  
218 trees in which they were left-out. Our main results (specificity, positive and negative  
219 predictive values) are reported with respect to a predefined sensitivity of 90%. The area under  
220 the ROC curve (AUC) is also reported. We further report results when the sensitivity is set to  
221 85% and 95%, as well as an overall sensitivity-specificity plot.

222

223 **Results**

224 **Univariate statistical analyses.** Supplementary Table S2 presents demographic data and  
 225 shows that around one third of the variables were statistically significant at the 0.01 level. The  
 226 four most significant variables were ascites via ultrasound, pleural effusion via ultrasound,  
 227 HDL cholesterol, and platelet count (all  $p < 0.0001$ ).

228

229 **Multivariate model.** Using the random forest algorithm with internal out-of-bag validation  
 230 on the full set of 56 variables (see Methods), we obtained: AUC = 0.94 (95%CI: [0.93, 0.95])  
 231 and for a sensitivity of 90%, the specificity was 89% (95%CI: [85, 91]) (Table 1). The PPV  
 232 was 74% (95%CI: [67, 79]) and the NPV was 96% (95%CI: [96, 96]). We also built and  
 233 tested a random forest on the model with the four most individually statistically significant  
 234 variables (Table 1).

235

236 **Table 1. Predicting severe dengue using random forest modeling**

Model	AUC	Sensitivity	Specificity	PPV	NPV
(ascites, pleural effusion, HDL, platelet count)	0.88 [0.88, 0.89]	90%	82% [78, 84]	64% [60, 67]	96% [96, 96]
multivariate (56 variables, 90% sensitivity)	<b>0.94</b> [0.93, 0.95]	90%	89% [85, 91]	74% [67, 79]	96% [96, 96]
multivariate (56 variables, 85% sensitivity)	<b>0.94</b> [0.93, 0.95]	85%	<b>94%</b> [89, 95]	<b>84%</b> [74, 87]	95% [95, 95]
multivariate (56 variables, 95% sensitivity)	<b>0.94</b> [0.93, 0.95]	95%	59% [48, 76]	45% [39, 58]	<b>97%</b> [97, 98]
multivariate (56 variables, $\leq 4$ days)	0.92 [0.90, 0.93]	90%	73% [68, 82]	46% [42, 57]	<b>97%</b> [96, 97]

237

238 Out-of-bag prediction results for three random forest models. AUC = area under ROC curve; PPV = positive  
239 predictive value; NPV = negative predictive value. [a, b] corresponds to a 95% CI. Units: HDL cholesterol:  
240 mmol/L; Platelets:  $10^9/L$ . Best values are shown in bold.

241  
242  
243 To further compare univariate and multivariate prediction performances, we calculated and  
244 plotted the sensitivity and specificity across all decision cut-offs between 0 and 1; see Fig. 1.

245

246

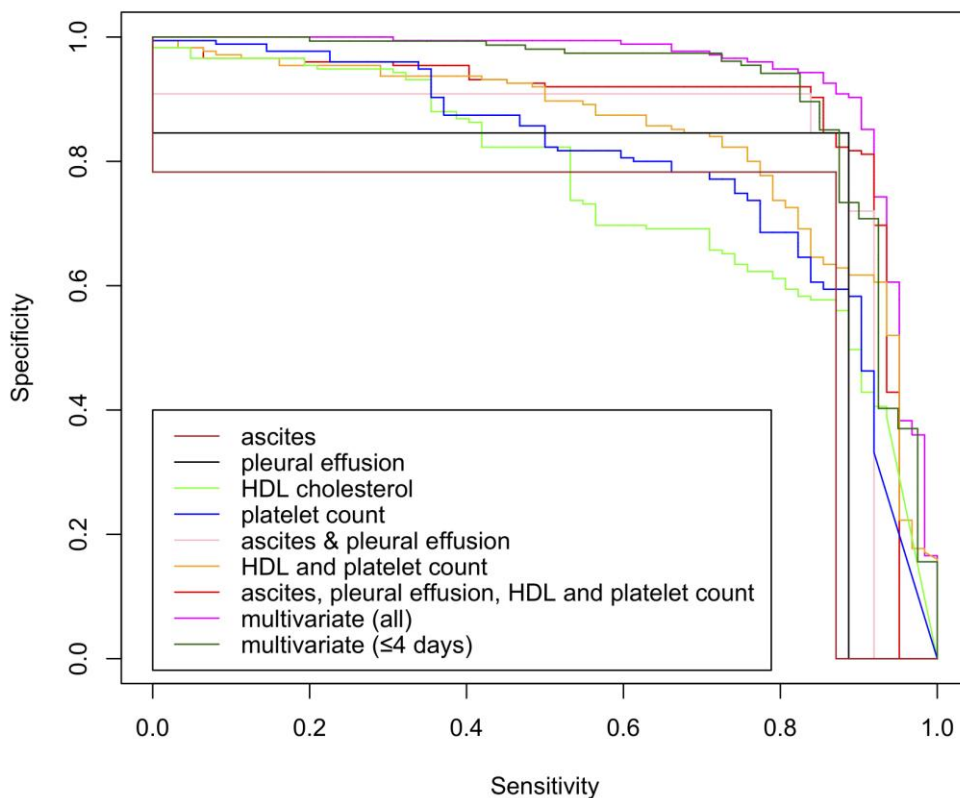
247

248

249

250

251 **Fig 1. Inter-model comparison of sensitivity vs specificity.**



252

253 **Analysis of SD individuals using adapted WHO 2009 classification predicted non-SD.**

254 Among the 62 SD patients in the Kampong Cham cohort, six were predicted non-SD by the  
255 random forest model when the sensitivity was fixed at 90%. All six of these dengue-  
256 confirmed patients were hospitalised between one and four days after onset of fever. All of  
257 them developed significant vascular leakage (as defined by clinical and biological variables)  
258 at defervescence and/or post-defervescence phases, characterised by detection of ascites and  
259 moderate pleural effusion by ultrasound and signs of severe plasma leakage.<sup>2</sup> Demographic,  
260 clinical and biological characteristics of these six SD patients with significant vascular  
261 leakage at admission, defervescence and/or post-defervescence phases are presented in  
262 Supplementary Table S3.

263

264 **At most four days after fever.** Recent studies have attempted to limit prediction of SD to  
265 individuals arriving at clinics/hospitals at most three or four days after fever onset.<sup>17,22</sup> To see  
266 whether inclusion of patients arriving later than four days after fever onset had any influence  
267 on our results, we ran a random forest algorithm on the subset of 194 individuals arriving at  
268 most four days after fever onset from our own study; this included 40 SD and 154 non-SD.  
269 The AUC dropped minimally to 0.92 while for a fixed sensitivity of 90%, the specificity fell  
270 slightly to 73% (95%CI: [68, 82]).

271

272 **Comparison with previous results.** We performed a comparison of our model and results  
273 with the state-of-the-art results in Nguyen *et al.*<sup>17</sup> First, their reported sensitivity, specificity,  
274 PPV and NPV results on their data were broadly similar to ours, despite involving a different  
275 model, country (Vietnam), predictive variables, and ratio of SD to non-SD patients. Second,  
276 we tested several scenarios using both our random forest and their (logistic) model on the  
277 Cambodian data for various sets of variables (Table 2). The four variables selected by Nguyen

278 *et al.*<sup>17</sup> (history of vomiting, platelet count, AST level, and NS1 rapid test status) performed  
279 acceptably on the data under both models, suggesting that these variables generalise to  
280 entirely new data. Our four-variable set performed better on the Cambodian data (for both  
281 random forests and logistic regression). It was interesting to note that the prediction results  
282 using *only platelet count* for a univariate logistic regression on the Cambodian data were  
283 essentially indistinguishable from the results for the four-variable model from Nguyen *et al.*<sup>17</sup>  
284 (Table 2), suggesting that perhaps most if not all of the predictive power in their model came  
285 from platelet count alone. Note finally that the full 56-variable random forest strongly  
286 outperformed all other models in predicting SD (Table 2).

287

288 **Table 2. Model comparisons**

Study	Variables	Model	AUC	Sensitivity	Specificity	PPV	NPV
-------	-----------	-------	-----	-------------	-------------	-----	-----

Nguyen <i>et al.</i> <sup>17</sup>	4 N2017	RF	0.85 [0.84, 0.87]	90%	63% [55, 67]	42% [41, 50]	95% [94, 95]
	4 N2017	logistic	0.87 [0.85, 0.87]	90%	65% [60, 67]	48% [44, 50]	95% [95, 95]
	Platelet	logistic	0.86 [0.85, 0.86]	90%	62% [57, 64]	46% [43, 47]	95% [94, 95]
Present study	4 B2021	RF	0.88 [0.88, 0.89]	90%	82% [78, 84]	64% [60, 67]	<b>96%</b> [96, 96]
	4 B2021	logistic	0.90 [0.89, 0.91]	90%	83% [75, 84]	65% [57, 67]	<b>96%</b> [96, 96]
	Full B2021	RF	<b>0.94</b> [0.93, 0.95]	90%	<b>89%</b> [85, 91]	<b>74%</b> [67, 79]	<b>96%</b> [96, 96]

289

290 AUC = area under ROC curve; PPV = positive predictive value; NPV = negative predictive value; RF = random

291 forest; 4 N2017 = the four variables from Nguyen *et al.*<sup>17</sup>: history of vomiting, platelet count, AST level, and

292 NS1 rapid test status; 4 B2021 = the four variables from the present study: ultrasound ascites, ultrasound pleural

293 effusion, HDL cholesterol level, and platelet count; full B2021 means the full 56-variable model from the

294 present study. [a, b] corresponds to a 95% CI. Best values are shown in bold.

295

296

297

298

299 **Discussion**

300 We have proposed a random forest model that is very effective in predicting future SD using  
301 an adapted WHO 2009 classification for SD case definition in Cambodian children at hospital  
302 arrival. Our results are comparable and often better than those in recent previous studies<sup>17,22</sup> in  
303 terms of AUC, sensitivity, specificity, and PPV and NPV. Furthermore, as Fig. 1 showed,  
304 moving from univariate to four variables and then to a full multivariate model significantly  
305 improved specificity for a fixed sensitivity (and vice versa) over a large range of sensitivities  
306 (60% to 90%) and specificities (70% to 90%), demonstrating the distinct advantage of a full  
307 multivariate random forest approach for predicting SD including significant vascular leakage.

308

309 There remain major complicating factors in dengue severity prediction. For instance, children  
310 and adults are not necessarily affected by dengue in the same way and with the same  
311 symptoms, making cross-comparisons between children and adults difficult.<sup>23—26,34,35</sup> We do  
312 not expect our results to necessarily generalise to adults. Further, it appears that the fraction of  
313 patients at clinic or hospital arrival with dengue-like symptoms who eventually develop  
314 severe dengue varies between countries, even in the same region, due to different virus  
315 serotypes/genotypes, human genetics, and clinical management.<sup>36</sup>

316

317 In our study, in addition to platelet count (note that below  $100 \times 10^9/L$  on day of admission  
318 gave 85% sensitivity for 66% specificity), which has been used in both WHO dengue case  
319 classifications, we identified three other highly univariately-significant prognostic markers:  
320 HDL cholesterol level, pleural effusion, and ascites, the latter two detected by  
321 ultrasonographic examination at admission. On their own, each performed fairly well for  
322 certain sensitivity/specificity trade-offs, rather than for a wide range of them. Using the four  
323 together to learn a random forest was better, though not as good as using all 56 variables. This



324 suggests that there is a trade-off to be made between interpretability (simple model) and  
325 predictive performance (full multivariate random forest model).

326

327 It has previously been noted that serial ultrasonography allows better identification of patients  
328 at risk of developing severe dengue.<sup>37</sup> Furthermore, we have demonstrated here that  
329 ultrasonographic examination at admission to detect extravascular fluid in lung and  
330 abdominal cavities improved prediction with respect to platelet count alone. Considering the  
331 affordable cost of mobile ultrasonogram units and their usefulness for other diseases, we  
332 recommend that this non-invasive and inexpensive procedure be used routinely, not  
333 exclusively in referral hospitals for overall assessment during the febrile phase, but in mid-  
334 level healthcare facilities as well, which are responsible for emergency and ambulatory triage  
335 assessment and treatment in endemic areas. Moreover, we suggest that the use of  
336 ultrasonographic examinations and training should also be included in general guidelines for  
337 dengue patient management.

338

339 Previous studies have shown a relationship between circulating lipids and dengue virus  
340 infection. Indeed, lower levels of plasma and serum cholesterol in severe compared to less  
341 severe dengue cases or healthy controls have been observed.<sup>37–40</sup> Total cholesterol is  
342 comprised of low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL)  
343 cholesterol, and to a lesser extent very low-density lipoprotein (VLDL) cholesterol and  
344 triglycerides. The assembly of NS1-HDL complexes triggers the production of pro-  
345 inflammatory cytokines in human primary macrophages while NS1 protein in complex with  
346 HDL and LDL particles was detected in the plasma of hospitalized dengue patients (Ref.).  
347 However, the relationships between severe dengue and total, HDL, and LDL cholesterol,  
348 respectively, remain unclear. In previous studies, total cholesterol and LDL cholesterol levels

349 decreased over the course of illness and were generally lower with increasing dengue severity,  
350 suggesting that LDL cholesterol levels might be more specific to dengue infection than HDL  
351 ones.<sup>41,42</sup> However, in this study, we have demonstrated that a low HDL cholesterol level,  
352 which has been observed in previous studies,<sup>15,38</sup> can indeed be used as a prognostic marker.  
353 Furthermore, even though we did not measure the LDL cholesterol level directly in our cohort  
354 (instead, we used the calculation method of Friedewald *et al.*<sup>45</sup>), this was indeed highly  
355 significantly related to severe dengue also (Supplementary Table S2). Finally, we remark that  
356 the role of lipid metabolism in the pathogenesis of dengue has also been identified in genetic  
357 studies.<sup>43,44</sup> However, the underlying mechanisms involved still require further investigation.

358

359 One further important issue in dengue severity prediction is the desire for “early” prediction.  
360 This is why more recent studies have focused on individuals arriving for treatment after at  
361 most three or four days of fever.<sup>17,22</sup> Though our study also included patients arriving later  
362 than this, we showed that on those arriving early (at most four days after fever) the predictive  
363 performance dropped only slightly. Note also that it is not realistic to expect that all patients  
364 will come to a clinic/hospital after only one or two days of fever, and thus only focusing on  
365 these patients may in fact add new biases to an analysis.

366

367 We also took a closer look at the six SD patients who were predicted to be non-SD by the  
368 random forest algorithm, to see if we could gain any knowledge as to why. We found that five  
369 of the individuals arrived at the hospital shortly after fever onset, and for all intents and  
370 purposes had clinical and laboratory data that strongly resembled that of non-SD individuals  
371 at early stages. The sixth patient was admitted later (day 5 after onset of fever) and presented  
372 clinical and laboratory signs suggestive of SD with significant vascular leakage on admission.  
373 Using an adapted WHO 2009 classification for SD case definition, this case therefore

374 represents the border between ‘obviously SD’ and ‘not obviously SD’. Generally speaking, it  
375 appears likely that some very-early dengue patients simply do not exhibit differences, in terms  
376 of currently-used clinical and laboratory variables, from non-SD patients.

377

378 The search for other laboratory markers and/or transcriptional markers<sup>46</sup> to complement or  
379 replace clinical variable studies therefore remains an important goal, as does transferring our  
380 prediction algorithm to practice for improving triage.

381

## 382 **Declaration**

383 *Ethics approval and consent to participate.* The study was approved by the Cambodian  
384 National Ethics Committee for Human Research (approval #087NECHR/2011). All patient  
385 inclusion and blood sampling happened only after obtaining written informed consent from  
386 the patient’s parents or legal guardians. The study was carried out in accordance with the  
387 ethical guidelines of the Pasteur Institute, Paris.

388

389 *Consent for publication.* Not applicable.

390

391 *Availability of data and materials.* The dataset used during the current study is available from  
392 the corresponding author on reasonable request.

393

394

395 *Competing interests.* Dr. Philippe Buchy is a former Head of Virology at Institut Pasteur du  
396 Cambodge and is currently an employee of GSK Vaccines, Singapore. The other authors  
397 declare no conflict of interest.

398

399 *Funding.* Data generation was funded by BioMérieux ([www.biomerieux.com](http://www.biomerieux.com)) and Institut  
400 Pasteur du Cambodge. This study received funding from the French government's  
401 *Investissement d'Avenir* program, Laboratoire d'Excellence "Integrative Biology of Emerging  
402 Infectious Diseases" (grant: ANR-10-LABX\_62°IBEID). TC is supported by a Wellcome-  
403 HHMI International Scholar award (Grant n°208710/Z/17/Z – [www.wellcome.ac.uk](http://www.wellcome.ac.uk)). The  
404 funders had no role in the study design, data collection and data analysis, decision to publish,  
405 or preparation of the manuscript.

406  
407 *Authors contributions.* KB analysed the data and implemented the statistical tests and  
408 prediction algorithms. PD, VD and CF verified the underlying data and classified patient  
409 status based on a precise interpretation of the WHO 2009 classification scheme. PD, VD, CF,  
410 TC, HS, DL, SL, PB, and AS provided input on all biological issues in dengue fever. AS  
411 coordinated the project. KB, PD, and AS wrote the manuscript. All authors read and approved  
412 the final manuscript.

413  
414 *Acknowledgements.* Not applicable

415

## 416 **References**

- 417 1. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The  
418 global distribution and burden of dengue. *Nature*. 2013;496(7446):504-7.
- 419 2. World Health Organization. *Dengue, guidelines for diagnosis, treatment, prevention*  
420 *and control*. Geneva: WHO. 2009.
- 421 3. Capeding MR, Tran NH, Hadinegoro SR, Ismail HI, Chotpitayasunondh T, Chua MN,  
422 et al. Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy

- 423 children in Asia: a phase 3, randomised, observer-masked, placebo-controlled trial.  
424 Lancet. 2014;384(9951):1358-65.
- 425 4. Hadinegoro SR, Arredondo-García JL, Capeding MR, Deseda C, Chotpitayasunondh  
426 T, Dietze Ret, et al. Efficacy and long-term safety of a dengue vaccine in regions of  
427 endemic disease. N Engl J Med. 2015;373(13):1195-206.
- 428 5. Sridhar S, Luedtke A, Langevin E, Zhu M, Bonaparte M, Machabert T, et al. Effect of  
429 Dengue Serostatus on Dengue Vaccine Safety and Efficacy. N Engl J Med.  
430 2018;379:327-340.
- 431 6. Gubler DJ. Dengue, Urbanization and globalization: the unholy trinity of the 21st  
432 century. Trop Med Health. 2011;39(4 suppl):3–11.
- 433 7. Achee NL, Gould F, Perkins TA, Reiner RC Jr, Morrison AC, Ritchie SA, et al. A  
434 critical assessment of vector control for dengue prevention. PLoS Negl Trop Dis.  
435 2015;9(5):e0003655.
- 436 8. Simmons CP, Farrar JJ, van Vinh Chau N, Wills B. Dengue. N Engl J Med.  
437 2012;366(15):1423-32.
- 438 9. World Health Organization. Dengue hemorrhagic fever: diagnosis, treatment,  
439 prevention and control. Geneva: WHO. 1997.
- 440 10. Horstick O, Ranzinger SR. Reporting progress on the use of the WHO 2009 dengue  
441 case classification: a review. Southeast Asian J Trop Med Public Health. 2015;46  
442 Suppl 1:49–54.
- 443 11. Gan VC, Lye DC, Thein TL, Dimatatac F, Tan AS, Leo Y-S. Implications of  
444 discordance in world health organization 1997 and 2009 dengue classifications in  
445 adult dengue. PloS One. 2013;8(4):e60946.
- 446 12. Libraty, D.H., Endy, T.P., Houn, H.S., Green, S., Kalayanarooj, S., Suntayakorn, S.,  
447 Chansiriwongs, W., Vaughn, D.W., Nisalak, A., Ennis, F.A., Rothman, A.L., 2002.

- 448 Differing influences of virus burden and immune activation on disease severity in  
449 secondary dengue-3 virus infections. *J Infect Dis* 185, 1213–21.
- 450 13. Phuong, C.X.T., Nhan, N.T., Kneen, R., Thuy, P.T.T., van Thien, C., Nga, N.T.T., Thuy,  
451 T.T., Solomon, T., Stepniewska, K., Wills, B., Dong Nai Study Group, 2004. Clinical  
452 diagnosis and assessment of severity of confirmed dengue infections in  
453 Vietnamese children: is the world health organization classification system  
454 helpful? *Am. J. Trop. Med. Hyg.* 70, 172–179.
- 455 14. Deen JL, Harris E, Wills B, Balmaseda A, Hammond SN, Rocha C, et al. The WHO  
456 dengue classification and case definitions: time for a reassessment. *Lancet.*  
457 2006;368:170–3.
- 458 15. Dussart P, Duong V, Bleakley K, Fortas C, Lorn Try P, Kim KS, et al. Comparison of  
459 dengue case classification schemes and evaluation of biological changes in different  
460 dengue clinical patterns in a longitudinal follow-up of hospitalized children in  
461 Cambodia. *PLOS Negl Trop Dis.* 2020; 14(9): e0008603
- 462
- 463 16. Alexander N, Balmaseda A, Coelho ICB, Dimaano E, Hien TT, Hung NT, et al.  
464 Multicentre prospective study on dengue classification in four South-east Asian and  
465 three Latin American countries. *Trop Med Int Health.* 2011;16(8):936–48
- 466 17. Nguyen MT, Ho TN, Nguyen VVC, Nguyen TH, Ha MT, Ta VT, et al. An evidence-  
467 based algorithm for early prognosis of severe dengue in the outpatient setting. *Clin*  
468 *Infect Dis* 2017;64(5):656-653.
- 469 18. Tanner L, Schreiber M, Low JG, Ong A, Tolfvenstam T, Lai YL, et al. Decision tree  
470 algorithms predict the diagnosis and outcome of dengue fever in the early phase of  
471 illness. *PLoS Negl Trop Dis* 2008;2(3):e196.
- 472 19. Phakhounthong K, Chaovalit P, Jittamala P, Blacksell SD, Carter MJ, Turner P, et al.  
473 Predicting the severity of dengue fever in children on admission based on clinical

- 474 features and laboratory indicators: application of classification tree analysis. *BMC*  
475 *Pediatr.* 2018; 18:109.
- 476 20. Potts JA, Gibbons RV, Rothman AL, Srikiatkachorn A, Thomas SJ, Supradish PO, et  
477 al. Prediction of dengue disease severity among pediatric Thai patients using early  
478 clinical laboratory indicators. *PLoS Negl Trop Dis.* 2010;4(8):e769.
- 479 21. Pang J, Lindblom A, Tolfvenstam T, Thein TL, Naim AN, Ling L, et al. Discovery  
480 and validation of prognostic biomarker models to guide triage among adult dengue  
481 patients at early infection. *PloS One.* 2016;11(6):e0155993.
- 482 22. Lee K, Liu JW, Chen YH, Chen YC, Tsai CY, Huang SY, Lin CY, et al. Development  
483 of a simple clinical risk score for early prediction of severe dengue in adult patients.  
484 *PloS One.* 2016;11(5):e0154772.
- 485 23. Kalayanaroj S, Nimmannitya S. Clinical presentations of dengue hemorrhagic fever  
486 in infants compared to children. *J Med Assoc Thail Chotmai het Thangphaet.* 2003;86  
487 *Suppl 3:S673-680.*
- 488 24. Wichmann O, Hongsiriwon S, Bowonwatanuwong C, Chotivanich K, Sukthana Y,  
489 Pukrittayakamee S. Risk factors and clinical features associated with severe dengue  
490 infection in adults and children during the 2001 epidemic in Chonburi, Thailand. *Trop*  
491 *Med Int Health.* 2004;9:1022–9.
- 492 25. Hammond SN, Balmaseda A, Perez L, Tellez Y, Saborio SI, Mercado JC, et al.  
493 Differences in dengue severity in infants, children, and adults in a 3-year hospital-  
494 based study in Nicaragua. *Am J Trop Med Hyg.* 2005;73:1063–70.
- 495 26. Thomas L, Moravie V, Besnier F, Valentino R, Kaidomar S, Coquet LV, et al.  
496 Clinical presentation of dengue among patients admitted to the adult emergency  
497 department of a tertiary care hospital in Martinique: implications for triage,  
498 management, and reporting. *Ann Emerg Med.* 2012;59:42–50.

- 499 27. Andries A-C, Duong V, Ly S, Cappelle J, Kim KS, Lorn Try P, et al. Value of routine  
500 dengue diagnostic tests in urine and saliva specimens. *PLoS Negl Trop Dis*.  
501 2015;9(9):e0004100.
- 502 28. Hue KDT, Tuan TV, Thi HTN, Bich CTN, Anh HHL, Wills BA, et al. Validation of  
503 an internally controlled one-step real-time multiplex RT-PCR assay for the detection  
504 and quantitation of dengue virus RNA in plasma. *J Virol Methods*. 2011;177(2):168–  
505 73.
- 506 29. Ly S, Sorn S, Tarantola A, Canier L, Buchy P, Duong V, et al. Chikungunya  
507 outbreak–Cambodia, February–March 2012. *MMWR Morb Mortal Wkly Rep*.  
508 2012;61(37):737–40.
- 509 30. Vong S, Khieu V, Glass O, Ly S, Duong V, Huy R, et al. Dengue incidence in urban  
510 and rural Cambodia: results from population-based active fever surveillance, 2006–  
511 2008. *PLoS Negl Trop Dis*. 2010;4(11):e903.
- 512 31. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32.
- 513 32. R core team. R: A language and environment for statistical computing. 2017. [www.R-](http://www.R-project.org)  
514 [project.org](http://www.R-project.org).
- 515 33. Scornet E. Trees, forests, and impurity-based variable importance. Preprint:  
516 <https://arxiv.org/abs/2001.04295> .
- 517 34. Gamble J, Bethell D, Day NP, Loc PP, Phu NH, Gartside IB, Farrar JF, White NJ.  
518 Age-related changes in microvascular permeability: a significant factor in the  
519 susceptibility of children to shock? *Clin. Sci*. 2000;98(2): 211–216.
- 520 35. Guzman MG, Kouri G, Bravo J, Valdes L, Vazquez S, Halstead SB. Effect of age on  
521 outcome of secondary dengue 2 infections. *Int J Infect Dis*. 2002;6: 118–24.
- 522 36. Guzman MG, Gubler DJ, Izquierdo A, Martinez E, Halstead SB. Dengue infection.  
523 *Nat Rev Dis Primer*. 2016;2:16055.



- 524 37. Michels M, Sumardi U, Mast Q de, Jusuf H, Puspita M, Dewi IMW, et al. The  
525 Predictive Diagnostic Value of Serial Daily Bedside Ultrasonography for Severe  
526 Dengue in Indonesian Adults. *PLoS Negl Trop Dis.* 2013;7(6):e2277.
- 527 38. van Gorp EC, Suharti C, Mairuhu AT, Dolmans WM, van Der Ven J, Demacker PN,  
528 et al. Changes in the plasma lipid profile as a potential predictor of clinical outcome in  
529 dengue hemorrhagic fever. *Clin Infect Dis.* 2002;34(8):1150–3.
- 530 39. Villar-Centeno LA, Díaz-Quijano FA, Martínez-Vega RA. Biochemical alterations as  
531 markers of dengue hemorrhagic fever. *Am J Trop Med Hyg.* 2008;78(3):370–4.
- 532 40. Suvarna JC, Rane PP. Serum lipid profile: a predictor of clinical outcome in dengue  
533 infection. *Trop Med Int Health.* 2009;14(5):576–85.
- 534 41. Biswas HH, Gordon A, Nuñez A, Perez MA, Balmaseda A, Harris E. Lower low-  
535 density lipoprotein cholesterol levels are associated with severe dengue outcome.  
536 *PLoS Negl Trop Dis.* 2015;9(9):e0003904.
- 537 42. Durán A, Carrero R, Parra B, González A, Delgado L, Mosquera J, et al. Association  
538 of lipid profile alterations with severe forms of dengue in humans. *Arch Virol.*  
539 2015;160(7):1687–92.
- 540 43. Khor, C.C., Chau, T.N.B., Pang, J., Davila, S., Long, H.T., Ong, R.T., Dunstan, S.J.,  
541 Wills, B., Farrar, J., Van Tram, T. and Gan, T.T. Genome-wide association study  
542 identifies susceptibility loci for dengue shock syndrome at MICB and PLCE1. *Nature*  
543 *genetics.* 2011;43(11):1139-41.
- 544 44. Sierra B, Triska P, Soares P, Garcia G, Perez AB, Aguirre E, Oliveira M, Cavadas B,  
545 Regnault B, Alvarez M, Ruiz D. OSBPL10, RXRA and lipid metabolism confer  
546 African-ancestry protection against dengue haemorrhagic fever in admixed Cubans.  
547 *PLoS Pathogens.* 2017;13(2):e1006220.

- 548 45. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-  
549 density lipoprotein cholesterol in plasma, without use of the preparative  
550 ultracentrifuge. *Clin Chem.* 1972;18(6):499–502.
- 551 46. Nikolayeva I, Bost P, Casademont I, Duong V, Koeth F, Prot M, et al. A Blood RNA  
552 Signature Detecting Severe Disease in Young Dengue Patients at Hospital Arrival. *J*  
553 *Infect Dis.* 2018; 217(11):1690-8.
- 554