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

Multivariate evidence-based pediatric dengue severity prediction at hospital arrival

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

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

* Corresponding author:

E-mail: kevin.bleakley@inria.fr

† these authors contributed equally to this work.

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

Abstract

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

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

Findings. In a random forest analysis using 56 available variables we obtained AUC = 0·94, and for a sensitivity of 90%: specificity = 89%, positive predictive value (PPV) = 74%, and negative predictive value (NPV) = 96%. Platelet count, HDL cholesterol, and ultrasound pleural effusion and ascites were the four variables most associated with severe dengue outcomes. A random forest on only these four variables gave AUC = 0·88, and for a sensitivity of 90%: specificity = 82%, PPV = 64%, and NPV = 96%.

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

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

Introduction

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

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

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

different ways to define dengue severity: the WHO 1997 [9] and WHO 2009 [2] classification criteria, have not helped.[10,11]

The WHO 1997 dengue case classification defines dengue fever (DF), dengue hemorrhagic fever (DHF), and dengue shock syndrome (DSS). It was developed in 1975 based on studies on Thai children in the 1950s and 1960s, then revised in 1997.[9] It looks for changes in haematocrit of more than 20% (percentage of cellular compartment in whole blood), a platelet count below $100x10^9$/L, and signs of plasma leakage, e.g., ascites and pleural effusion, to define DHF and DSS. However, it remains difficult to apply prospectively in terms of DHF and DSS. Moreover, DHF criteria vary widely in frequency of occurrence. In reality, the existence of a continuous spectrum of dengue disease rather than two separate disease entities suggests that the WHO 1997 scheme acts as a longitudinal dengue-case classifier, using prerequisite clinical and/or biological signs to move from one clinical phase (e.g., DF) to the next (e.g., DHF).[12,13] Also, it turns out to be overly stringent: some patients showing symptoms of severe disease can be misclassified as non-severe cases if the rules are followed exactly.[14,15]

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

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

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

Two recent articles provide more focused attempts to predict future SD cases using early clinical data. First, Lee *et al*.[22] used 85% of a retrospective cohort of 69 SD and 1184 non-SD adult patients from Taiwan to build logistic regression models, which were then tested on the remaining 15% of the cohort. The most significant variables, on which their models were based, were: age>65, minor gastrointestinal bleeding, leukocytosis, and platelet count below $100x10^9$/L. On the training set, they obtained an AUC (area under ROC curve) of 0.85 and sensitivity and specificity of 70% and 91% respectively. On the test set, they had an AUC of 0.90 and sensitivity and specificity of 89% and 76% respectively. The large difference between the training and testing results indicates possible problems in generalising the model to new patients. Logistic models also can miss interesting interactions between variables

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

Second, Nguyen *et al.*[17] performed a prospective study on 7,563 Vietnamese children presenting at outpatient clinics, of which 2,060 had lab-confirmed dengue and 117 reached the WHO 2009 criterion for SD. Using multivariate logistic modeling on up to 16 pre-defined variables, they discriminated between the 117 SD and the other 7,446 patients, obtaining an AUC of 0·95, sensitivity and specificity of 87% and 88% respectively, and positive (PPV) and negative (NPV) predictive values of 10% and 99·8%. Their final model included history of vomiting, platelet count, aspartate aminotransferase (AST) level, and NS1 rapid test status.

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

Methods

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

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

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

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

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

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

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

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

Results

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

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

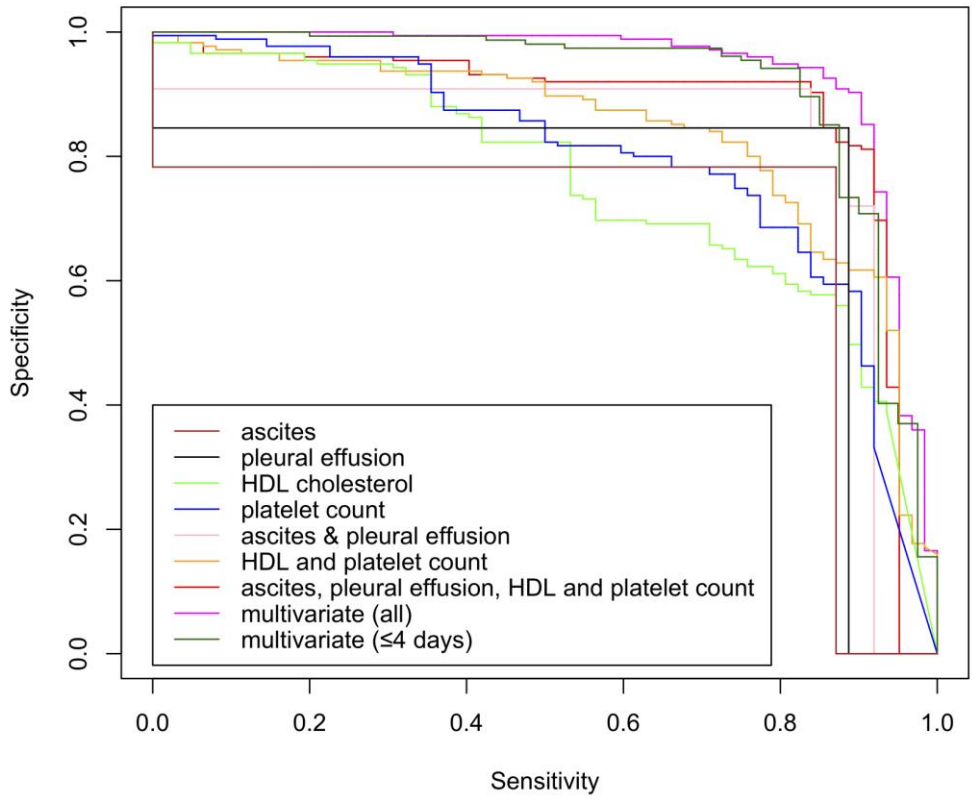
Table 1. Predicting severe dengue using random forest modeling

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)	**0.94** [0.93, 0.95]	*90%*	89% [85, 91]	74% [67, 79]	96% [96, 96]
multivariate (56 variables, 85% sensitivity)	**0.94** [0.93, 0.95]	85%	**94%** [89, 95]	**84%** [74, 87]	95% [95, 95]
multivariate (56 variables, (95% sensitivity)	**0.94** [0.93, 0.95]	95%	59% [48, 76]	45% [39, 58]	**97%** [97, 98]
multivariate (56 variables, ≤ 4 days)	0.92 [0.90, 0.93]	*90%*	73% [68, 82]	46% [42, 57]	**97%** [96, 97]

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

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

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



Analysis of SD individuals using adapted WHO 2009 classification predicted non-SD. Among the 62 SD patients in the Kampong Cham cohort, six were predicted non-SD by the random forest model when the sensitivity was fixed at 90%. All six of these dengue-confirmed patients were hospitalised between one and four days after onset of fever. All of them developed significant vascular leakage (as defined by clinical and biological variables) at defervescence and/or post-defervescence phases, characterised by detection of ascites and moderate pleural effusion by ultrasound and signs of severe plasma leakage.[2] Demographic, clinical and biological characteristics of these six SD patients with significant vascular leakage at admission, defervescence and/or post-defervescence phases are presented in Supplementary Table S3.

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

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

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

Table 2. Model comparisons

Study	Variables	Model	AUC	Sensitivity	Specificity	PPV	NPV

Nguyen *et al.*[17]	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]	**96%** [96, 96]
	4 B2021	logistic	0.90 [0.89, 0.91]	90%	83% [75, 84]	65% [57, 67]	**96%** [96, 96]
	Full B2021	RF	**0.94** [0.93, 0.95]	90%	**89%** [85, 91]	**74%** [67, 79]	**96%** [96, 96]

AUC = area under ROC curve; PPV = positive predictive value; NPV = negative predictive value; RF = random forest; 4 N2017 = the four variables from Nguyen *et al.*[17]: history of vomiting, platelet count, AST level, and NS1 rapid test status; 4 B2021 = the four variables from the present study: ultrasound ascites, ultrasound pleural effusion, HDL cholesterol level, and platelet count; full B2021 means the full 56-variable model from the present study. [a, b] corresponds to a 95% CI. Best values are shown in bold.

Discussion

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

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

In our study, in addition to platelet count (note that below $100x10^9$/L on day of admission gave 85% sensitivity for 66% specificity), which has been used in both WHO dengue case classifications, we identified three other highly univariately-significant prognostic markers: HDL cholesterol level, pleural effusion, and ascites, the latter two detected by ultrasonographic examination at admission. On their own, each performed fairly well for certain sensitivity/specificity trade-offs, rather than for a wide range of them. Using the four together to learn a random forest was better, though not as good as using all 56 variables. This

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

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

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

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

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

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

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

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

Declaration

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

Consent for publication. Not applicable.

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

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

Funding. Data generation was funded by BioMérieux (www.biomerieux.com) and Institut Pasteur du Cambodge. This study received funding from the French government's *Investissement d'Avenir* program, Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant: ANR-10-LABX_62°IBEID). TC is supported by a Wellcome-HHMI International Scholar award (Grant n°208710/Z/17/Z – www.wellcome.ac.uk). The funders had no role in the study design, data collection and data analysis, decision to publish, or preparation of the manuscript.

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

Acknowledgements. Not applicable

References

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

children in Asia: a phase 3, randomised, observer-masked, placebo-controlled trial. Lancet. 2014;384(9951):1358-65.

4. Hadinegoro SR, Arredondo-García JL, Capeding MR, Deseda C, Chotpitayasunondh T, Dietze Ret, et al. Efficacy and long-term safety of a dengue vaccine in regions of endemic disease. N Engl J Med. 2015;373(13):1195-206.
5. Sridhar S, Luedtke A, Langevin E, Zhu M, Bonaparte M, Machabert T, et al. Effect of Dengue Serostatus on Dengue Vaccine Safety and Efficacy. N Engl J Med. 2018;379:327-340.
6. Gubler DJ. Dengue, Urbanization and globalization: the unholy trinity of the 21st century. Trop Med Health. 2011;39(4 suppl):3–11.
7. Achee NL, Gould F, Perkins TA, Reiner RC Jr, Morrison AC, Ritchie SA, et al. A critical assessment of vector control for dengue prevention. PLoS Negl Trop Dis. 2015;9(5):e0003655.
8. Simmons CP, Farrar JJ, van Vinh Chau N, Wills B. Dengue. N Engl J Med. 2012;366(15):1423-32.
9. World Health Organization. Dengue hemorrhagic fever: diagnosis, treatment, prevention and control. Geneva: WHO. 1997.
10. Horstick O, Ranzinger SR. Reporting progress on the use of the WHO 2009 dengue case classification: a review. Southeast Asian J Trop Med Public Health. 2015;46 Suppl 1:49–54.
11. Gan VC, Lye DC, Thein TL, Dimatatac F, Tan AS, Leo Y-S. Implications of discordance in world health organization 1997 and 2009 dengue classifications in adult dengue. PloS One. 2013;8(4):e60946.
12. Libraty, D.H., Endy, T.P., Houng, H.S., Green, S., Kalayanarooj, S., Suntayakorn, S., 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

features and laboratory indicators: application of classification tree analysis. BMC Pediatr. 2018; 18:109.

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

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

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

45. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972;18(6):499–502.

46. Nikolayeva I, Bost P, Casademont I, Duong V, Koeth F, Prot M, et al. A Blood RNA Signature Detecting Severe Disease in Young Dengue Patients at Hospital Arrival. J Infect Dis. 2018; 217(11):1690-8.