Science Advances

NAAAS

Supplementary Materials for

Integration of population-level data sources into an individual-level clinical prediction model for dengue virus test positivity

Robert J. Williams *et al.*

Corresponding author: Daniel T. Leung, daniel.leung@utah.edu; Henrik Salje, hs743@cam.ac.uk

This PDF file includes:

Fig. S1
Tables S1 to S4

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	3
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	3-4
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	4-5
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	4-5
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	4
	5b	Describe eligibility criteria for participants.	4
	5c	Give details of treatments received, if relevant.	N/A
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	4
	6b	Report any actions to blind assessment of the outcome to be predicted.	N/A
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	4-6
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	N/A
Sample size	8	Explain how the study size was arrived at.	4
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	6
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	6
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	6
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	6
Risk groups	11	Provide details on how risk groups were created, if done.	N/A
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	7
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	7
Model development	14a	Specify the number of participants and outcome events in each analysis.	7
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	N/A
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	9-13
	15b	Explain how to the use the prediction model.	9-13
Model performance	16	Report performance measures (with CIs) for the prediction model.	9-13
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	15
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	14-15
Implications	20	Discuss the potential clinical use of the model and implications for future research.	14-15
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	19-22
Funding	22	Give the source of funding and the role of the funders for the present study.	15

Supplementary Table S1. TRIPOD checklist.

Clinical Predictors	DENV Negative	DENV Positive	OR	95% CI
Fever			2.15	1.87-2.47
No	971 (14%)	391 (6.8%)		
Yes	6,127 (86%)	5,337 (93%)		
Nausea			1.66	1.53-1.79
No	4,044 (57%)	2,543 (44%)		
Yes	3,054 (43%)	3,185 (56%)		
Headache			1.53	1.4-1.67
No	2,304 (32%)	1,371 (24%)		
Yes	4,794 (68%)	4,357 (76%)		
Emesis			1.50	1.39-1.62
No	4,126 (58%)	2,754 (48%)		
Yes	2,972 (42%)	2,974 (52%)		
Malaise			1.33	1.23-1.44
No	3,687 (52%)	2,569 (45%)		
Yes	3,411 (48%)	3,159 (55%)		
Anorexia			1.33	1.22-1.44
No	4,611 (65%)	3,340 (58%)		
Yes	2,487 (35%)	2,388 (42%)		
Abdominal Pain			1.27	1.17-1.38
No	4,815 (68%)	3,570 (62%)		
Yes	2,283 (32%)	2,158 (38%)		
Myalgias			1.26	1.16-1.36
No	3,638 (51%)	2,607 (46%)		
Yes	3,460 (49%)	3,121 (54%)		
Chills			1.20	1.11-1.3
No	4,139 (58%)	3,078 (54%)		
Yes	2,959 (42%)	2,650 (46%)		
Retro-orbital Pain			1.16	1.06-1.28
No	5,506 (78%)	4,284 (75%)		
Yes	1,592 (22%)	1,444 (25%)		
Hemorrhage			1.16	1.05-1.29
No	5,951 (84%)	4,677 (82%)		
Yes	1,147 (16%)	1,051 (18%)		
Diarrhea			1.06	0.97-1.16
No	5,401 (76%)	4,297 (75%)		
Yes	1,697 (24%)	1,431 (25%)		
Arthralgias			1.05	0.96-1.15
No	5,198 (73%)	4,134 (72%)		
Yes	1,900 (27%)	1,594 (28%)		

Rash			1.02	0.93-1.12
No	5,507 (78%)	4,426 (77%)		
Yes	1,591 (22%)	1,302 (23%)		
Age			0.98	0.98-0.98
	23 (18)	18 (11)		
Dark Urine			0.86	0.75-1
No	6,502 (92%)	5,306 (93%)		
Yes	596 (8.4%)	422 (7.4%)		
Seizure			0.83	0.69-0.99
No	6,704 (94%)	5,461 (95%)		
Yes	394 (5.6%)	267 (4.7%)		
Abnormal Movement			0.79	0.67-0.93
No	6,622 (93%)	5,423 (95%)		
Yes	476 (6.7%)	305 (5.3%)		
Nuchal Rigidity			0.77	0.63-0.94
No	6,764 (95%)	5,516 (96%)		
Yes	334 (4.7%)	212 (3.7%)		
Pharyngitis			0.76	0.7-0.83
No	4,978 (70%)	4,322 (75%)		
Yes	2,120 (30%)	1,406 (25%)		
Jaundice			0.63	0.51-0.78
No	6,761 (95%)	5,552 (97%)		
Yes	337 (4.7%)	176 (3.1%)		
Cough			0.55	0.51-0.6
No	4,039 (57%)	4,044 (71%)		
Yes	3,059 (43%)	1,684 (29%)		
Rhinitis			0.55	0.49-0.61
No	5,641 (79%)	5,017 (88%)		
Yes	1,457 (21%)	711 (12%)		

Supplementary Table S2. The relative frequencies, odds ratios, and confidence intervals for each clinical variable by DENV positivity.
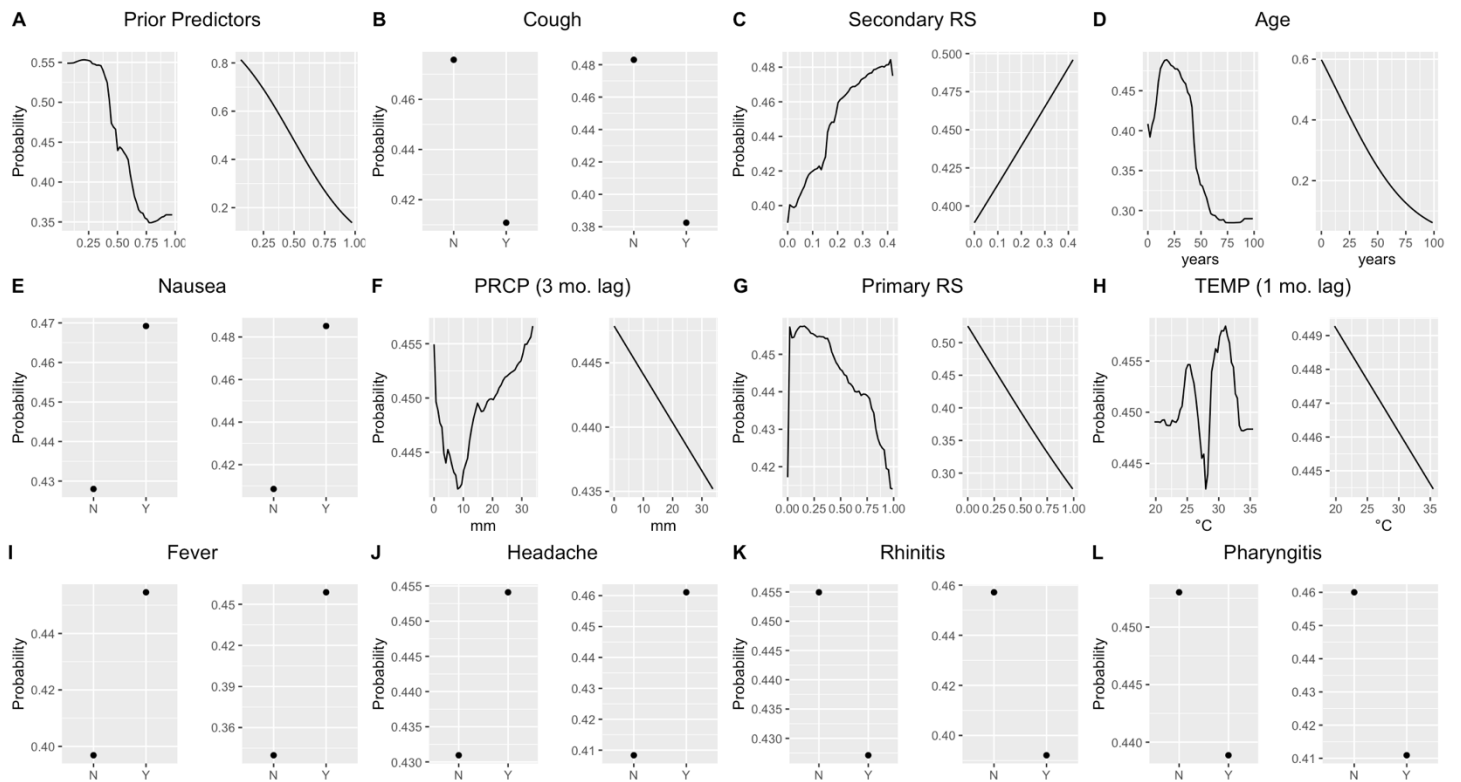
Supplementary Figure S1. Partial Dependency Plots for the top performing variables for predicting DENV infection by AUC. For each predictor, the graph on the left shows the partial dependency for a random forest model and the partial dependency for a logistic regression model is shown on the right. 'Y' indicates presence of the symptom and 'N' indicates absence of a symptom. 'PRCP' refers to precipitation, 'TEMP' refers to the environmental temperature, 'RS' refers to reconstructed susceptibility estimates.

	Overall, N = 12,826[1]	0-4 years, N = 954[1]	5-9 years, N = 2,033[1]	10-14 years, N = 2,971[1]	15-19 years, N = 2,271[1]	20-24 years, N = 1,174[1]	25-29 years, N = 875[1]	30-34 years, N = 624[1]	35-39 years, N = 448[1]	40+ years, N = 1,476[1]	p-value[2]
Nausea											<0.001
Y	6,239 (49)	341 (36)	952 (47)	1,515 (51)	1,239 (55)	646 (55)	447 (51)	320 (51)	208 (46)	571 (39)	
Cough											<0.001
Y	4,743 (37)	514 (54)	840 (41)	1,034 (35)	790 (35)	413 (35)	282 (32)	212 (34)	146 (33)	512 (35)	

[1]n (%)

[2]Pearson's Chi-squared test

Supplementary Table S3. The relative frequency of the top performing clinical variables stratified by age group. 'Y' indicates presence of the symptom and 'N' indicates absence of a symptom.

Climate Predictors (months lagged)	DENV Negative Mean (sd)	DENV Positive Mean (sd)	OR	95% CI
DEWPT	23.3°C (2.2)	23.7°C (1.8)	1.10	1.08-1.12
TEMP (1)	28.4°C (1.7)	28.6°C (1.5)	1.08	1.05-1.11
DEWPT (1)	23.3°C (2.3)	23.6°C (1.9)	1.08	1.06-1.1
VISIB	9.3 km (2.4)	9.5 km (2.1)	1.05	1.03-1.06
TEMP	28.3°C (1.6)	28.4°C (1.4)	1.04	1.01-1.06
PRCP	4.9 mm (4.5)	5.4 mm (4.6)	1.02	1.02-1.03
RH	75.3 (9.0)	76.6 (8.1)	1.02	1.01-1.02
RH (3)	70.7 (9.3)	69.8 (8.8)	0.99	0.98-0.99
PRCP (3)	3.7 mm (4.1)	3.2 mm (3.9)	0.97	0.96-0.98
SLP (1)	1008.1 mbar (2.9)	1007.8 mbar (2.6)	0.96	0.94-0.97
SLP	1008.1 mbar (2.9)	1007.7 mbar (2.7)	0.94	0.93-0.96
VISIB (3)	8.3 km (2.8)	7.7 km (2.9)	0.93	0.92-0.94
WDSP	0.6 m/s (0.3)	0.6 m/s (0.3)	0.74	0.65-0.85
WDSP (3)	0.7 m/s (0.3)	0.7 m/s (0.3)	0.73	0.64-0.83

Supplementary Table S4. The mean, standard deviation, odds ratio, and 95% CI intervals for each climate predictor.