Adv Health Behavior, 2018, 1(1), 12−16
DOI: 10.25082/AHB.2018.01.003

RESEARCH ARTICLE

Detecting Dengue Fever in Children: Using Sequencing Symptom Patterns for An Online Assessment Approach

Tsair-Wei Chien[1] Julie Chi Chow[2] Yu Chang[3] Willy Chou[4,5*]

Abstract: Background: Dengue fever (DF) is an important health problem in Asia. We examined it using its clinical symptoms to predict DF. Methods: We extracted statistically significant features from 17 DF-related clinical symptoms in 177 pediatric patients (69 diagnosed with DF) using the unweighted summation score and the non-parametric H^T person fit statistic, which jointly combine the weighted score (yielded by logistic regression) to predict DF risk. Results: Six symptoms (Family History, Fever $\geq$ 39°C, Skin Rash, Petechiae, Abdominal Pain, and Weakness) significantly predicted DF. When a cutoff point of 1.03 (p = 0.26) suggested combining the weighted score and the H^T coefficient, the sensitivity was 0.91 and the specificity was 0.76. The area under the ROC curve was 0.88, which was a better predictor: specificity was 5.56% higher than for the traditional logistic regression. Conclusions: Six simple symptoms analyzed using logistic regression were useful and valid for early detection of DF risk in children. A better predictive specificity increased after combining the non-parametric H^T coefficient to the weighted regression score. A self-assessment using patient smart phones is available to discriminate DF and may eliminate the need for a costly and time-consuming dengue laboratory test.

Keywords: dengue fever, H^T person mapping statistic, logistic regression, score summation, receiver operating characteristic curve

1 Introduction

Dengue fever (DF) is one of the most common arthropod-borne viral diseases worldwide,[1] especially in South East Asia, Africa, the Western Pacific, and the Americas.[2,3]

There is, however, no accurate and speedy diagnostic screening test for DF at an early stage because its signs and symptomse.g., fever, headache, and myalgiaare similar to those of other illnesses.[4–6] Some studies[4,5] that used a univariate analysis report that the presumptive diagnosis of DF is imprecise. Multivariate logistic regressions also do not significantly distinguish patients with dengue from those with other febrile illnesses.[7] The multivariate discrimination analyses reported a sensitivity and a specificity 0.76, and an area under the receiver operating characteristic (ROC) curve (AUC) of 0.93, but costly laboratory tests (Dengue Duo IgM & Rapid Strips; Panbio, Queensland, Australia)[8–11] were needed before DF was serologically confirmed.

DF symptoms are usually assessed using a dichotomous (i.e., absent versus present) evaluation. The dependent variable (DF^+ versus DF^-) predicted using independent evaluations with a weighted summation score is more accurate than that using simple evaluations with an unweighted summation score. So far, there has been no published study that has reported using the specific sequence of symptoms reported or observed in specific patients suspected of having DF. All published studies to date still report using only a standard group of symptoms with an unweighted summation score that apply to a general group of patients that might have DF.

The non-parametric HT fit statistic has been used in education and psychometrics to identify aberrant test respondents.[12,13] It is a transposed formulation of a scalability coefficient for items (e.g., symptoms in this study) and evidently the best among 36 person fit statistics for detecting abnormal behaviors.[14]

In the present study, we used the H^T coefficient com-

Received: June 13, 2018 **Accepted:** July 6, 2018 **Published:** July 9, 2018

*Correspondence to: Willy Chou, Ncphrology Department, Chi-Mei Medical Center, 901 Chung Hwa Road, Yung Kung Dist., Tainan 710, Taiwan; Email: hsienyi-wang@gmail.com

[1] Medical Research Department, Chi-Mei Medical Center, Tainan, Taiwan

[2] Department of Paediatrics, Chi-Mei medical center, Tainan, Taiwan

[3] National Taiwan University School of Medicine, Taiwan

[4] Department of Sports Management, College of Leisure and Recreation Management, Chia Nan University of Pharmacy and Science, Tainan, Taiwan

[5] Ncphrology Department, Chi-Mei Medical Center, Tainan, Taiwan

Citation: Chien TW, Chow JC, Chang Y, et al. Detecting Dengue Fever in Children: Using Sequencing Symptom Patterns for An Online Assessment Approach. *Adv Health Behavior*, 2018, **1**(1): 12−16

bined with weighted and unweighted variables to examine whether these combinations provide a valid and reliable approach for the early detection of DF in children.

2 Materials and methods

2.1 Sample and clinical symptoms

The sample of 177 pediatric patients ($\leq$ 16 years old; DF^+:69; DF^-:108) was the same as in our previous paper.[8] Guided by the literature,[5–7] we collected nineteen DF-related clinical symptoms from the patients medical records to develop the initial set of itemsdesignated as 0=absent or 1=present to screen for DF infection: (i) personal history of DF, (ii) family history of DF, (iii) mosquito bites within the previous 2 weeks, (iv) fever $\geq$ 39°C, (v) biphasic fever, (vi) rash, (vii) petechiae, (viii) retro-orbital pain, (ix) bone pain (arthralgia), (x) headache, (xi) myalgia, (xii) abdominal pain, (xiii) anorexia, (xiv) occult hematuria, (xv) stool occult blood, (xvi) cough, (xvii) sore throat, (xviii) soft (watery) stool, and (xix) flushed skin. Data from these patients charts were obtained and approved by the Research Ethics Review Board of the Chi-Mei Medical Center.

2.2 The H^T fit statistic

H^T is defined for the persons of a dichotomous dataset with L items (in columns) and N persons (in rows),[12] where X_{ni} is the scored (0,1) response of person n to item i, and $P_n = S_n/L$. Here, S_m is the raw score for person m, and S_n is the raw score for person n.

$$H^T(n) = \frac{\sum_{m=1,m\neq n}^{N}\left(\left[\sum_{i=1}^{L} X_{ni}X_{mi}\right]/L - P_nP_m\right)}{\sum_{m=1,m\neq n}^{N}(min[P_n(1-P_m).P_m(1-P_n)])} \quad (1)$$

H^T is the sum of the covariances between person n and the other persons divided by the maximum possible sum of those covariances, so that the range of H^T is -1 to +1. When the responses by person n are positively correlated with those of all the other persons, then $H^T(n)$ will be positive. In contrast, when the responses by person n are negatively correlated with those of all the other persons, then $H^T(n)$ will be negative. When person's responses are random, $H^T(n)$ will be close to zero[11]. We hypothesized that DF^+ patients have different H^T coefficients than do DF^- patients. All DF^+ group members were sequenced to the DF^- group members to obtain an H^T coefficient using equation (1).

2.3 Selecting symptoms and determining predictor variables

All symptoms were examined by the probability of Type I error using the following three steps in Figure 1 to determine predictor variables. First, each symptom was separately examined by the univariate approach using a χ^2 test and logistic regression, respectively, for identifying a significant association with DF. Second, two models (i.e., the univariate and the multivariate approaches) were investigated for determining valid predictor variables associated with DF when the probability of Type I error is less than 0.05. Third, the predictor variables were used in a weighted combination for discriminating patients suspected with dengue virus infection.

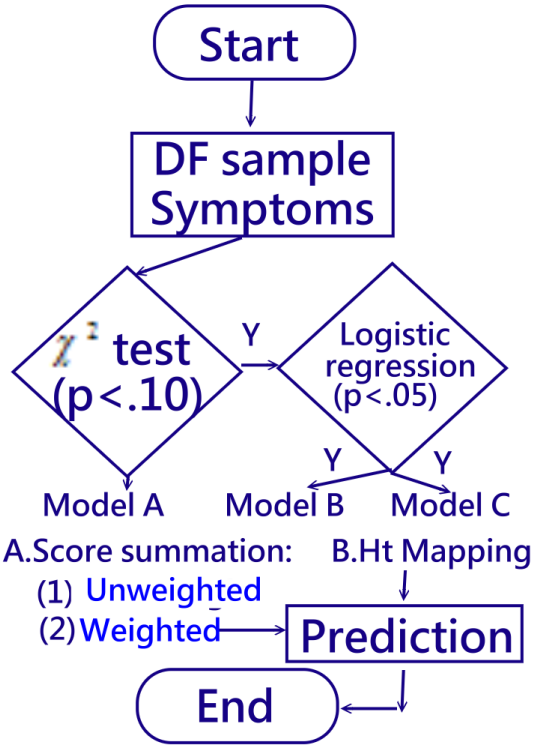

Figure 1. Overall study concept and the flow chart

2.4 Detecting dengue fever: a comparison of three models

The efficacy of three models (A, B, and C) for detecting dengue fever was examined: (i) A comparison was made using univariate logistic regression in Model A to examine effects through the AUC yielded by Unweighted (i.e., summed item) scores, Weighted (i.e., logistic regression) scores, and H^T coefficients, respectively, (ii) Multivariate logistic regression with the three aforementioned factors combined was used in Model B,

(iii) after selecting the significant variables in Model B, the combined predictive variables were analyzed using multivariate logistic regression in Model C to obtain effective weighted coefficients, and (iv) finally, we wanted to use a single continuous variable yielded by the combined predictive variables in Model C to compare the AUC with the counterparts in Model A and C.

2.5 Statistical tools and data analyses

SPSS 15.0 for Windows (SPSS Inc., Chicago, IL) and MedCalc 9.5.0.0 for Windows (MedCalc Software, Mariakerke, Belgium) were used to calculate (i) the probability of false positives (Type I error) using a χ^2 test and logistic regression, (ii) Youden J index (the higher, the better), AUC (area under the ROC curve), sensitivity, specificity, and the cutoff point at maximal summations of specificity and sensitivity, (iii) correlation coefficients among variables of unweighted, weighted, and H^T scores.

3 Results

Sixty-nine pediatric patients clinically diagnosed with DF and 108 with no evidence of DF infection were included in this study (Table 1). A χ^2 test and logistic regression analyses showed that only six symptoms (Family History, Fever $\geq$ 39°C, Skin Rash, Petechiae, Abdominal Pain, and Weakness) were significant for assessing the likelihood of DF (Table 2).

Table 1. Demographic characteristics of the study sample

Demographical Variables		DF(-)[1]		DF(+)[2]		Total		P-value[3]
		n	%	n	%	n	%	
Gender	Female	47	44	29	42	76	43	0.845
	Male	61	57	40	58	101	57	
Age(years)	0-4	48	44	11	16	59	34	0.005
	5-9	24	22	20	29	44	25	
	9-16	36	33	37	54	73	42	

[1]DF(+): patients with a positive dengue fever strip test
[2]DF(-): patients with a negative dengue fever strip test
[3]P-values were determined by the test

Table 2. Logistic analysis of symptoms for the patients suspected with dengue virus infection using the univariate approach

Symptom		DF(-)		DF(+)		Total		Chi-square test		Logistic	
Variable	Present	n	%	n	%	n	%	χ^2	P-value	B	P-value
Family history	No	79	73	40	58	119	67	3.74	0.053	1.35	0.002
	Yes	29	27	29	42	58	33				
High fever of 39°C	No	87	81	37	54	124	70	13.30	<.001	1.48	0.048
	Yes	21	19	32	46	53	30				
Skin rash	No	82	76	20	29	102	58	36.09	<.001	2.63	0.000
	Yes	26	24	49	71	75	42				
Petechiae	No	106	98	60	87	166	94	7.29	0.007	2.34	0.026
	Yes	2	1.9	9	13	11	6.2				
Abdominal pain	No	104	96	53	77	157	89	14.03	<.001	2.89	0.000
	Yes	4	3.7	16	23	20	11				
Weak sense	No	90	83	48	70	138	78	3.88	0.049	0.98	0.048
	Yes	18	17	21	30	39	22				
Constant										-3.3	

P-values were determined by the test and the Wald test of Logistic regression

Comparisons of the AUCs for the three study models (A, B, and C) showed that the weighted variable (derived by the Logistic regression) and the HT coefficient can be jointly used for predicting DF risk using equation (2):

$$Logit = -3.32 + 0.93 \times weighted\_score + 1.92 \times H^T\_coefficient \quad (2)$$

The risk probability can be computed using the transformed equation (3):

$$p = \frac{exp\,(logit)}{1 + exp\,(logit)} \quad (3)$$

where *logit* denotes a unit of log odds.

A cutoff point of 1.03 ($P = 0.26$) was determined using the combined predictive variables in Model C: sensitivity = 0.91, specificity = 0.76, and AUC = 0.88 (Figure 2 and Table 3). Predictive power was better: specificity was 5.56% (i.e., 75.93-70.37 shown in Table 3) higher than when using traditional logistic regression; however, the AUC was slightly lower (0.72) than when using the unweighted (0.84) and the weighted (0.87) variables (Table 2). The HT coefficients related to the weighted and unweighted scores were 0.26 and 0.22, respectively. The weighted score has a higher correlation coefficient than does the unweighted score to the H^T coefficients.

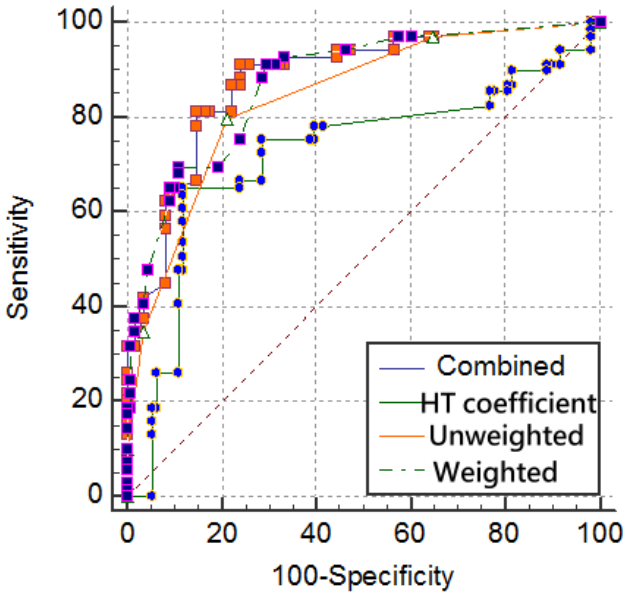

Figure 2. Four models plotted by ROC curves

A snapshot on a smart phone responding to questions (Figure 3, top) was generated and the results for assessing whether the patient has DF (Figure 3, bottom) were determined, which indicated that patients suspected of having DF can directly scan the QR-code to obtain

Table 3. Comparisons of AUC for the study models

Approach	Logistic		ROC curve analysis				
Steps	B[a]	P-value	AUC	Youden J[b]	Cut point	Sensitivity	Specificity
(1) Model A:Univariate approach with a single variable comparing to the DF using Logistic regression and ROC analysis							
Unweight[c]	1.60*	<0.001	0.84	0.58	>1.00	79.7	78.7
Weight[d]	0.97*	<0.001	0.87	0.61	>-0.93	91.3	69.4
H^t coeff.[e]	3.75*	<0.001	0.72	0.53	>0.15	65.2	88.0
(2) Model B: Multivariate approach with combined these three variables in regressing the DF using Logistic regression							
Unweight	0.31	0.595					
Weight	0.77*	0.014					
H^t coeff.	3.08*	0.001					
Constant	-1.03	0.350					
(3) Model C: Combined these two significant predictor variables using Logistic regression							
Weight	0.919*	<0.001					
H^t coeff.	2.962*	0.001					
Constant	-0.463	0.75					
(4) A single continuous variable yielded by the combined predictor variables							
Combined[f]	1	<0.001	0.89	0.67	>-0.65	87	79.6

[a]: coefficient of Logistic regression
[b]: Youden J index
[c]: item-score summation method
[d]: multiplying item-score with the weighted regression coefficient
[e]: the Ht coefficient
[f]: using the two combined variables to predict patients DF
*:p<0.05

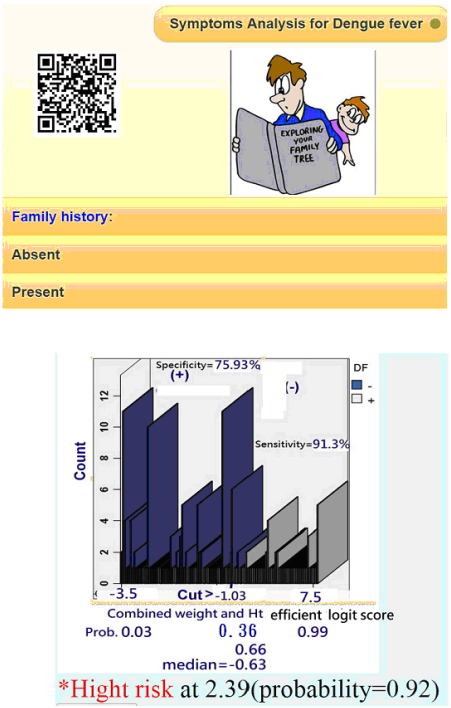

Figure 3. Figure 3 Snapshots on a smart phone responding questions (top) and the result (bottom) for assessing the patient DF

their DF *logit* scores (or the risk probability) and examine whether these 6 symptoms are useful for predicting a high DF risk (>1.03 *logits* or $P \geqslant 0.26 = exp(-1.03\,logits)/(1+exp(-1.03\,logits))$.

4 Discussion

We found that using the H^T coefficient yielded predictions that were 5.56% more specific (i.e., 75.93-70.37 shown in Table 3) than those of traditional logistic regression. The H^T index is promising when the patient sequence symptom pattern is compared with the DF$^+$ group to detect dengue fever in children. It can be combined with the weighted summation score to jointly predict the DF risk and then to report that risk on smartphones.

The H^T coefficient has been used in education and psychometrics to identify aberrant test respondents.[12, 13] Although some have used item response theory (IRT) fit statistics (e.g., outfit mean square error > 2.0) to select abnormal responses that indicate cheating, careless responding, lucky guessing, creative responding, or random responding,[15] our literature review revealed no published papers that reported using the H^T coefficient in medical settings, especially for detecting individual aberrant response patterns different from the study reference sample, or, like the current study, identifying the DF risk by comparing their sequence symptom pattern to that of the DF$^+$ group.

A diagnosis of DF is usually confirmed by three steps: (i) observing DF-related symptoms, (ii) testing laboratory data such as white blood cells (WBCs) and platelets (PLTs), and (iii) serologically verifying DF using dengue IgM and IgG antibodies, polymerase chain reaction (PCR) analysis, and virus isolation tests. The latter two are relatively expensive. It is needed to develop a self-assessment approach (e.g., scanning QR-code, responding questions, and obtaining the DF risk on his/her smart phone) (1) helping patients for consultation at an earlier stage, (2) prompting doctors sampling patient laboratory data when he/her DF risk reaches a cutpoint of P=0.26=exp(-1.03 *logits*)/(1+exp(-1.03 *logits*)).

We found that the weighted score was a better predictor than was the unweighted score (see Model A and Model B in Table 3). However, we still see so many scales in medical setting using unweighted summation scores to determine the presence or absence of disease. Along with the smartphones popularly used in the technical age, the way of obtaining the DF risk on smartphones using the combined H^T coefficient and weighted scores is available and worth recommending to healthcare providers to use for detecting the risk for DF.

This study has some limitations. First, the DF cutpoint based on the symptoms of our study sample might be biased toward that population. Moreover, we did not remove abnormal data when the H^T coefficient was less than the critical value of 0.22, which best identifies aberrantly responding examinees.[14] Second, although the sample size was small, using the Rasch H^T coefficient

combined with the AUC yielded highly accurate discriminatory screening. This finding, however, requires confirmation in prospective studies of other regions with a substantial incidence of DF.

5 Conclusions

Analyzing six simple symptoms using logistic regression is useful and valid for the early detection of DF risk in children. Combining the Rasch H^T coefficient with the weighted score yields a prediction that is 5.56% more specific than does traditional logistic regression. A self-assessment app using patient smartphones is available to help people suspected of having DF, and it might eliminate the need for costly and time-consuming laboratory tests.

6 Competing interests

The authors declare that they have no competing interests.

7 Authors contributions

T.-W.C. and S-C.K. conceived and designed the study, performed the statistical analyses and were in charge of recruiting study participants. W.-S.L. and T.-W.C. helped design the study, collected information and interpreted data. All authors read and approved the final article. This research was supported by grant Chi-Mei Foundation Hospital research CMFCR10593 from the Chi-Mei Medical Center. The authors have no other funding or conflicts of interest to disclose.

References

[1] World Health Organization. Dengue and Dengue Haemorrhagic Fever. Fact sheet No. 117. Geneva: World Health Organization, 2002.

[2] Henchal EAand Putnak JB. The dengue viruses. *Clin Microbiol Rev*, 1990, **3**(4): 376-396. https://doi.org/10.1128/CMR.3.4.376

[3] Gubler DJ. Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev*, 1998, **11**(3): 480-496. https://doi.org/10.1016/S1045-1870(97)80003-9

[4] Phuong HL, de Vries PJ, Nga TT, *et al*. Dengue as a cause of acute undifferentiated fever in Vietnam. *BMC Infect Dis*, 2006, **6**:123. https://doi.org/10.1186/1471-2334-6-123

[5] Nunes-Arajo FR, Ferreira MS and Nishioka SD. Dengue fever in Brazilian adults and children: assessment of clinical find-ings and their validity for diagnosis. *Ann Trop Med Para-sitol,* 2003, **97**(4): 415-419. https://doi.org/10.1179/000349803235002263

[6] Hammond SN, Balmaseda A, Prez L, *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**(6): 1063-1070. https://doi.org/10.4269/ajtmh.2005.73.1063

[7] Potts JA and Rothman AI. Clinical and laboratory features that distinguish dengue from other febrile illnesses in endemic populations. *Trop Med Int Health*, 2008, **13**(11): 1328-1340. https://doi.org/10.1111/j.1365-3156.2008.02151.x

[8] Lai WP, Chien TW, Lin HJ, *et al.* A screening tool for dengue fever in children. *Pediatr Infect Dis J*, 2013, **32**(4): 320-324. https://doi.org/10.1097/INF.0b013e31827e111e

[9] Lai WP, Chien TW, Lin HJ, *et al.* An approach for early and appropriate prediction of dengue fever using white blood cells and platelets. *HealthMed*, 2012, **6**(7): 806-812.

[10] Kittigul L, Suankeow K. Use of a rapid immunochromatographic test for early diagnosis of dengue virus infection. *Eur J Clin Microbiol Infect Dis*, 2002, **21**(3): 224-226. https://doi.org/10.1007/s10096-001-0691-z

[11] Vaughn DW, Nisalak A, Kalayanarooj S, *et al.* Evaluation of a rapid immunochromatographic test for diagnosis of dengue virus infection. *J Clin Microbiol*, 1998, **36**(1): 234-238.

[12] Sijtsma K. A coefficient of deviant response patterns. *Kwantitatieve Methoden*, 1986, **7**(22): 131-145.

[13] Linacre JM. A comment on the HT person fit statistic. *Rasch Meas Trans*, 2012, **26**(1):1358.

[14] Karabatsos G. Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Appl Meas Educ*, 2003, **16**(4): 277-298. https://doi.org/10.1207/S15324818AME1604_2

[15] Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas, 2002, **3**(1): 85-106.