UMC Utrecht and Erasmus MC

Statistical Analysis Plan

Feverkidstool and host-based Assay

26-3-2018

Table of contents

van Houten C, *et al. BMJ Paediatrics Open* 2019; 3:e000416. doi: 10.1136/bmjpo-2018-000416

General

This is a multi-centre, prospective study to validate a clinical prediction rule (Feverkidstool, FKT)[1] and to assess the accuracy of this tool when CRP is replaced with a new host-response based diagnostics (ImmunoXpert™) for distinguishing between bacterial and viral aetiologies in paediatric patients with lower respiratory tract infections and fever without source. The assay is based on serum concentrations of the three host-proteins: TNF-related apoptosis-inducing ligand (TRAIL), Interferon gamma-induced protein-10 (IP-10), and C-reactive protein (CRP).[2, 3] We use the term Assay throughout this document to refer to the aforementioned host-response based diagnostic test. This statistical analysis plan (SAP) was written in accordance with the International Conference on Harmonization topic E9 (Statistical principles for clinical trials, 1998), and the STARD statement.[4] The SAP was locked in advance of looking at the outcome data.

Study Objectives and Endpoints

Study Objectives

To assess the accuracy of a clinical prediction rule (“Feverkidstool”) [1] when CRP is replaced with a host-response based diagnostics ("Assay") to assess the risks of different serious bacterial infections (SBIs) in paediatric patients aged 1 to 60 months with lower respiratory tract infections (LRTI) or fever without source (FWS).

Primary Endpoints

For each patient:

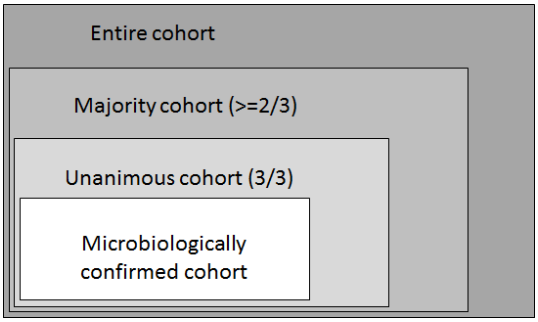
- Diagnosis assigned by three independent senior paediatricians (reference standard). Bacterial cases will be divided into pneumonia and other SBI based on clinical syndromes.
- Predictive probability for pneumonia or other SBI based on Feverkidstool, including CRP.
- Predictive probability for pneumonia or other SBI based on Feverkidstool when CRP is replaced with the Assay, hereafter called ‘the adjusted Feverkidstool’.

van Houten C, *et al. BMJ Paediatrics Open* 2019; 3:e000416. doi: 10.1136/bmjpo-2018-000416

Reference standard

We define three types of diagnostic reference standards that capture increasing portions of the patient cohort as illustrated in Figure 1.[5, 6]

Figure 1. Types of diagnostic reference standards



Microbiologically confirmed diagnosis

A sub-group of the patients have a microbiologically confirmed diagnosis. Bacterially labelled patients will be unanimously diagnosed by all three panel members as bacterial (or mixed) infection AND present one of the clinical indications delineated in Table 1A AND adhere to the respective criteria in Table 1B. Clinical indications are based on the diagnosis of the attending physician. Viral labelled patients will be unanimously diagnosed by panel members and have at least one laboratory detected virus. Non-infectious aetiology labelled patients will be unanimously labelled as such by the expert panel.

Table 1. List of clinical indications for infections and their predetermined criteria.

1A. Clinical indication	1B. Predetermined Criteria
Pneumonia	A positive sputum culture.
Bacteraemia	A positive blood culture excluding the following probable contaminants: 1. Coagulase-negative staphylococci, 2. Corynebacterium species 3. Bacillus sp. other than Bacilllus Anthracis 4. Propionibacterium acnes 5. Micrococcus species 6. Viridans group streptococci 7. Enterococci
Bacterial meningitis	A positive CSF culture and CSF pleocytosis with PMN predominance.
Lower/Upper UTI	Positive urine culture with >50,000 CFU of a single urinary pathogen AND leukocyte esterase AND/OR nitrite positive urinalysis.
Bacterial tonsillitis	Signs and symptoms suggestive of pharyngitis / tonsillitis and a positive throat culture of group A/G/C streptococci in the absence of other suggestive agent (Adenovirus or EBV).
Septic arthritis	Culture positive arthrocentesis.
Infective endocarditis	Fulfilment of Duke criteria.

Unanimous diagnosis

A sub-group of the patients have a unanimous diagnosis by the expert panel. A bacterial label will be assigned to patients that are unanimously diagnosed as bacterial by all three paediatricians. A viral label will be assigned to patients that are unanimously diagnosed as viral by all three paediatricians. Patients assigned as mixed infection were later classified as bacterial because they are clinically managed similarly. Bacterial cases will be divided into pneumonia and other SBI based on clinical syndromes assigned by the attending physician. The unanimous reference diagnosis will therefore be one of five options: reference Bacterial pneumonia; reference Bacterial other SBI; reference Viral; reference Non-infectious; reference Indeterminate, (in case where all three paediatricians diagnose the patient as Indeterminate, or each of the paediatricians assigns a different diagnosis and therefore there is no unanimous diagnosis).

Majority diagnosis

A sub-group of the patients have a majority diagnosis by the expert. **This sub-group will be used for the primary analysis.** A bacterial label will be assigned to patients where at least two of the three paediatricians diagnosed as bacterial. A viral label will be assigned to patients where at least two of the three paediatricians diagnose as viral. Patients assigned as mixed infection were later classified as bacterial because they are clinically managed similarly. Bacterial cases will be divided into pneumonia and other SBI based on clinical syndromes assigned by the attending physician. The majority reference diagnosis will therefore be one of five options: reference Bacterial pneumonia; reference Bacterial other SBI; reference Viral; reference Non-infectious; reference Indeterminate (in case majority of paediatricians assign an Indeterminate diagnosis, or each of the paediatricians assigns a different diagnosis and therefore there is no majority).

Indeterminate diagnosis

A sub-group of the patients in which a diagnosis reference could not be reliably established. This group includes patients with: (i) no majority diagnosis or (ii) an Indeterminate label assigned by the majority of experts.

To enabled analysis of as many included participants as possible, indeterminate cases will be imputed 10 times using the MICE algorithm, R statistical software. The imputation model included all Feverkidstool variables, the majority reference standard outcome and several relevant variables describing case mix of the patients, such as country, gestational age at birth and dehydration signs.

van Houten C, *et al. BMJ Paediatrics Open* 2019; 3:e000416. doi: 10.1136/bmjpo-2018-000416

Predictive probabilities

Feverkidstool

Based the variables as described by Nijman et al. [1]] the predictive probability of two states will be computed. The states are: pneumonia and other SBI.

Table 2 shows the variables and definitions that will be used for the Feverkidstool. Peripheral capillary refill is not part of the case record form from the new cohort, therefore these missing values will be replaced by mean prevalence in the initial Feverkidstool derivation cohort (=0.039).

To enable analysis of as many included participants as possible, other missing variables for clinical signs and symptoms will be imputed 10 times using the MICE algorithm, R statistical software. The imputation model included all Feverkidstool variables, the majority reference standard outcome and several relevant variables describing case mix of the patients, such as country, gestational age at birth and dehydration signs. Analyses were performed separately in the 10 imputed data sets and the results were combined using Rubin's rules.

We will perform sensitivity analyses by leaving out patients who were missing four or more variables used by the Feverkidstool.

Table 2. Variables of Feverkidstool

Variable	Definition
Age	Age (max 1 year, in years) + Age (if >1 year: age in years – 1)
Gender	Sex (female)
Temperature	Degrees Celsius
Duration of fever	Days, truncated at a maximum of 6 days
Tachypnea	Defined according to age specific APLS cut-offs
Tachycardia	Defined according to age specific APLS cut-offs
Desaturation	Oxygen saturation <94%
Prolonged capillary refill	Peripheral capillary refill time >3 seconds
Retractions	Presence of chest wall retractions, nasal flaring, groaning
General appearancee	Ill appearance
CRP	Ln(CRP) (mg/l), truncated at a CRP value of 225mg/l

Assay

The Assay uses the concentrations of TRAIL, IP-10 and CRP to calculate the probability of a bacterial infection. Figure 2 shows the distribution of the scores on a scale of 0-100, in which values closer to 0 have a higher probability of a viral infection and values closer to 100 are more likely bacterial.

Figure 2, Assay score

Analysis Methods

Primary outcome analysis

There will be four primary outcomes:

1. The diagnostic accuracy of the Feverkidstool to identify patients with pneumonia.
2. The diagnostic accuracy of the Feverkidstool to identify patients with other SBI.
3. The diagnostic accuracy of the adjusted Feverkidstool to identify patients with pneumonia.
4. The diagnostic accuracy of the adjusted Feverkidstool to identify patients with other SBI.

Definitions of diagnostic accuracy are described under the subheading Measures of accuracy.

The primary analysis will include non-ICU patients with a bacterial (or mixed) infection and viral infection, patients with a non-infectious aetiology and healthy controls will be excluded. The predictive probability's will be evaluated against reference standard with all patients predicted bacterial or viral infection, after eliminating patients with a non-infectious aetiology or healthy controls and after imputation of the patients with indeterminate reference standard.

Predictive probability's

As we want to find out whether the Assay can replace CRP, the diagnostic accuracy of both tests has to be compared. Therefore, we will use the replacement diagnostic pathway.[9]
Both in the original Feverkidstool and in the Assay, CRP is one of the variables. We will neutralize the effect of CRP in the adjusted Feverkidstool, therefore we will use the median CRP value in the Feverkidstool for this adjusted model. For the primary outcomes (see above) we will fit a logistic regression model including 2 variables; for outcomes 1 and 2 the Feverkidstool probability with CRP, for outcome 3 and 4 the adjusted Feverkidstool probability.

This will result in the following (polytomous) logistic regression models:

LP1 = Intercept + β_1 (LP pneumonia Feverkidstool)

LP2 = Intercept + β_2 (LP other SBI Feverkidstool)

LP3 = Intercept + β_{3a} (LP pneumonia Feverkidstool, with median CRP for all patients)+ β_{3b} (score Assay)

LP4 = Intercept + β_{4a} (LP other SBI Feverkidstool, with median CRP for all patients)+ β_{4b} (score Assay)

Whereas the intercept and the coefficients (β) will be fit, the intercept and coefficients within the Feverkidstool are fixed from the original prediction model.

The LPs for pneumonia and other SBI Feverkidstool are defined as follows:

LP (pneumonia Feverkidstool*) = -17.9 (Intercept) + 1.02 * Age (max 1 year, in years) + 0.01 * Age (if >1 year: age in years - 1)+ 0.13 * Sex (female) + 0.29 * Temperature (ºC) + 0.21 * Duration of fever (days) + 0.44 * Presence of tachypnoea - 0.04 * Presence of tachycardia + 1.59 * Oxygen saturation <94% - 0.18 * Capillary refill time (>3 s) + 0.47 * Presence of chest wall retractions + 0.16* ill appearance + 0.64 * Ln(CRP) (mg/l)*

LP (other SBI Feverkidstool*) = -4.7 (Intercept) -1.73 * Age (max 1 year, in years) + 0.11 * Age (if >1 year: age in years – 1)+ 0.70 * Sex (female) - 0.02 * Temperature (ºC) - 0.03 * Duration of fever (days) - 0.11 * Presence of tachypnoea - 0.02 * Presence of tachycardia - 3.29 * Oxygen saturation <94% + 0.30 * Capillary refill time - 3.78 * Presence of chest wall retractions + 0.27 * ill appearance + 1.14 * Ln(CRP) (mg/l)*

Probabilities of the outcomes are calculated with:

Feverkidstool Risk (pneumonia) = $e^{LP1}/ (1 + e^{LP1} + e^{LP2})$

Feverkidstool Risk (other SBI) = $e^{LP2} / (1 + e^{LP1} + e^{LP2})$,

Adjusted Feverkidstool Risk (pneumonia) = $e^{LP3} / (1 + e^{LP3} + e^{LP4})$

Adjusted Feverkidstool Risk (other SBI) = $e^{LP4} / (1 + e^{LP3} + e^{LP4})$, [7, 8]

where LP refers to the linear predictor in a (polytomous) logistic regression model.

A sensitivity analysis with a fractional polynomials recalibration will be performed.

Measures of accuracy

All measures will be performed for the risk of pneumonia and the risk for other SBI separately.

We will perform two primary analyses. For the first primary analysis we use cut-off independent accuracy measurements (e.g. c-statistics). The ordinal c-index (ORC) will be computed on the above mentioned patients (i.e. all non-ICU patients with bacterial and viral infections). [10]

As second primary analysis we will perform a decision curve analysis to help the reader to interpret the differences between the models along the wide range of predicted probabilities.[11]

Secondary analyses will include sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), total accuracy, positive likelihood ratio (LR+), negative likelihood ratio (LR-), NRI and diagnostic odds ratio (DOR). For these analysis we will calculate cut-off values for several risk thresholds, the most relevant thresholds will be based on the decision curve analysis. Calibration plots of predicted risks of the outcomes and observed frequencies will be presented by fifths of predicted risks. A histogram of predicted risks will be included at the bottom of the plot. To calculate the Net Reclassification Index (NRI)[12] to perform an intuitive comparison between the Feverkidstool with and without the Assay we will use several risk thresholds. These measures are defined as follows:

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$total\ accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

$$PPV = \frac{TP}{TP + FP} = \frac{sensitivity \cdot prevalence}{sensitivity \cdot prevalence + (1 - specificity) \cdot (1 - prevalence)}$$

$$NPV = \frac{TN}{TN + FN} = \frac{specificity \cdot (1 - prevalence)}{specificity \cdot (1 - prevalence) + (1 - sensitivity) \cdot (prevalence)}$$

$$LR+ = \frac{Sensitivity}{1 - Specificity}$$

$$LR- = \frac{1 - Sensitivity}{Specificity}$$

$$DOR = \frac{LR+}{LR-}$$

P, N, TP, FP, TN, FN are positives, negatives, true-positives, false-positives, true-negatives, and false-negatives, respectively. Prevalence is the relative frequency of the positive class (i.e., prevalence = P/(P + N)).

The 95% CIs of all accuracy measures will be reported throughout the analysis.

SPSS variables

Variables	Name in SPSS	Values in SPSS
General		
Remarks	Remarks1	0=none, 1=<2months, 3=control, 4=excluded
Admission site	Recruitmentsite	1=secondary, 2=tertiary, 3=PICU
Feverkidstool		
Age (max 1 year, in years)	Age_ST1y	
Age (if >1 year: age in years – 1)	Age_GT1y	
Sex (female)	Sex_num	0=male, 1=female
Temperature	Temperature	
Duration of fever	Fever_duration	
Tachypnea	Tachypnea	0=no, 1=yes
Tachycardia	Tachycardia	0=no, 1=yes
Desaturation	Hypoxia94	0=no, 1=yes
Prolonged capillary refill	Cap_refill2	
Retractions	Retractions	0=no, 1=yes
General appearancee	Ill_app	0=well appearing, 1=ill appearing
Ln(CRP)	LnCRP	
Assay		
Score (0-100)	Score	
Diagnosis	Dxbactvirmarg	1=bacterial, 2=viral, 3=marginal
Concentration TRAIL	TRAIL	
Concentration CRP	CRPdx	
Concentration IP10	IP10	
Outcome		
Logit FKT pneumonia	Logit_PNEU	
Logit Feverkidstool SBI	Logit_SBI	
Logit adjusted FKT pneumonia	Logit_PNEU2	
Logit adjusted FKT SBI	Logit_SBI2	
FKT Probability pneumonia	ProbPneu	
FKT Probability SBI	ProbSBI	
Adjusted FKT Probability pneumonia	ProbPneu2	
Adjusted FKT Probability SBI	ProbSBI2	
Reference standard		
Diagnosis	Expertpanel	0=viral, 1=bacterial, 2=non-infectious, 3=indetermined
Consensus	Consensus	1=3/3, 2=2/3, 0=1/3
Pneumonia or SBI	SBIcat	0=no SBI, 1=pneumonia, 2=other SBI

References

1. Nijman RG, Vergouwe Y, Thompson M, et al. Clinical prediction model to aid emergency doctors managing febrile children at risk of serious bacterial infections: diagnostic study. BMJ **2013**; 346: f1706.
2. Oved K, Cohen A, Boico O, et al. A Novel Host-Proteome Signature for Distinguishing between Acute Bacterial and Viral Infections. PLoS One **2015**; 10(3): e0120012.
3. van Houten CB, de Groot JA, Klein A, et al. A host-protein based assay to differentiate between bacterial and viral infections in preschool children (OPPORTUNITY): a double-blind, multicentre, validation study. Lancet Infect Dis **2017**; 17(4): 431-40.
4. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. The Standards for Reporting of Diagnostic Accuracy Group. Croat Med J **2003**; 44(5): 639-50.
5. Bertens LC, Broekhuizen BD, Naaktgeboren CA, et al. Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. PLoS Med **2013**; 10(10): e1001531.
6. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. Health Technol Assess **2007**; 11(50): iii, ix-51.
7. Begg CB, Gray R. Calculation of polychotomous logistic-regression parameters using individualized regressions. Biometrika **1984**; (71): 11-8.
8. Wijesinha A, Begg CB, Funkenstein HH, McNeil BJ. Methodology for the differential diagnosis of a complex data set. A case study using data from routine CT scan examinations. Medical decision making : an international journal of the Society for Medical Decision Making **1983**; 3(2): 133-54.
9. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. Bmj **2006**; 332(7549): 1089-92.
10. Van Calster B, Van Belle V, Vergouwe Y, Steyerberg EW. Discrimination ability of prediction models for ordinal outcomes: relationships between existing measures and a new measure. Biometrical journal Biometrische Zeitschrift **2012**; 54(5): 674-85.
11. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. Bmj **2016**; 352: i6.
12. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med **2011**; 30(1): 11-21.