BMJ Open

Cross-sectional evaluation of host biomarkers for guiding antibiotic use in bacterial and non-bacterial acute febrile illness in low- and middle-income tropical settings

B Leticia Fernandez-Carballo,[1] Michele Atzeni,[2] Camille Escadafal ,[1] Martina Vettoretti,[2] Steffen Geis,[3,4,5] Selidji Todagbe Agnandji,[6,7] Andre M Siqueira,[8] Jullita Kenala Malava,[3] Louis Banda,[3] Anita L Kabwende,[6] Ayodele Alabi,[6,9] Juste Christin Bie Ondo,[6] Marguerite Massinga-Loembe,[6] Paulin N Essone,[6] José Moreira,[10,11] Aline da Rocha Matos,[12] Braulia C Caetano,[12] Marilda M Siqueira,[12] Ana Maria Bispo de Filippis,[13] Érica Aparecida dos Santos Ribeiro da Silva,[14] Maria Cristina S Lourenço,[14] Julia Häring,[15] Anna Günther,[15] Meike Jakobi,[15] Nicole Schneiderhan-Marra,[15] Christine Hoogland,[16] Patrícia Brasil ,[8] Sunil Pokharel,[17] Stefano Ongarello,[16] Victoria Harris,[1] Aurélien Macé,[16] Sue J Lee ,[17,18] Barbara Di Camillo,[2] Sabine Dittrich [1,17,19]

To cite: Fernandez-Carballo BL, Atzeni M, Escadafal C, *et al*. Cross-sectional evaluation of host biomarkers for guiding antibiotic use in bacterial and non-bacterial acute febrile illness in low- and middle-income tropical settings. *BMJ Open* 2025;**15**:e086912. doi:10.1136/bmjopen-2024-086912

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2024-086912).

BLF-C and MA contributed equally.
CE and MV contributed equally.

Received 26 March 2024
Accepted 10 January 2025

For numbered affiliations see end of article.

Correspondence to
Dr Sabine Dittrich;
sabine.dittrich@th-deg.de

ABSTRACT

Objectives To evaluate the effectiveness of 18 different host biomarkers in differentiating bacterial from non-bacterial acute febrile illness (AFI) in resource-limited settings, specifically in Brazil, Malawi and Gabon.

Design Multinational, cross-sectional study.

Setting The study was carried out across multiple primary healthcare facilities, including urban and rural settings, with a total of three participating centres. Recruitment took place from October 2018 to July 2019 in Brazil, May to November 2019 in Gabon and April 2017 to April 2018 in Malawi.

Participants A total of 1915 participants, including children and adults aged 21–65 years with a fever of≤7 days, were recruited through convenience sampling from outpatient clinics in Brazil, Gabon and Malawi. Individuals with signs of severe illness were excluded. Written consent was obtained from all participants or their guardians.

Intervention This is not applicable as the study primarily focused on biomarker evaluation without specific therapeutic interventions.

Primary and secondary outcome measures The primary outcome measure was the ability of each host biomarker to differentiate between bacterial and non-bacterial AFI, as evaluated by area under the receiver operating characteristic (AUROC) curves. Secondary outcomes included the performance of individual biomarkers across the different study sites and in a multivariable setting.

Results A Kruskal-Wallis test, adjusted by Benjamini-Hochberg, was performed for each biomarker to identify covariates with a significant difference in the distribution of biomarker values. The analysis revealed that country of origin (Brazil, Gabon, Malawi), age, sex and malaria status significantly impacted biomarker distribution (p≤0.001). The most widely known biomarkers, such as white blood cell (WBC) count and C-reactive protein (CRP), demonstrated the best performance in distinguishing between bacterial and non-bacterial infections, with

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Diverse evaluation: This study is an extensive evaluation of 18 host biomarkers across low- and middle-income countries to differentiate bacterial from non-bacterial infections.
⇒ Methodological alignment: The study protocol aligns with Food and Drug Administration-approved classifications for distinguishing between bacterial and non-bacterial infections, enhancing methodological rigour.
⇒ No control group: The absence of a control group limits the ability to establish baseline biomarker performance or to assess asymptomatic carriers.
⇒ Time and geographical variability: The short enrolment period and heterogeneity of acute febrile illness causes may limit the generalisability of findings across different times and geographical contexts, particularly in Asia.
⇒ Subjectivity in classification: The two-step clinical classification process may introduce subjectivity, particularly as clinicians had access to haematology biomarker results during classification, potentially biasing results.

AUROCs reaching up to 0.83 (0.77–0.88) for WBC count and 0.71 (0.59–0.82) for CRP. However, none of the evaluated novel host biomarkers exhibited high performance (AUROC<0.70 in most cases) and variations in biomarker performance were observed across the three settings. Multivariable analyses demonstrated that while the best combination of biomarkers achieved higher AUROCs, the increase was modest (1–13%), suggesting that the interaction of biomarkers contributed minimally to predictive accuracy.

Conclusions There is a continued need for innovation in the host-biomarker space as the available markers do not meet the needs of diverse populations around the globe. This highlights the importance of targeted evaluations in non-severe patients in multiple settings to understand the true potential for real-life use. The findings highlight that not one-marker fits all settings and novel innovations remain urgently needed.

Trial registration number Clinical trial number: NCT03047642.

INTRODUCTION

Globally, acute febrile illness (AFI) is one of the leading reasons individuals, particularly children aged less than 5 years, present to primary healthcare facilities.[1] AFI has various causes, both infectious and non-infectious, that vary according to geography, age group and season.[1] In malaria-endemic settings, malaria was long considered the primary cause of all fevers; however, the introduction of rapid diagnostic tests (RDTs) for malaria in the past decade has disproved this. Modelling estimates suggest that approximately 70% of all fevers can be attributed to non-malarial causes, even in malaria-endemic settings.[2] In the Integrated Management of Childhood Illness, introduced by the WHO and UNICEF in the mid-1990s and subsequently implemented in more than 100 countries, the standard 'fever' algorithm currently includes a malaria RDT but no diagnostic test for other infections.[3] Hence, at the primary care level, the only evidence-based treatment decision that can be made relies on the malaria RDT, resulting in extremely high levels of antibiotic use in malaria-negative patients.[4] In this context of limited knowledge about the causes of AFI and limited diagnostic and human capacity, it is unsurprising that healthcare providers prescribe antibiotics to avoid negative outcomes in their patients.

To assist healthcare providers with clinical decision-making, a simple diagnostic tool is required to differentiate patients with AFI of bacterial and non-bacterial aetiology and provide appropriate care. In well-resourced settings, in both high-income countries (HICs) and low- and middle-income countries (LMICs), some non-specific host biomarkers are used for this purpose, most frequently C-reactive protein (CRP) and procalcitonin (PCT), although these biomarkers are less useful in settings with a higher frequency of comorbidities.[5] Thus, in 2015, an international group of experts was convened to define the target product profile (TPP) of such a tool, specifically for low-resource settings, to guide product development and implementation as part of integrated treatment management guidelines.[6] Since then, the ongoing viral pandemic (SARS-CoV-2) has further highlighted the challenge of differential diagnosis and shows yet again that better antimicrobial stewardship interventions are needed to counter the overprescribing of antibiotics in patients with viral infections.[7]

Host biomarkers other than CRP and PCT have been evaluated for distinguishing bacterial from non-bacterial infections, including human neutrophil lipocalin (HNL), heparin-binding protein and chitinase 3-like protein 1.[8] There are also some commercially available tests. ImmunoXpert, from MeMed, uses a biomarker combination comprising CRP, interferon gamma-inducible protein 10 (IP-10) and TNF-related apoptosis-inducing ligand, while FebriDx, from Lumos Diagnostics, uses an MxA and CRP biomarker combination. While these biomarker signatures show promise, they have only been evaluated in limited settings. Any potential impact of co-infections or comorbidities, common in LMICs, on their effectiveness is unknown. Other characteristics of host-biomarker studies that hamper direct comparisons include: (1) just one/a few biomarkers in the study; (2) small sample sizes, increasing the probability of recruiting unrepresentative study populations; (3) narrow population subgroups (eg, children only, hospitalised only, respiratory infections only), limiting the generalisability of study results to the broader AFI population; (4) studies conducted in one country, so co-infections/comorbidities may not be comparable with those of other countries; (5) retrospective studies that used convenience sampling and case-control study designs, increasing the risk of bias; and (6) the lack of standard definitions for classifying bacterial versus non-bacterial infections.[9]

Here, we describe the Biomarker for Fever Diagnostic (BFF-Dx) study, specifically designed to evaluate host biomarkers to distinguish bacterial from non-bacterial infections in line with the published TPP and the final use case of such diagnostic tests. To our knowledge, this is the only study to evaluate host biomarkers in the intended target population (non-severe patients), prospectively, in multiple settings with a large sample set. We evaluated 18 host biomarkers in three distinct settings, in Brazil, Gabon and Malawi with the main objective of providing a performance comparison of host biomarkers in the non-severe AFI population from resource-limited settings, with the goal of overcoming many of the previously described limitations (eg, sample size, retrospective vs prospective, focused populations, biased analysis).[10] The described comparison was conducted within the pragmatic context of diagnostic product development and aimed to identify host biomarkers or biomarker combinations for utilisation in next-generation RDTs.

METHODOLOGY

Study settings

This multinational, cross-sectional study was conducted in Brazil, Gabon and Malawi; Gabon and Malawi were selected as high-malaria endemicity settings, while Brazil was selected as a low-malaria endemic setting. The study

sites were UPA Manguinhos and Family Health Clinics Armando Palhares in Rio de Janeiro, Brazil; the Clinical Trials Unit Centre of Medical Research Lambaréné (CERMEL), Lambaréné, Gabon; and Malawi Epidemiology and Intervention Research Unit (MEIRU), Chilumba campus, Malawi. The enrolment sites were an urban primary healthcare facility, a hospital in a semi-rural setting and a rural primary healthcare facility in Brazil, Gabon and Malawi, respectively. Participants were recruited from October 2018 to July 2019, May to November 2019 and April 2017 to April 2018, in Brazil, Gabon and Malawi, respectively. Reporting complies with the STARD-15 (Standards for Reporting of Diagnostic Accuracy Studies 2015) checklist.

Study population and study procedure

Participants were obtained through convenience sampling and included both children and adults, aged between 2 and 65 years, who presented at the outpatient clinics with a history of fever of ≤7 days duration (Brazil and Gabon) or fever at presentation (Malawi). Patients with signs of severe illness were not included in the study. The overarching study protocol was slightly adapted to each site due to local requirements (logistical or ethical). Detailed criteria for inclusion by study sites have been published previously.[10] Outcomes were based on the TPP criteria and while no patient input was used, external expert input was used to define the target population and criteria. Only patients who met the eligibility criteria and who provided written consent (patient or guardian for children) were enrolled in the study. Data and samples were systematically collected and analysed as previously described. To ensure consistent quality and comparability of data, the same standard operating procedures were used at all sites (for data collection and laboratory testing).[10]

Patient and public involvement statement

None

Bacterial/non-bacterial classification and biomarker selection and testing

A two-step process was used to classify the patients into 'bacterial' and 'non-bacterial' groups. First, the cause of fever (bacterial/non-bacterial) was classified according to laboratory-determined parameters ('electronic group'). The electronic group was based on predefined and widely accepted laboratory parameters, including direct pathogen detection, a fourfold increase in antibody titre or a positive PCR or antigen RDT result. The list of tests performed is described in detail in by Escadafal *et al*.[10] Next, cases that could not be classified by laboratory-determined parameters were assessed by a panel of three independent clinical experts. The patient's history and clinical and laboratory data were provided to the experts. Clinical expert's assessments were then compared. If the three-panel members unanimously assigned a diagnostic label, patients were considered to have 'bacterial' or 'non-bacterial' infections; if two out of three-panel members reported a classification of 'bacterial' or 'non-bacterial', these patients were considered to have 'probable bacterial infection' or 'probable non-bacterial infection', respectively.

Data were analysed based on three groups of patients: (1) the 'electronic group', that is, subjects with a cause of fever defined based on laboratory parameters; (2) the 'strict group' which comprised the electronic group and the patients that were unanimously classified by the clinical panel of three experts; and (3) the 'loose group' which comprised the electronic and strict groups as well as those patients for whom two of the clinical experts agreed they had either probable bacterial or probable non-bacterial infection. Subjects with undetermined causes of fever according to the three classification criteria considered ('electronic group', 'strict group', 'loose group') were excluded from the statistical analysis. This outcome-oriented approach, based on methods developed for host-biomarker studies previously, was used to ensure the total intended-use population of any future test was represented in the final analysis.[10 11]

The evaluated biomarkers were selected based on previously reported performances and haematological markers as well as CRP were included as comparators (table 1 and online supplemental table 1 and 2).[8 12]

At the end of data collection, all biomarker data were analysed to assess the percentage of missing values and the percentage of values below the lower limit or above the upper limit of detection of the used tests. Biomarkers with more than 50% of missing data or more than 95% of saturated values below the lower limit of quantification of the used test, were excluded from the following statistical analysis.

Statistical analysis

Kruskal-Wallis analysis and definition of covariates influence on biomarkers

A Kruskal-Wallis test, adjusted by Benjamini-Hochberg, was conducted for each biomarker to determine which covariates exhibited statistically significant differences in the distribution of biomarker values. The covariates studied were country (ie, the country of origin of the patients), age, sex, malaria status, comorbidities (ie, presence of one or more diseases among cardiovascular, neurological, respiratory, renal, genitourinary, connective tissue, cancer or infectious diseases), malnutrition status calculated based on WHO body mass index criteria, self-reported use of antibiotics prior to visiting the health facility, axillary temperature≥38°C and positive result to Chikungunya test. The Kruskal-Wallis test was performed for each of the three patient groups defined in the previous section ('electronic', 'strict', 'loose'). The results of the Kruskal-Wallis test allowed the identification of covariates that most significantly impacted the biomarker distribution ($p \leq 0.001$, adjusted by Benjamini-Hochberg). The most significant covariates were considered for defining subgroups of patients in which the following univariate

Table 1 Novel biomarkers were identified in the literature and evaluated in the BFF-Dx study, including sample type used, evaluation method and sample origin

Abbreviation	Biomarker name	Sample type	Evaluation method	Sample origin
AGP	A-1-acid glycoprotein	EDTA-plasma	Luminex	B, G, M
C2	Complement 2	EDTA-plasma	Luminex	B, G, M
C4b	Complement C4b	EDTA-plasma	Luminex	B, G, M
CHI3L1	Chitinase-3-like protein 1	EDTA-plasma	Luminex	B, G, M
CRP	C-reactive protein	EDTA-plasma	CRP NycoCard/ NycoCardReader II, ELISA	B, G, M
Gal-9	Galectin-9	EDTA-plasma	Luminex	B, G, M
HBP	Heparin-binding protein	EDTA-plasma	ELISA	B, M
HNL	Human neutrophil lipocalin	Heparin-activated plasma time-controlled activation*	ELISA	M
		EDTA-plasma	ELISA	B, G, M
HP	Haptoglobin	EDTA-plasma	Luminex	B, G, M
IFN-gamma	Interferon gamma	EDTA-plasma	Luminex	B, G, M
IL-4	Interleukin-4	EDTA-plasma	Luminex	B, G, M
IL-6	Interleukin-6	EDTA-plasma	Luminex	B, G, M
IP-10	Gamma-induced protein 10	EDTA-plasma	Luminex	B, G, M
LBP	Lipopolysaccharide binding protein	EDTA-plasma	Luminex	B, G, M
NGAL	Neutrophil gelatinase-associated lipocalin	Frozen heparin-activated plasma	Luminex	M
		EDTA-plasma	Luminex	B, G, M
PCT	Procalcitonin	EDTA-plasma	Luminex; ELISA	B, G, M
sPLA2	Secretory phospholipase 2	EDTA-plasma	Luminex	B, G, M
sTREM-1	Soluble triggering receptor expressed on myeloid cells 1	EDTA-plasma	Luminex	B, G, M
TRAIL	TNF-related apoptosis-inducing ligand	EDTA-plasma	Luminex	B, G, M

*Whole blood samples were collected in lithium heparin tubes and activation was performed within 60 min prior to freezing and subsequent ELISA testing.[29] All biomarkers were tested using the same standard operating procedures (SOPs) and all sites were trained on the SOPSs. For CRP and PCT, different devices were used at different sites, repeat testing was performed at the central facility (NMI).
B, Brazil; BFF-Dx, Biomarker for Fever Diagnostic; G, Gabon; M, Malawi.

analyses were performed or included as covariates in the multivariable analyses.

Univariate analysis

As an exploratory step, the ability of each biomarker to discriminate between bacterial and non-bacterial infections was assessed by the area under the receiver operating characteristic curve (AUROC). In particular, subjects were ranked based on the values of the single variable of interest (ie, based on ordered values) and, using this as a score, calculated the ROC curve and the corresponding area under the curve. Such univariate analysis was conducted for each patient group ('electronic', 'strict', 'loose') and specific patient subgroups (Malaria status, Country and Age).

However, since the univariate analyses did not yield satisfactory results, we also explored multivariable models to potentially improve the predictive capabilities by incorporating a broader range of information.

Multivariable analysis

Multivariable classification models were developed to assess the discrimination ability of combinations of biomarkers and covariates. For the multivariable analysis, both linear (logistic regression) and non-linear classification models (RuleFit) were explored.[13] The candidate features for each model included a group of host biomarkers and some additional covariates (age, temperature, fever duration, diastolic blood pressure, respiration rate and pulse rate). Regarding host-biomarkers, three different groups of biomarkers were considered: haematology biomarkers only (ie, white blood cell (WBC), neutrophil, red blood cell, lymphocyte counts), protein biomarkers only (ie, novel biomarkers+CRP) and haematology plus protein biomarkers (ie, all biomarkers).

For each patient subgroup and each candidate feature set, three multivariable models were developed: (1) a logistic regression model with stepwise (SW) feature

selection; (2) a logistic regression model with features selected based on recursive feature addition (RFA; a variant of the method proposed in[14]); (3) RuleFit, a non-linear model in which a set of rules from an ensemble of decision trees (typically from a tree-based model like a random forest or gradient boosted trees) is generated and then fit a sparse linear regression model (regularised with least absolute shrinkage and selection operator (LASSO)), where the features are the rules generated from the trees.[13 14]

To further tackle the number of biomarkers and variables included in the best models, we introduced an additional selection step, employing a plateau-seeking approach. The primary objective of this approach was to pinpoint a concise set of variables capable of attaining an AUROC score similar to that of our comprehensive model which already incorporated the most impactful and previously selected variables. This was to ensure that our model was not only effective in terms of performance but also efficient in its variable inclusion.

Each model was trained and tested using the following pipeline. The data were randomly split into training and test sets (80% and 20% of the data, respectively) stratifying by the outcome variable. Missing data in the training and test sets were imputed using the MICE (multiple imputation by chained equation) algorithm. The n_imp parameter for MICE imputation was set to 1, resulting in a single imputed data set; however, the imputation process was integrated into a robust bootstrapping pipeline, generating ten independent data sets. This approach ensured variability in our results, stemming not only from the MICE imputation but also from the bootstrapping process. This dual approach guarantees that each imputed data set is distinct.[15] All quantitative variables were scaled into the range (0,1) by subtracting their minimum value and dividing by the difference between the maximum and minimum values in the training set. The categorical variables with n categories were encoded using n-1 binary 'dummy' variables. The model was then trained on the imputed and scaled training set and its performance was assessed on the imputed and scaled test set by computing the AUROC. The AUROC on the test set was also calculated for single-host biomarkers, to allow a fair comparison of the performance of the multivariable classification models versus single-host biomarkers.

To assess the robustness and variability in the results of the developed models, the entire pipeline were bootstrapped, that is, it was run 10 times with different random training-test set splits. Finally, the mean and the SD or the minimum and maximum reached of the AUROC across the 10 training-test splits were calculated for each multivariable model and each single host biomarker.

Software

All statistical analyses and model development were performed using the R programming language (V.4.1.2). Specifically, the mice package was used for data imputation, while the pre and stats packages were used for RuleFit and logistic regression model development, respectively.

RESULTS

Study population

In total, 1915 patients with AFI were included in the study (Brazil: n=500; Gabon: n=415; Malawi: n=1000). Just under half (862/1915, 45%) of participants at each study site were male. Children aged <5 years comprised 45/500 (9%), 182/415 (43.9%) and 367/1000 (36.7%) participants in Brazil, Gabon and Malawi, respectively; the median (range) age was 3 (2–4) years (table 2). Detailed baseline characteristics of patients and analyses of differences will be described in a separate manuscript (Alabi *et al* in preparation).

Bacterial and non-bacterial outcomes by classification groups

Using the electronic classification grouping, 15.1% (290/1915) of cases were bacterial infections, 20.2% (387/1915) were non-bacterial infections and 64.5% (1238/1915) had an undetermined cause of fever (figure 1). Under the strict classification grouping, 24.3% (366/1509), 66.9% (1010/1509) and 9.0% (133/1509) were classified as bacterial, non-bacterial and undetermined infections, respectively, while using the loose classification grouping 25.7% (491/1915), 67.3% (1286/1915) and 7.0% (133/1915) were classified as bacterial, non-bacterial and undetermined infections, respectively (figure 1). Subjects with undetermined causes of fever/infections were excluded from the following univariate and multivariable analyses.

Exclusion of biomarkers with too many missing or saturated values

The biomarkers C4b, HNL and PCT had more than 50% missing values and were therefore excluded. The high number of missing values is due to the fact that biomarkers were analysed in groups based on the required dilution using the Luminex platform. For some biomarkers, the dilution was not optimal and it was only possible to remeasure biomarkers with a different dilution a limited number of times. IFN-gamma and sTREM-1 were excluded due to more than 95% of values saturated to the minimum/maximum level detectable by the measurement instrument. All the biomarkers retained in the analysis had less than 12% missing values (online supplemental table 3).

Identification of relevant subgroups for analyses

According to the Kruskal-Wallis analysis on the 'electronic group', the variables 'country', 'malaria status' and 'age' showed statistically significant differences in the distributions of many host biomarkers ($p \leq 0.001$ for strong differences, $0.001 \leq p < 0.01$ for high differences; Online supplemental table 4). The variables 'sex', 'comorbidities', 'history of antibiotic use' showed no ($p > 0.05$) or

Table 2 Baseline characteristics of patients

	Brazil	Gabon	Malawi	All
0–5 years (median, IQR, n)	3, (2–4), 45	3, (2–5),182	3, (2–4), 367	3, (2–4), 594
5–15 years (median, IQR, n)	11, (8–14), 85	9, (7–12), 214	9, (7–12), 276	9, (7–12), 575
>15 years (median, IQR, n)	34, (24-45), 370	16, (16–16.5), 19	28, (21-36), 357	30, (21-42), 746
Male (%, n)	49.6, 248	45.1, 187	42.7, 427	45.0, 862
Temperature, °C (median, IQR, n)	37.7, (36.7–38.4), 500	36.8, (36.4–37.4), 415	38.1, (37.7–38.8), 999	37.8, (37.3–38.5), 1914
WBC count, 10^9/L (median, IQR, n)	7.28, (5.47–10.39), 494	7.7, (5.7–10), 411	6.7, (5.1–9.3), 985	7.1, (5.3–9.8), 1890
Neutrophil count, 10^9/L (median, IQR, n)	4.97, (3.63–7.4), 494	2.77, (1.96–3.9), 408	4.3, (3–6.18), 906	4.1, (2.8–6), 1812
RBC count, 10^9/L (median, IQR, n)	40.1, (36.5–43.2), 494	33.2, (29.4–35.8), 412	36.2, (33.2–39.5), 984	36.3, (33–40.2), 1892
Lymphocyte count, 10^9/L (median, IQR, n)	1.15, (0.7–1.99), 493	2.73, (1.8–4.16), 411	1.5, (1–2.2), 982	1.63, (1–2.6), 1883
CRP NycoCard*—mg/L (median, IQR, n)	70.5, (35–98.75), 498	28, (5–73), 415	47, (12–106.5), 987	49, (13–98), 1900
Malaria-positive by RDT on-site (% all, n)	0.2, 1	56.4, 234	45.9, 458	36.2,693
Malaria-positive by qPCR or microscopy (% all, n)	-	-	50.5, 505	-
HIV-positive by RDT (% all, n)	1.4, 7	1.2, 5	4.2, 42	2.8,54
History of antibiotic-use pre-presentation (% all, n)	8.8, 44	2.41, 10	7.2, 70	6.5,124
History of antipyretic-use pre-presentation (% all, n)	83.2, 416	79.76, 331	55.1, 551	62.2,1298
Cough (%, n)	35.8, 179	30.1, 125	48.2, 482	41, 786
Diarrhoea or vomiting (%, n)	31.8, 159	28.9, 120	27.5, 275	28.9, 554
Dysuria or urinary urgency (%, n)	0.9, 45	5.12, 21	7.6, 76	7.4, 142
Headache (%, n)	76.4, 382	46.5, 193	71.1, 711	67.2, 1286
Sore throat or swallow pain (%, n)	39, 195	8.92, 37	15.8, 158	20, 390
Rash (%, n)	24.4, 122	4.1, 17	2.5, 25	8.6, 164

*NycoCard was found to be equivalent to reference testing in the relevant range (online supplemental figure 1).
-, data not available; CRP, C-reactive protein; qPCR, quantitative PCR; RBC, red blood cell; RDT, rapid diagnostic test; WBC, white blood cell.

slight (p≤0.05) differences in all the host biomarkers. The effects of 'chikungunya status' and 'fever above 38°C' were generally significant (p≤0.01), but the sample sizes for these groups were either too small or exhibited an imbalance. Additionally, while we conducted subgroup analyses by clinical syndromes (ie, cough, diarrhoea or vomiting, dysuria or urinary urgency, headache, sore throat or swallow pain, rash), the resulting data sets were similarly limited in size, restricting our ability to make robust interpretations from these analyses. The primary focus remained centred on populations grouped by study country and malaria status variables—both of which showed big statistical differences with the value of the biomarkers in the 'strict' and 'loose' groups (online supplemental tables 5 and 6)—other significant covariates were also included in the multivariable analysis. This inclusion was due to their influence and factors like the study country were considered as variables in the overall scenario.

Individual host-biomarker performance: univariate analysis

The performance of 18 host biomarkers was consistent across the three patient classification groups in each of the settings (table 3 and online supplemental tables 7–9). WBC and neutrophil counts were the most effective biomarkers for differentiating bacterial and non-bacterial infections. For the malaria-negative population, the mean (95% CI) of AUROC for WBCs was between 0.60 (0.48 to 0.72) and 0.83 (0.77 to 0.88) and for neutrophils, it was between 0.67 (0.57 to 0.77) and 0.80 (0.74 to .86) across the three countries and the three groups ('electronic', 'strict', 'loose'). Neutrophil and WBC counts showed the highest AUROCs in the Brazilian population, between 0.80 (0.74 to 0.86) and 0.83 (0.77 to 0.88), respectively. All protein biomarkers showed relatively poor performances (<0.7 in most cases, table 4) in all three settings. Galactin-9, CRP, IP-10 and NGAL were the best-performing protein biomarkers across the three settings and criteria. Protein biomarkers showed better performances in

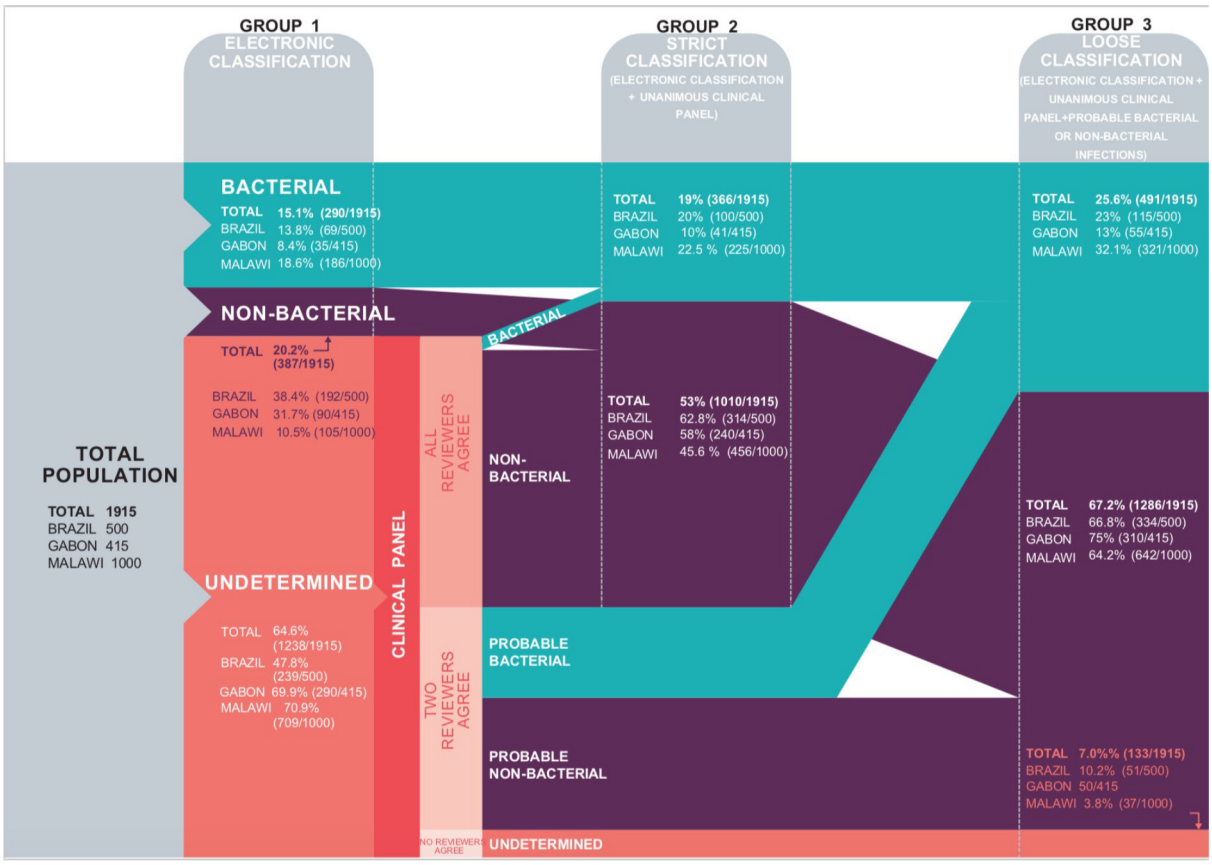

Figure 1 Classification criteria to assign bacterial versus non-bacterial infection categories for the analysis. The flows in different colours (turquoise=bacteria, purple=non-bacterial, red=undetermined) represent the proportion of patients that were assigned into the respective group (bacteria/non-bacteria/undetermined) after each classification step. Group 1 representing only patients assigned using laboratory data; group 2 representing patients with a unanimous decision after review by the clinical panel; group 3 after clinical panel review and group 3 including all patients, even if only two panel members agreed on the probable cause.

Malawi and Gabon, as in Brazil most protein biomarkers showed performances of<0.6. When the biomarker results were stratified by age, the AUROCs were slightly higher for children (≤15 years) compared with those seen for adults in the malaria-negative population (online supplemental tables 10–15). Among the malaria-positive population, WBC, lymphocyte and neutrophil counts were the best-performing biomarkers in both Gabon and Malawi (in most cases between 0.6 and 0.7).

Combinations of host biomarkers and additional covariates: multivariable analysis

The best-performing biomarkers in the univariate analysis were compared with the best performances from the multivariable analyses with several feature-selected biomarkers and covariates (table 4 and online supplemental tables 16–21). In most cases, the best combination of biomarkers showed higher AUROCs than the top-performing individual biomarkers, with a low/moderate 'gain' (range 1–13%). The best-performing AUROCs were very similar, irrespective of the multivariable model used, especially for the 'strict' and 'loose' groups (difference in AUROC range 0.02–0.03 for Malawi and Brazil). Biomarkers identified as top performing by the multivariable analyses differed depending on the model used. While SW and RFA selected three to five biomarkers or combinations, RuleFit selected more biomarkers (10 variables on average) to be part of the signature. The relatively low increase in AUROC when comparing the top-performing single biomarker with multivariable models indicates that biomarkers in addition to the single best-performing biomarker do not make a major contribution.

DISCUSSION

We present the most extensive and diverse host biomarker evaluation study to differentiate bacterial from non-bacterial infections in LMICs. The study aimed to identify if next-generation host biomarkers for distinguishing bacterial from non-bacterial cases of AFI which could replace existing biomarkers such as CRP, PCT and WBC/neutrophil assessments. The data show that none of the promising host-biomarkers exhibited high AUROCs in our non-severe AFI population in either low malaria prevalence (Brazil) or high malaria prevalence (Gabon, Malawi) settings. Haematology biomarkers and CRP were included as a baseline to identify better-performing

Table 3 Univariate analysis of 18 individual biomarkers‡ among malaria-negative patients for all three countries (A–C)

	(A) Brazil AUROC* (CI), N		
	Electronic	**Strict**	**Loose**
Haematological biomarkers			
Lymphocyte count	0.67 (0.59 to 0.74), 257	0.66 (0.59 to 0.72), 408	0.66 (0.6 to 0.72), 442
Neutrophil count	0.77 (0.7 to 0.84), 257	0.8 (0.74 to 0.86), 408	0.79 (0.73 to 0.84), 442
RBC count	0.61 (0.52 to 0.69), 258	0.58 (0.51 to 0.65), 408	0.58 (0.51 to 0.64), 442
WBC count	0.81 (0.75 to 0.87), 257	0.83 (0.77 to 0.88), 408	0.82 (0.77 to 0.87), 442
Protein biomarkers			
AGP	0.59 (0.51 to 0.68), 252	0.54 (0.47 to 0.61), 402	0.52 (0.46 to 0.59), 434
Chitinase 3-like 1	0.58 (0.5 to 0.66), 246	0.54 (0.47 to 0.6), 394	0.55 (0.49 to 0.61), 424
CRP†	0.61 (0.52 to 0.69), 259	0.61 (0.54 to 0.68), 412	0.62 (0.55 to 0.68), 446
IP-10/IP-10/CRG-2	0.6 (0.52 to 0.68), 252	0.53 (0.46 to 0.59), 402	0.53 (0.47 to 0.59), 434
Galectin-9	0.63 (0.55 to 0.71), 252	0.56 (0.49 to 0.63), 401	0.57 (0.5 to 0.63), 433
hCC2	0.51 (0.43 to 0.6), 244	0.51 (0.44 to 0.58), 392	0.52 (0.46 to 0.59), 424
HBP‡	0.67 (0.52 to 0.81), 113	0.68 (0.55 to 0.8), 144	0.64 (0.51 to 0.76), 151
HPTGN	0.48 (0.4 to 0.57), 248	0.51 (0.44 to 0.58), 398	0.51 (0.45 to 0.58), 430
IL-4	0.58 (0.5 to 0.65), 249	0.53 (0.47 to 0.59), 398	0.54 (0.48 to 0.59), 429
IL-6	0.49 (0.43 to 0.54), 247	0.49 (0.44 to 0.54), 395	0.48 (0.43 to 0.52), 426
LBP	0.58 (0.5 to 0.66), 248	0.54 (0.48 to 0.61), 397	0.52 (0.46 to 0.58), 429
Lipocalin-2/NGAL	0.49 (0.41 to 0.57), 249	0.51 (0.44 to 0.57), 396	0.51 (0.44 to 0.57), 428
sPLA/Lp-PLA2	0.54 (0.46 to 0.62), 252	0.53 (0.46 to 0.59), 402	0.52 (0.45 to 0.58), 434
TRAIL	0.56 (0.49 to 0.64), 252	0.53 (0.47 to 0.59), 402	0.53 (0.48 to 0.59), 434
	(B) Gabon AUROC (CI), N**		
	Electronic	**Strict**	**Loose**
Haematological biomarkers			
Lymphocyte count	0.58 (0.45 to 0.71), 81	0.52 (0.4 to 0.63), 167	0.55 (0.45 to 0.65), 222
Neutrophil count	0.78 (0.66 to 0.89), 80	0.72 (0.62 to 0.83), 165	0.67 (0.57 to 0.77), 219
RBC count	0.55 (0.41 to 0.68), 81	0.52 (0.41 to 0.63), 167	0.53 (0.43 to 0.63), 222
WBC count	0.67 (0.54 to 0.79), 81	0.6 (0.48 to 0.72), 167	0.61 (0.5 to 0.71), 222
Protein biomarkers			
AGP	0.77 (0.65 to 0.9), 80	0.7 (0.59 to 0.82), 163	0.65 (0.55 to 0.75), 220
Chitinase 3-like 1	0.6 (0.46 to 0.74), 79	0.6 (0.48 to 0.72), 162	0.62 (0.52 to 0.72), 217
CRP†	0.71 (0.59 to 0.82), 81	0.65 (0.55 to 0.75), 167	0.63 (0.53 to 0.72), 224
IP-10/IP-10/CRG-2	0.6 (0.48 to 0.73), 80	0.51 (0.4 to 0.62), 164	0.52 (0.43 to 0.62), 221
Galectin-9	0.7 (0.58 to 0.83), 80	0.6 (0.48 to 0.71), 163	0.54 (0.43 to 0.64), 219
hCC2	0.55 (0.41 to 0.69), 77	0.52 (0.4 to 0.64), 159	0.51 (0.41 to 0.61), 216
HBP‡			
HPTGN	0.64 (0.5 to 0.78), 77	0.62 (0.51 to 0.74), 159	0.55 (0.45 to 0.66), 214
IL-4	0.46 (0.4 to 0.52), 79	0.49 (0.45 to 0.53), 163	0.51 (0.47 to 0.55), 220
IL-6	0.51 (0.47 to 0.55), 80	0.51 (0.48 to 0.55), 164	0.51 (0.47 to 0.55), 221
LBP	0.69 (0.56 to 0.83), 78	0.67 (0.55 to 0.78), 160	0.6 (0.5 to 0.71), 217
Lipocalin-2/NGAL	0.67 (0.54 to 0.8), 79	0.6 (0.49 to 0.72), 163	0.58 (0.48 to 0.68), 219
sPLA/Lp-PLA2	0.58 (0.44 to 0.71), 80	0.54 (0.43 to 0.65), 164	0.58 (0.48 to 0.68), 221
TRAIL	0.5 (0.5 to 0.5), 74	0.5 (0.49 to 0.5), 156	0.49 (0.48 to 0.5), 212

Continued

Table 3 Continued

	(C) Malawi AUROC** (CI), N		
	Electronic	**Strict**	**Loose**
Haematological biomarkers			
Lymphocyte count	0.56 (0.47 to 0.66), 154	0.51 (0.45 to 0.58), 303	0.52 (0.47 to 0.58), 461
Neutrophil count	0.67 (0.58 to 0.77), 143	0.73 (0.67 to 0.79), 273	0.7 (0.65 to 0.76), 414
RBC count	0.46 (0.36 to 0.56), 155	0.53 (0.46 to 0.59), 305	0.56 (0.5 to 0.61), 463
WBC count	0.69 (0.6 to 0.78), 155	0.72 (0.66 to 0.78), 304	0.68 (0.63 to 0.73), 461
Protein biomarkers			
AGP	0.56 (0.46 to 0.66), 158	0.54 (0.48 to 0.6), 309	0.54 (0.49 to 0.59), 466
Chitinase 3-like 1	0.49 (0.39 to 0.59), 155	0.5 (0.43 to 0.56), 304	0.5 (0.44 to 0.55), 462
CRP†	0.55 (0.45 to 0.65), 156	0.6 (0.54 to 0.67), 305	0.58 (0.53 to 0.63), 462
IP-10/IP-10/CRG-2	0.66 (0.56 to 0.75), 158	0.6 (0.53 to 0.66), 309	0.61 (0.56 to 0.66), 466
Galectin-9	0.71 (0.62 to 0.8), 158	0.61 (0.55 to 0.67), 309	0.63 (0.57 to 0.68), 466
hCC2	0.59 (0.49 to 0.69), 158	0.55 (0.49 to 0.62), 309	0.55 (0.5 to 0.6), 466
HBP‡	0.53 (0.39 to 0.68), 63	0.55 (0.44 to 0.66), 106	0.52 (0.41 to 0.63), 124
HPTGN	0.54 (0.45 to 0.64), 157	0.51 (0.45 to 0.58), 307	0.51 (0.46 to 0.57), 464
IL-4	0.48 (0.4 to 0.57), 157	0.48 (0.42 to 0.53), 306	0.47 (0.42 to 0.51), 463
IL-6	0.56 (0.47 to 0.65), 158	0.61 (0.55 to 0.67), 307	0.59 (0.54 to 0.64), 465
LBP	0.52 (0.42 to 0.61), 157	0.54 (0.47 to 0.61), 267	0.53 (0.47 to 0.59), 394
Lipocalin-2/NGAL	0.56 (0.46 to 0.66), 156	0.65 (0.59 to 0.72), 265	0.61 (0.56 to 0.67), 392
sPLA/Lp-PLA2	0.58 (0.47 to 0.68), 158	0.55 (0.49 to 0.61), 308	0.56 (0.51 to 0.61), 466
TRAIL	0.61 (0.51 to 0.71), 157	0.62 (0.56 to 0.68), 306	0.62 (0.57 to 0.67), 463

Common biomarkers such as CRP and haematological biomarkers were included for reference. In this context, we defined performance as follows: dark blue (AUROC≥0.7), light blue (AUROC>0.65 and <0.7), orange (AUROC 0.6–0.65) and red (AUROC<0.6)
*AUROC has a value between 0 and 1, where 1 corresponds to an effect classifier, 0.5–1 that assigns classes randomly.
†CRP was measured with a NycoCard device.
‡Freeze-thaw experiments to evaluate the stability of the biomarkers after five cycles (referred to as 'treated') were performed with Luminex 9-plexes and 2-plexes. Three samples each were freeze-thawed up to six times and compared with samples after the first thawing (referred to as 'untreated'; biomarkers were considered stable with 80–120% recovery). Samples were analysed in triplicate and showed good stability up to five freeze-thaw cycles for all analytes showing acceptable results, except for the C2 and C4b biomarkers (C2: 2/3 (66.7%) samples were stable; C4b: two samples failed the sixth freeze-thaw cycle). As a result, these biomarkers were excluded as they would never be suitable as the basis of a diagnostic test.
§HBP was evaluated in a small group of patients in Malawi and Brazil; however, HBP did not show promise and was not evaluated further.
AGP, A-1-acid glycoprotein; AUROC, area under the receiver operating characteristic; CRP, C-reactive protein; HBP, heparin-binding protein; IL-4, interleukin-4; IP-10, gamma-induced protein 10; LBP, lipopolysaccharide binding protein; NGAL, neutrophil gelatinase-associated lipocalin; RBC, red blood cell; TRAIL, TNF-related apoptosis-inducing ligand; WBC, white blood cell.

markers; however, they remain those with the highest AUROC values (approximately 0.60–0.70 AUROC) in our population.

Overall, the performance of all markers was underwhelming, yet not surprising. It aligns with previous data where a marked reduction in performance was observed when shifting the population from inpatients to outpatients.[16–18] Previously, it was hypothesised that the decrease in performance in host biomarkers between HIC and LMIC settings or even between Africa and Asia, was due to the untreated comorbidities (eg, diabetes, malaria, neglected tropical diseases) which contribute to inflammation and the non-specific triggering of host biomarkers, unrelated to the current acute presentation.[18 19] In our data the performance was indeed poorer in malaria-positive patients (AUROC<0.6); however, even in the malaria-negative population, biomarkers showed low performances (~0.6–0.7) in our cohort. Similarly, sex and arboviral status appeared to have no major effect on biomarker performance. Our data notably indicated that combining biomarkers can enhance performance. However, this improvement was not consistently observed. When combining several biomarkers and additional covariates, the 'gain' in AUROC values was low/moderate (range 1–13%) compared with the top-performing individual biomarkers. From a diagnostic development perspective, a low gain in performance would not justify the additional complexity and cost of developing a simple multiplex test.

Table 4 Multivariable analysis of biomarkers among malaria-negative patients, including the gain/loss of performance when comparing multivariable analysis and single-host biomarkers comprising both haematological and protein host biomarkers

Classification group	Best multivariable model/models: mean (min–max) AUROC	Best host biomarker: mean (min–max) AUROC	Multivariable AUROC gain/loss (%) ‡ multivariable and single host biomarkers ratio
Overall (Brazil + Gabon + Malawi)†			
L	SW/RFA/RF:0.75 (0.69–0.81)	WBC count: 0.7 (0.64–0.76)	+7
S	SW:0.83 (0.75–0.91)	WBC count: 0.78 (0.72–0.84)	+6
E	SW/RFA:0.83 (0.77–0.89)	WBC count: 0.77 (0.69–0.85)	+8
Brazil			
L	SW: 0.82 (0.70–0.94)	WBC count: 0.8 (0.68–0.92)	+2.5
S	RFA: 0.82 (0.70–0.94)	WBC count: 0.8 (0.68–0.92)	+2.5
E	SW: 0.85 (0.73–0.97)	WBC count: 0.83 (0.69–0.97)	+2
Gabon*			
L	SW/RFA: 0.7 (0.46–0.94)	WBC count: 0.7 (0.64–0.76)	
S	SW/RFA: 0.76 (0.52–0.96)	WBC count: 0.78 (0.72–0.84)	–3
E	RFA: 0.77 (0.63–0.91)	WBC count: 0.77 (0.69–0.85)	
Malawi			
L	SW/RFA: 0.74 (0.62–0.86)	neutrophil count: 0.72 (0.66–0.78)	+3
S	SW: 0.73 (0.61–0.85)	neutrophil count: 0.72 (0.58–0.86)	+ 1
E	RFA: 0.72 (0.60–0.84)	WBC count: 0.7 (0.56, 0.84)	+ 2

Dark Blue (gain, ie, the multivariable models show better performances than univariate models); red (loss, ie, the univariate models show better performances than multivariable models).
*Multivariable performances for Gabon were computed using as a predictor model the model trained in the 'Overall' scenario (all participants from the three analysed countries) then evaluated using Gabon data only. Indeed, the sample size of Gabon data was not sufficient to allow the development of a reliable model specific to this country.
†In the 'Overall' scenario, the model was developed using the data of all countries and the variable indicating the country was used as a covariate in the model.
‡Performance comparison was computed as: ((multivariable AUROC – univariate AUROC) / univariate AUROC) × 100.
AUROC, area under the receiver operating characteristic; E, electronic classification group; L, loose classification group; RF, RuleFit; RFA, logistic recursive feature addition; S, strict classification group; SW, stepwise logistic regression; WBC, white blood cell.

Adding to the challenges of host biomarker studies is the lack of consistent reference standards and that most studies have focused their analyses solely on the subpopulation of patients with a microbiologically confirmed diagnosis. This approach ignores the largest group (>70%) of patients and the intended-use population of any future test.[20] The group with laboratory-confirmed diagnosis will decrease further in the non-severe AFI population; presenting at the primary care level. Going forward more clarity will likely follow as a recent host-biomarker test (BVtest, MeMed, Israel) was approved by the Food and Drug Administration (FDA) and subsequent guidance will prescribe more clearly how studies have to be designed to standardise the classification of 'bacterial' versus 'non-bacterial' evaluated to guide prescribing for bacterial or non-bacterial infections.[9 21] Our protocol is aligned with the FDA-approved classification hence we are confident our methodology is robust.

While our study aimed to mitigate the challenges described, it still had several limitations. The study did not include a control group, so no baseline information was available for biomarker performance or asymptomatic carrier populations. The enrolment period in Brazil and Gabon lasted for less than 1 year and given the heterogeneity of causes of AFI across time, the performance of the biomarkers may not be generalisable to different times of the year and geographical settings, particularly in Asia. The study used a two-step process to classify outcomes and the clinical classification based on recorded clinical information may have introduced subjectivity. Notably, clinicians had access to the haematology biomarker results (WBCs, neutrophils) during outcome classification which might have introduced a bias in favour of these biomarkers. However, comparing AUROCs between all classification groups (E, L, S) suggests this potential bias had no major impact as the results are similar across groups. There were some heterogeneities in the inclusion criteria across the various study sites, including age groups and fever criteria. In Brazil and Gabon, the inclusion criterion was a history of fever in the past 7 days, while it was fever at presentation in Malawi. Studies have found that acute fever at presentation has implications for the interpretation of host biomarkers;[22] however, our subanalysis by acute fever showed no differences, so we do not consider

that these different inclusion criteria impacted interpretation. Despite best efforts to standardise procedures, there was a level of adaptability required in the choice of testing methods by the clinical teams in each country, for arbovirus and respiratory pathogen detection. Further, the choice to follow the TPP and focus on non-severe patients in the recruitment was based on the need's definition by the WHO and others, while this still holds as a major priority, in hindsight this focus did not allow us to stratify by severity (eg, Sequential Organ Failure Assessment (SOFA) score).

Overall, the results of this diverse study highlight the difficulties in identifying single-host biomarkers or simple-host biomarker combinations that can help solve the problem of undifferentiated prescribing in primary healthcare, particularly to be used across diverse global settings. On the eighth birthday of the original TPP for a diagnostic assay to distinguish bacterial and non-bacterial infections in resource-limited settings, a more recent consultation confirmed that the need for such an assay remains and is in fact increasingly urgent.[6 23] Yet again, the consultation concluded primary healthcare clinics and their equivalents must have the ability to perform tests other than just malaria RDTs.[23] The lack of diagnostics infrastructure at the lower levels of health systems is well documented and requires urgent improvement to support medical staff in their decision-making. While no novel host biomarker assay meets these needs, evidence for existing biomarkers, for example, CRP, and various haematology biomarkers, should be used to drive such improvements, although using slightly different approaches and cut-offs across settings. In addition to using existing tools, increased investment into lower-level health infrastructures is critical and the first step to improved care. Recent studies have shown that even simple-host biomarkers, such as CRP, can have a major impact on how clinical staff use antibiotics.[24–26] The current study confirms that the existing biomarkers are imperfect and hence should only be used as guidance, in conjunction with expanded clinical algorithms.[27 28] Such guidelines, alongside adopted policies, strengthened infrastructures and accessible haematology/biochemistry data could enable healthcare workers to use simple tools to gain additional data points to help form a more evidence-based diagnosis that has to be guided by the local epidemiology. Optimising existing haematology or biochemistry tools and their maintenance requirements to meet the needs of low-resourced settings could be one step towards more expanded use of these well-known markers. In conclusion, our study reinforces the continued need for innovation in the host biomarker space and highlights the importance of targeted evaluations of such innovations, in diverse intended-use settings, to fully understand their true value.

Author affiliations
[1]Malaria and Fever Programme, FIND, Geneva, Switzerland
[2]Department of Information Engineering, University of Padova, Padova, Italy
[3]Malawi Epidemiology and Intervention Research Unit (MEIRU), Chilumba, Karonga, Malawi
[4]Faculty of Epidemiology and Population Health, London School of Hygine and Tropical Medicine (LSTMH), London, UK
[5]Infection Prevention and Control Unit, University Hospital Düsseldorf, Düsseldorf, Germany
[6]Centre de Recherches Medicales de Lambarene (CERMEL), Lambarene, Moyen-Ogooué, Gabon
[7]Department of Parasitology, University Medical Center, Leiden, Leiden, Netherlands
[8]Evandro Chagas Natl Inst Infect Dis, Rio de Janeiro, Brazil
[9]Universitätsklinikum Tübingen, Tubingen, Germany
[10]Instituto Nacional de Infectologia Evandro Chagas (INI), FIOCRUZ, Rio de Janeiro, Brazil
[11]Clinical Trials & Pharmacovigilance, Instituto Butantan, São Paulo, São Paulo, Brazil
[12]Laboratorio de Vírus Respiratórios e Sarampo, FIOCRUZ, Rio de Janeiro, Brazil
[13]Laboratorio de Flavivirus, FIOCRUZ, Rio de Janeiro, Brazil
[14]Laboratorio de Bacteriologia e Bioensaios em Microbacterias, Instituto Nacional de Infectologia Evandro Chagas,Fundação Oswaldo Cruz, Rio de Janeiro, Brazil
[15]NMI Natural and Medical Sciences Institute at the University of Tübingen, Reutlingen, Germany
[16]Data Science Unit, FIND, Geneva, Switzerland
[17]University of Oxford Centre for Tropical Medicine and Global Health, Oxford, UK
[18]Mahidol Oxford Tropical Medicine Research Unit, Bangkok, Thailand
[19]European Campus Rottal Inn, Deggendorf Institut of Technology, Pfarrkirchen, Germany

X Andre M Siqueira @siqueira__andre, José Moreira @j_moreira_ and Sabine Dittrich @sabinedittrich

Acknowledgements We would like to thank a group of dedicated colleagues (Quique Bassat, Heidi Hopkins, Valerie D'Acremont) for critical review of the data and continued discussions regarding analysis and interpretation. Further thanks to Luciano S. Oliveira, Cintia Damasceno dos Santos Rodrigues and Carolina Cardoso dos Santos for supporting the work in Brazil. Medical writing as well as editorial support, under the direction of the authors, was provided by Adam Bodley, funded by FIND, the global alliance for diagnostic in accordance with Good Publication Practice guidelines.

Contributors SD, CE, SO, AM, AMS, SG, STA, MM-L conceptualised the study and study design; CE, SG, STA, AMS, JKM, VH, ALK, AA, JCBO, MM-L, PNE, JAM, PB, LB, AdRM, BCC, MMS, AMBdF, EAdSRdS, MCSL, JH, AG, MJ, NS-M, CH implemented the study and data collection; MA, MV, SL, SO, BDC, BLF-C, SD, SP, SG, AMS, STA, SJL conducted data analysis and interpretation. BLF-C, SD wrote the first draft of the manuscript and all authors contributed to the final version of the manuscript. Guarantor is SD.

Funding Funding for this work was provided to FIND, the global alliance for diagnostics, by the governments of the Netherlands, the Foreign, Commonwealth and Development Office of the UK, and Australia Aid. The funding organisations had no role in the study design, data collection, analysis and interpretation of data. Further they had no role in writing of the report or decision to submit for publication

Competing interests SD, BLFC, CE, VH, SO, CH, AM, SL are or were employed by FIND, the global alliance for diagnostic during the study period. All other authors do not declare any competing interests.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The study protocol was submitted to ClinicalTrials.gov (NCT03047642) and ethical approval was obtained from all relevant institutional committees in Brazil (Research Ethics Committee of INI-FIOCRUZ and Comissão Nacional de Ética em Pesquisa, Ref:2.235.565; National Research Ethics Committee), Gabon (Comité National d'Ethique pour la Recherche, RefNr:N°0078/2019PR/SG/CNER) and Malawi (National Health Science Research Committee, Approval Nr: 16/9/1668; Observational and Intervention Research Ethics Committee of the London School of Hygiene and Tropical Medicine, UK, LSTMH Ref: 11974) and all details of the design have been previously published.[10] Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data are available upon request from the study sites (Malawi, Brazil, Gabon) or the study sponsor (FIND). All data were shared with the local teams to allow further analysis and investigations and acknowledge ownership.

ORCID iDs
Camille Escadafal http://orcid.org/0000-0002-0268-750X
Patrícia Brasil http://orcid.org/0000-0001-9555-7976
Sue J Lee http://orcid.org/0000-0002-8409-4248
Sabine Dittrich http://orcid.org/0000-0002-4522-2788

REFERENCES

1 Rhee C, Kharod GA, Schaad N, *et al*. Global knowledge gaps in acute febrile illness etiologic investigations: A scoping review. *PLoS Negl Trop Dis* 2019;13:e0007792.
2 Dalrymple U, Cameron E, Bhatt S, *et al*. Quantifying the contribution of *Plasmodium falciparum* malaria to febrile illness amongst African children. *Elife* 2017;6:e29198.
3 Boschi-Pinto C, Labadie G, Dilip TR, *et al*. Global implementation survey of Integrated Management of Childhood Illness (IMCI): 20 years on. *BMJ Open* 2018;8:e019079.
4 Hopkins H, Bruxvoort KJ, Cairns ME, *et al*. Impact of introduction of rapid diagnostic tests for malaria on antibiotic prescribing: analysis of observational and randomised studies in public and private healthcare settings. *BMJ* 2017;356:j1054.
5 van Griensven J, Cnops L, De Weggheleire A, *et al*. Point-of-Care Biomarkers to Guide Antibiotic Prescription for Acute Febrile Illness in Sub-Saharan Africa: Promises and Caveats. *Open Forum Infect Dis* 2020;7:ofaa260.
6 Dittrich S, Tadesse BT, Moussy F, *et al*. Target Product Profile for a Diagnostic Assay to Differentiate between Bacterial and Non-Bacterial Infections and Reduce Antimicrobial Overuse in Resource-Limited Settings: An Expert Consensus. *PLoS One* 2016;11:e0161721.
7 Langford BJ, So M, Raybardhan S, *et al*. Antibiotic prescribing in patients with COVID-19: rapid review and meta-analysis. *Clin Microbiol Infect* 2021;27:520–31.
8 Leticia Fernandez-Carballo B, Escadafal C, MacLean E, *et al*. Distinguishing bacterial versus non-bacterial causes of febrile illness - A systematic review of host biomarkers. *J Infect* 2021;82:1–10.
9 Escadafal C, Nsanzabana C, Archer J, *et al*. New Biomarkers and Diagnostic Tools for the Management of Fever in Low- and Middle-Income Countries: An Overview of the Challenges. *Diagnostics (Basel)* 2017;7:44.
10 Escadafal C, Geis S, Siqueira AM, *et al*. Bacterial versus non-bacterial infections: a methodology to support use-case-driven product development of diagnostics. *BMJ Glob Health* 2020;5:e003141.
11 Ashkenazi-Hoffnung L, Oved K, Navon R, *et al*. A host-protein signature is superior to other biomarkers for differentiating between bacterial and viral disease in patients with respiratory infection and fever without source: a prospective observational study. *Eur J Clin Microbiol Infect Dis* 2018;37:1361–71.
12 Kapasi AJ, Dittrich S, González IJ, *et al*. Host Biomarkers for Distinguishing Bacterial from Non-Bacterial Causes of Acute Febrile Illness: A Comprehensive Review. *PLoS One* 2016;11:e0160278.
13 Friedman JH, Popescu BE. Predictive learning via rule ensembles. *Ann Appl Stat* 2008;2:916–54.
14 Vettoretti M, Di Camillo B. A Variable Ranking Method for Machine Learning Models with Correlated Features: In-Silico Validation and Application for Diabetes Prediction. *Appl Sci (Basel)* 2021;11:7740.
15 Buuren S, mice G-O. Multivariate Imputation by Chained Equations in R. *J Stat Softw* 2011;45:1–67.
16 Díez-Padrisa N, Bassat Q, Morais L, *et al*. Procalcitonin and C-reactive protein as predictors of blood culture positivity among hospitalised children with severe pneumonia in Mozambique. *Trop Med Int Health* 2012;17:1100–7.
17 Hildenwall H, Muro F, Jansson J, *et al*. Point-of-care assessment of C-reactive protein and white blood cell count to identify bacterial aetiologies in malaria-negative paediatric fevers in Tanzania. *Trop Med Int Health* 2017;22:286–93.
18 Lubell Y, Blacksell SD, Dunachie S, *et al*. Performance of C-reactive protein and procalcitonin to distinguish viral from bacterial and malarial causes of fever in Southeast Asia. *BMC Infect Dis* 2015;15:511.
19 Díez-Padrisa N, Bassat Q, Machevo S, *et al*. Procalcitonin and C-reactive protein for invasive bacterial pneumonia diagnosis among children in Mozambique, a malaria-endemic area. *PLoS One* 2010;5:e13226.
20 Mueller TC, Siv S, Khim N, *et al*. Acute undifferentiated febrile illness in rural Cambodia: a 3-year prospective observational study. *PLoS One* 2014;9:e95868.
21 Food, Administration Department. 510(k) substantial equivalence determination decision summary - assay and instrument, Available: https://www.accessdata.fda.gov/cdrh_docs/reviews/K210254.pdf
22 Shapiro NI, Self WH, Rosen J, *et al*. A prospective, multi-centre US clinical trial to determine accuracy of FebriDx point-of-care testing for acute upper respiratory infections with and without a confirmed fever. *Ann Med* 2018;50:420–9.
23 Unitaid, FIND. Biomarkers for acute febrile illness at the point-of-care in low-resource settings, 2021. Available: https://www.finddx.org/wp-content/uploads/2021/05/Meeting-report_Biomarkers-for-acute-febrile-illness-at-the-point-of-care-in-low-resource-settings.pdf
24 Althaus T, Greer RC, Swe MMM, *et al*. Effect of point-of-care C-reactive protein testing on antibiotic prescription in febrile patients attending primary care in Thailand and Myanmar: an open-label, randomised, controlled trial. *Lancet Glob Health* 2019;7:S2214-109X(18)30444-3:e119–31:.
25 Thi Thuy Do N, Greer RC, Lubell Y, *et al*. Implementation of C-reactive protein point of care testing to improve antibiotic targeting in respiratory illness in Vietnamese primary care (ICAT): a study protocol for a cluster randomised controlled trial. *BMJ Open* 2020;10:e040977.
26 Van Duffel L, Yansouni CP, Jacobs J, *et al*. Accuracy of C-reactive Protein and Procalcitonin for Diagnosing Bacterial Infections Among Subjects With Persistent Fever in the Tropics. *Open Forum Infect Dis* 2022;9:ofac434ofac434.
27 Keitel K, D'Acremont V. Electronic clinical decision algorithms for the integrated primary care management of febrile children in low-resource settings: review of existing tools. *Clin Microbiol Infect* 2018;24:S1198-743X(18)30352-5:845–55:.
28 Pellé KG, Rambaud-Althaus C, D'Acremont V, *et al*. Electronic clinical decision support algorithms incorporating point-of-care diagnostic tests in low-resource settings: a target product profile. *BMJ Glob Health* 2020;5:e002067.
29 Venge P, Douhan-Håkansson L, Garwicz D, *et al*. Human Neutrophil Lipocalin as a Superior Diagnostic Means To Distinguish between Acute Bacterial and Viral Infections. *Clin Vaccine Immunol* 2015;22:1025–32.