

1 **Analytical performance of 17 commercially available**
2 **point-of-care tests for CRP to support patient**
3 **management at lower levels of the health system**

4

5

6 Serafina Calarco^{1*}, B. Leticia Fernandez-Carballo¹, Thomas Keller², Stephan Weber²,
7 Meike Jakobi³, Patrick Marsall³, Nicole Schneiderhan-Marra³, Sabine Dittrich¹

8

9

10 ¹FIND, Geneva, Switzerland

11 ²ACOMED statistik, Leipzig, Germany

12 ³NMI Natural and Medical Sciences Institute at the University of Tübingen, Reutlingen,
13 Germany

14

15 *Corresponding author

16 Serafina.calarco@finddx.org

17

18 **Abstract**

19 Accurate and precise point-of-care (POC) testing for C-reactive protein (CRP) can help
20 support healthcare providers in the clinical management of patients. Here, we compared
21 the analytical performance of 17 commercially available POC CRP tests to enable more
22 decentralized use of the tool. The following CRP tests were evaluated. Eight quantitative
23 tests: QuikRead go (Aidian), INCLIX (Sugentech), SpinIt (Biosurfit), LS4000 (Lansionbio),
24 GS 1200 (Gensure Biotech), Standard F200 (SD Biosensor), Epithod 616 (DxGen), IFP-
25 3000 (Xincheng Biological); and nine semi-quantitative tests: Actim CRP (ACTIM),
26 NADAL Dipstick (nal von minden), NADAL cassette (nal von minden), ALLTEST Dipstick
27 (Hangzhou Alltest Biotech), ALLTEST Cassette cut-off 10-40-80 (Hangzhou Alltest
28 Biotech), ALLTEST Cassette cut-off 10-30 (Hangzhou Alltest Biotech), Biotest (Hangzhou
29 Biotest Biotech), BTNX Quad Line (BTNX), BTNX Tri Line (BTNX). Stored samples
30 (n=660) had previously been tested for CRP using Cobas 8000 Modular analyzer (Roche
31 Diagnostics International AG, Rotkreuz, Switzerland (reference standards). CRP values
32 represented the clinically relevant range (10-100 mg/L) and were grouped into four
33 categories (<10 mg/L, 10–40 mg/L or 10-30 mg/L, 40–80 mg/L or 30-80 mg/L, and >
34 80mg/L) for majority of the semi-quantitative tests. Among the eight quantitative POC
35 tests evaluated, QuikRead go and SpinIt exhibited better agreement with the reference
36 method, showing slopes of 0.963 and 0.921, respectively. Semi-quantitative tests with the
37 four categories showed a poor percentage agreement for the intermediate categories and
38 higher percentage agreement for the lower and upper limit categories. Analytical
39 performance varied considerably for the semi-quantitative tests, especially among the
40 different categories of CRP values. Our findings suggest that quantitative tests might

41 represent the best choice for a variety of use cases, as they can be used across a broad
42 range of CRP categories.

43

44 Introduction

45 C-reactive protein (CRP) is known to be an acute phase reactant, a glycoprotein
46 produced by the liver and released into the blood stream within a few hours of a tissue
47 injury occurring, at the start of an infection, or due to other sources of inflammation [1].
48 CRP levels are typically below 3 mg/L in healthy patients, from 10 to 100 mg/L during a
49 mild infection, and as high as 500 mg/L in patients experiencing a severe inflammatory
50 response.

51 Point-of-care (POC) tests for CRP are increasingly being used in primary care to assist
52 general practitioners (GPs) in the diagnostic workup for various health complaints,
53 including acute cough and abdominal pain, and to differentiate between mild and severe
54 respiratory tract infections [2][3][4]. These tests can be performed using capillary blood
55 samples, with results available within minutes. The use of POC tests for CRP has been
56 widely implemented and is standard practice in many high-income countries to guide the
57 use of treatment for respiratory tract infections, including in Norway and Sweden [5], while
58 in England these tests are recommended by Public Health England and the National
59 Institute for Health and Care Excellence (NICE) [6]. The measurement of CRP values has
60 also proved useful for excluding severe appendicitis or diverticulitis in patients attending
61 with abdominal pain at emergency rooms [7][8][9].

62 The diagnostic relevance of CRP testing has also been extensively documented for
63 distinguishing between bacterial and viral infections [10][11]. A study from 2013 of Peng,
64 highlighted that CRP levels > 19.6 mg/L (in the serum) might indicate a bacterial infection
65 and guide antibiotic prescription in patient presenting chronic obstructive pulmonary
66 disease [12]. A study from Korppi and coworkers point out that 40 mg/L could be more

67 efficient than 20 mg/L or 80 mg/L for differentiation between viral and bacterial infection,
68 in children with middle or lower respiratory tract infection [13]. These levels provide
69 clinicians with an additional data point, assisting them in forming a correct medical
70 diagnosis. In the Netherlands, for example, a patient's CRP value must be known before
71 antibiotics can be prescribed to patients with a lower respiratory tract infection [14].

72 Uncertainty about the cause of an infection can lead to inappropriate antibiotic
73 prescribing, overuse of resources, and disease complications [15] [16]. This is particularly
74 true in low- and middle-income settings, where the presence of wide-ranging
75 etiologies (parasitic, fungal, bacterial, or viral pathogens) and the lack of adequate
76 diagnostic facilities often leads to the choice of therapy being based on empirical
77 knowledge [3][17]. Antibiotics are often dispensed without diagnostic guidance, leading
78 to large quantities of antibiotics being used, which is linked to increasing levels of
79 antimicrobial resistance (AMR) [18][19].

80 In addition to guiding the use of antibiotics around the world to tackle the growing AMR
81 crisis [17], CRP has recently also been recommended by the World Health Organization
82 (WHO) [20] as a screening tool for tuberculosis (TB) [21]. For this specific use case a cut
83 off of > 5 mg/L was advised because it is the lowest threshold indicating anomaly in many
84 clinical settings [20].

85 This greater understanding globally of the potential benefits of using CRP in resource-
86 limited settings for different use cases means that POC devices are becoming
87 increasingly important for the global South. As with all diagnostics, the beneficial effects
88 of measuring CRP are linked to the quality of data produced when using the diagnostic
89 product [22][23][24]. Currently, there is limited evidence available as to how well

90 commonly used POC tests perform when compared with central reference laboratory
91 testing [25][26], critical information that is required when planning any expansion of the
92 use of such tests.

93 The purpose of this study, therefore, is to evaluate the analytical performance of selected
94 POC CRP tests under ideal conditions. We hope that our data will help to guide decision-
95 makers when selecting the most appropriate test for a specific use case. We compared
96 semi-quantitative and quantitative POC tests for CRP with an enzyme-linked
97 immunosorbent assay (ELISA) for CRP and determined the coefficient of variance. As
98 each CRP use-case requires slightly different cut-off values [27], one overarching
99 analysis using a common cut-off value was performed, with the goal of informing
100 programs and interventions across disease areas.

101 **Materials and methods**

102 **Criteria for selecting tests**

103 We conducted extensive searches of public databases and websites of commercial
104 companies to identify CRP tests currently available on the market. In total, 33 quantitative
105 POC tests, 21 semi-quantitative POC tests, and 2 qualitative lateral flow tests were pre-
106 selected. Predefined go/no go criteria around the availability of tests were then applied to
107 determine whether a particular POC test should be included in the study (Fig 1). The tests
108 that passed this first screening were then evaluated based on cost/market requirements
109 and test characteristics (Fig 1), each criterion was scored and ranked, and tests with the
110 overall highest scores were included in the study. The maximum number of tests to be

111 included in the study was determined by logistical and budgetary considerations. We
112 evaluated eight quantitative tests: QuikRead go (Aidian), INCLIX (Sugentech), Spinit
113 (Biosurfit), LS4000 (Lansionbio), GS 1200 (Gensure Biotech), Standard F200 (SD
114 Biosensor), Epithod 616 (DxGen), and IFP-3000 (Xincheng Biological); and nine semi-
115 quantitative tests: Actim CRP (ACTIM), NADAL Dipstick (Nal von minden), NADAL
116 cassette (Nal von minden), ALLTEST Dipstick (Hangzhou Alltest Biotech), ALLTEST
117 Cassette cut-off 10-40-80 (Hangzhou Alltest Biotech), ALLTEST Cassette cut-off 10-30
118 (Hangzhou Alltest Biotech), Biotest (Hangzhou Biotest Biotech), BTNX Quad Line
119 (BTNX), and BTNX Tri Line (BTNX). Details about each test are provided in Table 1.

120 **Fig 1. Flowchart for tests selection.** Go/no go criteria included both objective and subjective criteria:
121 Whether the test was CE-marked, relevant detection range (from 10 to 100 mg/L CRP for a quantitative
122 test, and one cut-off value of ≥ 20 mg/L for qualitative and semi-quantitative tests), worldwide distribution,
123 and responsiveness of the manufacturer. Ranking criteria included: Additional regulatory certification
124 (higher score for US Food and Drug Administration-approved products); cost of consumables (lower score
125 for test price >10 USD, higher score for test price <1 USD); assay sample type (lower score for tests
126 requiring serum samples only, higher score for tests that can be used with capillary whole blood, venous
127 blood, or plasma); sample volume (lower score for tests requiring a volume of >50 μ L, higher score for tests
128 requiring a volume of <10 μ L); time to result (lower score for tests requiring >20 min, higher score for tests
129 requiring <5 min); storage temperature (lower score for tests requiring $<4^{\circ}\text{C}$, higher score for tests stable
130 up to 40°C); shelf-life (lower score for tests stable for <12 months, higher score for tests stable for >24
131 months).

132 **Table 1. Detailed information on evaluated index tests, A) quantitative and B) semi-quantitative**

Type of test	Quantitative							
Analyzer name (manufacturer)	QuikRead go (Aidian)	INCLIX (Sugentech)	SpinIt (Biosurfit)	LS 4000 (LANSIONBIO)	GS1200 (GenSure Biotech)	STANDARD F 200 (SD BIOSENSOR)	EPITHOD 616 (DXGEN)	IFP-3000 (Xincheng Biological)
Analytical range (AR) (mg/L)	5–120 (plasma and serum) 5–150 (whole blood)	2.5–300	2–300	0.5–200	0.5–200	1–130 (plasma and serum) 1–150 (whole blood)	2.5–200	0.5–200
Sample type	WB*, serum, plasma	WB, serum, plasma	WB, serum, plasma	WB, serum, plasma	WB, serum, plasma	WB, serum, plasma	WB, serum, plasma	WB, serum, plasma
Sample volume	20 µL	5 µL	8 µL	5 µL	5 µL	5 µL	5 µL	5 µL
Storage temperature for consumables	2–8°C	2–30°C	2–8°C	2–30°C	2–30°C	2–30°C	2–10 °C	2–30°C
Shelf-life of consumables		24 months	12 months			18 months		
Estimated preanalytical time	<1 min	1 min	5 min	1 min	3 min	3 min	2 min	1 min
Time to result	2 min	5 min	4 min	2 min	3 min	3 min	10 sec	3 min
Ability to determine other biomarker(s), yes/no, specification	Yes CRP+Hb, wrCRP wrCRP+Hb, HbA1c, STREP A, iFOBT,	Yes hsCRP, PCT, Troponin I, HbA1c, β-hCG, Total IgE	Yes BC, HbA1c, COVID-19 Antibody	Yes PCT, SAA, IL-6, HbA1c, PSA, PG I and PG II, TT3, COVID-19 antigen	Yes PCT, SAA, IL-6, HBP, TNF α, Fungal 1,3 β-D glucan, cTn I, NT-proBNP, BNP, cTnI/Myo/CK-MB, HFABP, D-D, MPO, Lp-PLA2, ST2, S100-β, IGFBP-1, PLGF, β- HCG, FSH, LH, PROG, TES, HGH, AMH, INH B, PRL, E2, VB12, FA, 25-OH- VD, SF, TRF, BGP, PTH, PG I, PG II, GP-17, NGAL, Cys C, β2-MG, RBP, MAU, T3, T4, TSH, TG, HbA1c, INS, C-peptide	Yes U-Albumin FIA, HbA1c	Yes U-Albumin, HbA1c, Hb, COVID-19 antigen, COVID-19 IgM/IgG, COVID-19 IgM	Yes HsCRP, PCT, IL-6, SAA, cTnI, MYO/CK-MB/cTnI, NT-proBNP, D-Dimer, COVID-19 IgG/M, COVID-19 Antigen, COVID-19 Neutralizing Antibody
Retail price, consumables (USD)	<3	<3	<6	<1	<1	<3	<2	<1
Retail price, analyzer (USD)	<1300	<1000	<2000	<600	<1000	<2000	<500	<500
Analyzer size (mm)	190 x 140 x 80	230 x 250 x 250	n/a	191 x 84.5 x 45	n/a	200 x 240 x 205	145 x 210 x 240	168 x 70 x 44
Analyzer weight	0.4287 kg	1.93 kg	n/a	0.6 kg	n/a	2.7 kg	1.8 kg	0.3 kg
Data export and connectivity	UBS LIS ^b /HIS ^c	USB (PDF, Excel) Email and LIS	UBS, LIS/HIS	Data transfer via USB not possible	Data transfer as PDF or xls via USB not possible Storage via LIS	UBS, LIS/HIS	USB, Wifi LIS/HIS	Data transfer via USB not possible
Practical aspects of the test ^a	- Analyzer easy to set up and operate - Many languages available - High-throughput measurements are difficult to perform because of the incubation time of 2 min	-Analyzer easy to handle -“Quick Mode” suitable for testing many samples	- Only operated in standard mode, less suitable for many samples	-Portable device -“Quick Mode” allows processing of many samples. -Low battery capacity -Touchscreen difficult to use with gloves	-Simple interface -Very fast workflow possible due to having six cassette slots	-Only operated in standard mode, less suitable for many samples -Single results cannot be exported	-Digital data export not possible. -Sample preparation with three different reagents (time consuming and comparatively high consumption of pipette tips)	-Fast test processing, suitable for many tests (Quick Mode) -No USB. -Touchscreen keyboard small -Sample preparation complex

134

135 B)

Type of test	Semi-quantitative								
Test name (manufacturer)	ACTIM CRP dipstick (ACTIM)	NADAL Dipstick (Nal von minden)	NADAL Cassette (Nal von minden)	ALLTEST Dipstick (Hangzhou Alltest Biotech)	ALLTEST Cassette (Hangzhou Alltest Biotech)	ALLTEST Cassette (Hangzhou Alltest Biotech)	BIOTEST Cassette (Hangzhou Biotest Biotech)	BTNX Quad Line dipstick (BTNX)	BTNX Tri Line Cassette (BTNX)
Cut-off (mg/L)	≥10 – 40 – ≤80	≥10 – 40 – ≤80	≥10 – 40 – ≤80	≥10 – 40 – ≤80	≥10 – 40 – ≤80	≥10 – ≤30	≥10 – 40 – ≤80	≥10 – 40 – ≤80	≥10 – 30 – ≤80
Sample type	WB, serum, plasma	WB, serum, plasma	WB, serum, plasma	WB, serum, plasma	WB, serum, plasma	WB, serum, plasma	WB, serum, plasma	WB, serum, plasma	WB, serum, plasma
Sample volume	10 µL	10 µL WB 5 µL S/P	10 µL WB 5 µL S/P	10 µL	10 µL	10 µL	10 µL WB 5 µL S/P	10 µL WB 5 µL S/P	10 µL WB 5 µL S/P
Storage temperature for consumables	2–25°C	2–30°C	2–30°C	2–30°C	2–30°C	2–30°C	2–30°C	2–30°C	2–30°C
Shelf-life of consumables									
Estimated preanalytical time	2 min	2 min	<1 min	3 min	5 min	<1	<1 min	<1 min	1 min
Time to result	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min	5 min
Retail price, consumables (USD)	<1	<1	<1	<1	<1	<1	<1	<5	<5
Practical aspects of the test ^a	-Non-homogeneous test line intensity (sometimes the middle of the line is stronger than the edges and therefore difficult to interpret)	Easily readable band width with homogeneous staining and uniform, consistent color	Low intensity test lines but mostly homogeneous	-First band intensity differs considerably from the control line (assessment difficult) -Band intensity often very non-homogeneous	-Band width non-homogeneous - Variations in color intensity	-Bright red background appears (complicates interpretation of line intensity)	-Good band intensity - Homogeneity partly irregular	-Homogeneous bands -Intense coloration	-Homogeneous bands -Intense band staining

136 n/a: not applicable

137 WB: whole blood

138 ^a Practical aspects were defined based on subjective observations of test performers using the form shown in Fig S1 to systematically collect

139 feedback for all tests.

140 ^b LIS: Laboratory information system141 ^c HIS: Hospital information system

142

143 **Sample composition**

144 Serum samples used for this study were selected from the biobank curated by FIND, the
145 global alliance for diagnostics [28] [29]. All patients from whom these samples were
146 obtained gave consent for their samples to be stored in a biobank and used for research
147 into diagnostic tools for fever management. Patients included in this study had fever at
148 presentation and were aged from 2 to 65 years [28]. Standardized guidance for sample
149 transport and storage prior to laboratory evaluation was followed [28], and all samples
150 were preserved under temperature-controlled conditions at -20°C until CRP testing.
151 Reference testing for CRP was conducted as part of the original sample characterization
152 using Cobas 8000 Modular analyzer (Roche Diagnostics International AG, Rotkreuz,
153 Switzerland) [28].

154 To evaluate the samples' stability over time and confirm that the previously generated
155 CRP values were reliable, a comparison study was performed (n = 33). Samples were
156 thawed (timepoint 2), and CRP was measured using a highly sensitive ELISA (C-reactive
157 protein high sensitive ELISA, IBL International, Hamburg, Germany, reference method
158 RM2). The resulting measurements were then compared with the original CRP values
159 (timepoint 1) obtained using Cobas 8000 Modular analyzer (Roche Diagnostics
160 International AG, Rotkreuz, Switzerland, reference method RM1). As different methods
161 were used at the two timepoints, the analysis acts as a sample stability and method
162 comparison of RM1 and RM2. A Bland–Altman (BA) plot was used to test for equivalence
163 (two one-sided t-test, TOST), applying predefined +/-8% limits for bias at a +/-90%

164 confidence interval (CI). If necessary, Clinical and Laboratory Standards Institute (CLSI)
165 guidelines EP9 allow separate concentration ranges to be analyzed.

166 **Sample size**

167 For the quantitative comparison, the number of samples (n = 40) was chosen according
168 to the verification protocol described in CLSI guideline EP09-A3 [30]. For the semi-
169 quantitative tests, a sample size of n = 100 (n = 25 per category, see below) was chosen,
170 to achieve sufficient widths of confidence intervals for the agreement measures used in
171 the study.

172 **Sample selection and study design**

173 For the evaluation of quantitative tests, samples were selected based on the distribution
174 range of CRP values of 961 samples collected for a previous study [28]. The distribution
175 of the samples had the following deciles (10% .. 90%, in mg/L): 4.4, 10, 20, 30, 45, 65,
176 85, 113, and 164. As the CLSI guideline requires 40 samples for the quantitative
177 comparison analysis, four samples were randomly selected from each of the following ten
178 ranges (mg/L): >0–4.4, >4.4–10, >10–20, >20–30 >30–45, >45–65, >65–85, >85–113,
179 >113–164, and >164. The samples were organized into four sets of 40 samples per set
180 and the same set was used with two different methods (e.g., set 1 for methods 1 and 2,
181 set 2 for methods 3 and 4, etc.). Samples were measured in duplicate.

182 For semi-quantitative tests, samples were selected to be equally distributed over all test
183 categories. For example, for a test with four categories (0–10–40–80 mg/L) the following
184 distribution of samples was chosen: 25 samples in the range 0–10 mg/L, 25 samples in
185 the range 10–40 mg/L, 25 samples in the range 40–80 mg/L, and 25 samples in the range

186 >80 mg/L. The samples were organized into five sets of 100 samples per set, and the
187 same set was used for two different methods (e.g., set 1 for methods 1 and 2, set 2 for
188 methods 3 and 4, etc.). If part of the measurement process involved visual inspection by
189 a reader, each sample was measured three times, and the respective cassettes were
190 read by two readers (blinded; six reads per sample overall). At the end, 660 samples were
191 selected for this study out of the 961 available.

192 **CRP testing procedures**

193 All tests were processed according to the manufacturer's instructions. Briefly, all
194 components were brought to room temperature, and sample dilution was performed as
195 defined by the manufacturer, using the corresponding buffer provided with the kit. After
196 thorough mixing, the diluted sample was transferred, using the appropriate applicator, to
197 the application field of either a cassette or a test strip. In some cases, the test strip was
198 dipped into the diluted sample (e.g., for a dipstick test). Following the defined incubation
199 time specified by the manufacturer, semi-quantitative tests were independently read and
200 evaluated by two readers, directly after each other, who were blinded to each other's
201 result. For the assessment of test-line intensity, a grid, ranging from 0 to 10, was used. For
202 the quantitative POC tests, the test device (cassette, strip, etc.) was read in an analyzer
203 supplied with the respective tests, with the CRP concentration directly displayed by the
204 analyzer.

205 **Statistical analysis**

206 Statistical analyses were performed using SAS software (version 9.4). If the equivalence
207 of the index method with the reference method is considered, i.e., if it is examined whether

208 the bias is near to zero, the estimates (slope of Passing–Bablok (PB) regression or mean
209 bias resulting from a BA plot) are presented with their 90% CI, allowing statements similar
210 to the TOST at an alpha level of 0.05, whereby acceptance criteria of +/-10% were
211 applied, otherwise, 95% CI are presented for estimates.

212 In terms of kappa values, we present both simple and weighted kappa values. Weighted
213 kappa, on the one hand, accounts for similarity between neighboring categories, thus,
214 allowing for a more comprehensive assessment of agreement in terms of actual
215 concentration values, compared to the binary “black & white” viewpoint of simple kappa
216 evaluation, which on the other hand may be more relevant with respect to the clinical
217 decision making. Moreover, the interpretation of kappa values is often based on the
218 proposals of Landis and Koch [31] and Altman [32], where a kappa of >0.8 is considered
219 very good/almost perfect, >0.6 is good/substantial, and >0.4 is moderate. Applying these
220 criteria, the use of the weighted kappa alone would lead to an overly optimistic
221 interpretation.

222 **Quantitative tests**

223 Method comparison

224 To compare the index test methods with the reference methods, PB regression [33] and
225 BA plots [34] (relative differences) were applied to all measured values, whereby the
226 means of duplicates measurements have been investigated. Visual inspection was
227 performed to exclude obtrusive outliers [30]. Values outside the limits of the analytical
228 ranges (ARs as reported in table 1) were not included in the calculations but are shown
229 in the respective figures. For PB regression, the slope and intercept were estimated

230 together with the respective 90% CI. A CUSUM (cumulative sum control chart) test was
231 also applied to detect any deviations from linearity.

232 The BA plots were applied to ranges with a homogeneous distribution of differences. The
233 bias, along with its 90% CI, was estimated using the means of duplicate measurements.
234 The limits of agreement (LoA) were estimated as (95%,95%)-tolerance intervals of the
235 observed differences [35], using the variance estimate obtained from single
236 measurements differences to reflect the precision of a real world measurement in
237 assessment of accuracy as presented by LoA.

238 Precision quantitative tests

239 For each duplicate, except those out of AR, the precision profile of the percent coefficient
240 of variation (%CV) was examined visually. Repeatability was estimated using a random
241 effects ANOVA, the pooled %CV and its 95% CI were calculated for each test.

242 **Semi-quantitative tests**

243 Method comparison

244 The reference method measurement results were categorized in the same way as the
245 respective index tests. The percentage agreement of the category within the specific
246 ranges of RM was calculated.

247 The descriptive statistics (tabulation of raw data, scatterplots, and agreement plots) [36],
248 as well as the analysis (kappa, linear weighted kappa, and percentage agreement for a
249 category related to the respective range of reference values), were performed on the 600
250 single reads.

251 Reliability of semi-quantitative tests

252 The agreement of results from the two readers were investigated by simple and linear
253 weighted kappa and percentage concordant and percentage discordant values. In total,
254 300 paired measurements were evaluated for each index test.

255 **Quantitative and semi-quantitative tests**

256 Binary test results

257 To compare all tests at a cut-off of 10 mg/L, available for both quantitative and semi-
258 quantitative tests, the binary test results were assessed according to the CLSI guideline
259 EP12 [37], so positive and negative percent agreement (PPA, NPA) were estimated
260 together with their Clopper–Pearson 95% CI.

261 Practical evaluation

262 In addition to their technical performance, the practical aspects of the various POC CRP
263 tests play a role in the reliability of the CRP result. Therefore, to assess the usability of
264 the different tests, we conducted a practical evaluation in the laboratory. The following
265 aspects were considered: Required sample volume, estimated preanalytical time, and
266 duration of the analysis. Moreover, band width and homogeneity were evaluated for the
267 semi-quantitative tests, as was band intensity, using a reference grade provided by FIND
268 (from 0 to 10). The general usability of quantitative test analyzers was also evaluated,
269 based on the subjective interpretation of the two users. Overall, for each semi-quantitative
270 test, 600 observations were performed, while for each quantitative test, 80 observations
271 were evaluated. The form used to record the results of the practical evaluation is shown
272 in the supporting information (S1 Fig).

273 **Results**

274 Our search identified 56 tests, which were then assessed according to the predefined
275 criteria (Fig 1). Following this assessment, 17 tests (8 quantitative and 9 semi-
276 quantitative) from 13 companies were included in the study (Table 1).

277 **Equivalence of reference methods**

278 We were able to confirm equivalence (using TOSTs, with acceptance criteria +/-8%) for
279 the results measured using RM1 (Cobas 8000 Modular analyzer Roche Diagnostic) and
280 RM2 (C-reactive protein high sensitive ELISA, IBL International) for the CRP values in
281 the range >10 mg/L (bias=2.1% (90%CI: -0.0% .. 4.3%, 1 outlier removed). The
282 corresponding BA plot is shown in the supporting information (S2 Fig). Low
283 concentrations of CRP, in the range <10 mg/L, were evaluated separately, according to
284 CLSI EP9 guidelines; for this range, a media bias of -0.48 mg/L (90% CI) must be taken
285 into account, which is acceptable for this study.

286 **Quantitative tests**

287 **Method comparison**

288 Both method comparison analysis (PB regression and BA plot) detected considerable
289 (and significant) proportional biases for the majority of the compared tests (PB: slopes
290 from 0.822 to 1.571, BA: bias from -28.0% to 31.7%) (Table 2). For three devices
291 (QuikRead go, Spinit, and INCLIX), an acceptable agreement with the reference method
292 was seen. For the PB regression analysis (Fig 2), the slope was in the 0.91–1.1 range
293 (meaning +/-10%) for QuikRead go and Spinit. In the BA plot analysis (Fig 3), the mean
294 bias was -1.2 (90% CI: -4.7 to 2.3%) and -7.5% (90% CI: 9.3–5.7%) for INCLIX and

295 QuikRead go, respectively. The regression analysis (Fig 2) showed that an
 296 overestimation of CRP values occurred with STANDARD F 200, EPITHOD 616, and IFP-
 297 3000 when compared with the reference method.

298 Overall, QuikRead go and Spinit (and, to a lesser extent, INCLIX and GS 1200) showed
 299 best agreement with the reference method (Fig 2).

300
 301 **Table 2. Results for Passing–Bablok regression (slope, intercept) and Bland–Altman plot analysis**
 302 **(percentage bias, limits of agreement (LoA)).**

Index test	Number of samples	PB regression		Number of samples	BA plot analysis	
		Slope (two-sided 95%CI)	Intercept (two-sided 95%CI)		Mean percentage difference (two-sided 90%CI)	LoA (k-factor approach)
QuikRead go*	34	0.963 (0.930, 0.988)	-0.483 (-0.998, -0.039)	34	-7.2% (-9.0%, -5.3%)	-24.8% to 10.5%
Spinit*	38	0.921 (0.875, 0.936)	-0.543 (-0.807, 0.369)	38	-11.1% (-12.5%, -9.7%)	-25.7% to 3.5%
INCLIX	37	0.896 (0.809, 0.975)	3.042 (0.214, 6.635)	37	-1.2% (-4.7%, 2.3%)	-41.0% to 38.6%
GS1200	38	0.843 (0.772, 0.904)	-1.246 (-3.576, -0.516)	38	-26.7% (-31.7%, -21.6%)	-74.3% to 20.9%
LS 4000	39	0.822 (0.735, 0.939)	0.495 (-1.401, -1.329)	39	-12.5% (-16.7%, -8.2%)	-50.3% to 25.4%
IFP-3000	35	1.345 (1.277, 1.411)	-1.251 (-3.947, 1.409)	31	23.4% (17.7%, 29.0%)	-28.8% to 75.5%
EPITHOD 616	36	1.390 (1.390, 1.303)	-0.937 (-3.679, 0.765)	36	31.3% (25.0%, 37.5%)	-29.1% to 91.6%
Standard F 200	22	1.571 (1.379, 1.948)	-4.278 (-10.358, -2.468)	22	28.0% (14.3%, 41.6%)	-79.2% to 135%

304 Tests were ranked based on the best slope (i.e., the nearest to 1).

305 The number of samples, each measured in duplicate, refers to the number included in the specific analysis

306 out of 40 samples. Where this number differs from 40, it means that some samples were excluded from the

307 analysis because the values were (for PB and BA) outside the AR or (for BA) violated the assumption of
308 homogeneity of variances (mean <3 mg/L, IFP 3000).

309 *SpinIt and QuikRead go showed the best slope and intercept, near to 1 and 0, respectively.

310 NB: One outlier found with both replicates was excluded for LS4000. For EPITHOD 616 and GS1200,
311 strong deviating replicates were found for one sample in both cases, so a third measurement was then
312 performed and the pair with similar values was chosen for analysis.

313

314 **Fig 2. Passing–Bablok regression scatterplots for quantitative tests.** Passing–Bablok regression
315 analysis to compare the various quantitative tests and CRP concentrations determined using the RM. Plots
316 are sorted by agreement of the slope with 1, from the upper left to the bottom right. The regression line is
317 represented by a solid black line; dashed gray lines indicate the line of identity. Black dots represent
318 samples included in the analysis, while the other symbols represent samples excluded from the regression
319 analysis measurements (values out of AR and outliers). Gray areas represent the 90% confidence bands,
320 whereby the range covers the observed concentration within AR of the respective index method.

321

322 **Fig 3. Bland–Altman plots for quantitative tests.** Bland–Altman plots comparing CRP POC test results
323 measured using the various quantitative index tests and CRP results measured using the reference method
324 (RM). The X-axes depict the CRP values of the RM, and the Y-axes depict the relative difference between
325 CRP results measured by the POC test under study and the RM. The thick black lines represent the bias
326 and the dotted line its 90% CI; the LoA are represented by the vertical expansion of the gray areas, their
327 horizontal range covers the observed concentrations within AR of the respective index method*. Black dots
328 represent samples included in the analysis, while the other symbols represent excluded data points (values
329 out of AR and outliers). Note: For Standard F200 and IFP-3000, the assumption of concentration-
330 independent relative bias, which is essential for the validity of the BA plot analysis, is questionable.

331 *For IPF3000, three samples were excluded in the low concentration range (<3 mg/L) to achieve constant
332 CV (variance homogeneity) in the analyzed range. As a consequence, the lower limit of the AR is
333 approximately 4 mg/L and not 0.5 mg/L (denoted by the light-blue area in the figure).

334 Precision

335 The repeatability of the data was analyzed to evaluate the agreement between measured
336 values obtained by replicate measurements (Table 3). The values, expressed as mean
337 % CV, ranged from about 5% up to 16.9%. The tests with the best repeatability were
338 Spinit, LS 4000, and QuikRead go (range approximately 5%), while the largest variability
339 was observed with the SD Biosensor test, with a CV of >15%.

340 **Table 3. Repeatability of quantitative tests.**

Repeatability (%CV), 95% CI	
Spinit	4.7% [3.8% to 6.0%]
LS 4000	4.8% [3.9% to 6.1%]
QuikRead go	5.3% [4.3% to 6.9%]
IFP-3000	9.5% [7.7% to 12.4%]
EPITHOD 616	13.7% [11.1% to 17.7%]
GS1200	13.8% [11.3% to 17.8%]
INCLIX	15.4% [12.5% to 19.9%]
STANDARD F 200	16.9% [13.1% to 23.9%]

341 Tests are ranked based on their repeatability, expressed as % CV, 95% CI.

342 Semi-quantitative tests

343 Method comparison

344 The agreement of the semi-quantitative strips with the RM is shown in Table 4. The
345 percentage agreements ranged from high values (100%) to as low as 5.1%, depending
346 on the test and the range. Based on the kappa values, the tests that showed the best
347 agreement were BTNX Quad Line, Biotest, NADAL cassette, and BTNX Tri Line (simple
348 kappa values >0.6). If we also consider the different percentages of agreement for the

349 single categories, the NADAL CRP test and BTNX Quad Line were the two tests that
 350 gave the best performance, showing simple kappa values of >0.6 (good agreement) and
 351 being the only two tests where we observed a percentage of agreement >50% for all four
 352 categories. The Bangdiwala plots (Fig 4), used for analysis visualization, also showed a
 353 greater agreement among all the categories for the NADAL cassette and BTNX Quad
 354 Line tests.

355 Most of the quantitative tests with four categories showed a lower percentage of
 356 agreement for the intermediate categories (10–40 and 40–80 mg/L), while a higher
 357 percentage of agreement was observed for the lower and upper categories (0–10 and
 358 >80 mg/L) (Table 4). For example, for the cut-off <10 mg/L, we observed among all
 359 analyzed tests a percentage of agreement that ranged from 54% to 99%, while for the
 360 cut-offs >40 and <80 mg/L, the percentage of agreement was overall quite low, ranging
 361 from 11% to 56%. Biotest was the best performing test for the ≤10 mg/L category (99.3%
 362 agreement), NADAL cassette for the >10 and <40 mg/L category (83.3% agreement),
 363 BTNX Quad Line for the >40 and <80 mg/L category (56% of agreement), and ALLTEST
 364 Dipstick for the ≥80 mg/L category, showing 100% agreement.

365 **Table 4. Results of the method comparison for semi-quantitative methods.**

Test	Kappa		Percentage agreement of the category within the specific range of the reference method			
	Simple [95%CI]	Weighted [95%CI]	Negative (<10 mg/L)	≥10 to <40 mg/L	≥40 to <80 mg/L	≥80 mg/L
BTNX Quad Line*	0.676 [0.630, 0.721]	0.800 [0.769, 0.830]	76.0% (114/150) [68.4%, 82.6%]	74.7% (112/150) [66.9%, 81.4%]	56.0% (84/150) [47.7%, 64.1%]	96.0% (144/150) [91.5%, 98.5%]
NADAL cassette*	0.651 [0.604, 0.698]	0.789 [0.759, 0.819]	88.0% (132/150) [81.7%, 92.7%]	83.3% (125/150) [76.4%, 88.9%]	53.3% (80/150) [45.0%, 61.5%]	70.7% (106/150) [62.7%, 77.8%]
Biotest	0.649 [0.603, 0.694]	0.798 [0.770, 0.827]	99.3% (149/150) [96.3%, 100%]	45.3% (68/150) [37.2%, 53.7%]	55.3% (83/150) [47.0%, 63.4%]	94.7% (142/150) [89.8%, 97.7%]
ACTIM	0.476 [0.428, 0.523]	0.690 [0.658, 0.723]	54.7% (82/150) [46.3%, 62.8%]	72.0% (108/150) [64.1%, 79.0%]	16.0% (24/150) [10.5%, 22.9%]	100% (150/150) [97.6%, 100%]
ALLTEST dipstick	0.443 [0.395, 0.491]	0.671 [0.636, 0.707]	86.7% (130/150) [80.2%, 91.7%]	37.2% (58/156) [29.6%, 45.3%]	7.0% (10/142) [3.4%, 12.6%]	100% (150/150) [97.6%, 100%]

NADAL dipstick	0.422 [0.373, 0.472]	0.646 [0.610, 0.681]	98.7% (148/150) [95.3%, 99.8%]	24.0% (36/150) [17.4%, 31.6%]	55.3% (83/150) [47.0%, 63.4%]	48.7% (73/150) [40.4%, 57.0%]
ALL TEST cassette	0.413 [0.365, 0.462]	0.636 [0.598, 0.674]	76.7% (115/150) [69.1%, 83.2%]	37.2% (58/156) [29.6%, 45.3%]	11.1% (16/144) [6.5%, 17.4%]	98.0% (147/150) [94.3%, 99.6%]
			Negative (<10 mg/L)	≥10 to <30 mg/L	≥30 to <80 mg/L	≥80 mg/L
BTNX Tri Line	0.645 [0.597, 0.693]	0.750 [0.714, 0.786]	81.9% (149/182) [75.5%, 87.2%]	86.6% (123/142) [79.9%, 91.7%]	80.7% (121/150) [73.4%, 86.7%]	27.8% (25/90) [18.9%, 38.2%]
			Negative (<10 mg/L)	≥10 to <27.7 mg/L	≥27.7 to <32.5 mg/L	≥32.5 mg/L
ALLTEST cassette	0.289 [0.248, 0.329]	0.420 [0.380, 0.460]	54.0% (107/198) [46.8%, 61.1%]	87.8% (158/180) [82.1%, 92.2%]	11.1% (2/18) [1.4%, 34.7%]	5.1% (10/198) [2.4%, 9.1%]

366

367 Kappa (simple, linear weighted) and percentage agreement within specific ranges of the reference method,
 368 according to the cut-off of the investigated method. Within equal cut-off groups, tests are ranked based on
 369 the simple kappa values.

370 *Semi-quantitative tests (BTNX CRP quad line and NADAL cassette) showed a good agreement (simple
 371 kappa >0.6) and percentage of agreement for each individual category of >50%.

372

373 **Fig 4. Agreement plots (Bangdiwala plots) for semi-quantitative tests.** Y-axis: categories using
 374 measurement values of the reference method, X-axis: categories according to the different index tests (units
 375 are mg/L). The horizontal width refers to the number of measurement results in the category. The gray
 376 shading refers to varying degrees of agreement (darker gray = a better degree of agreement). Overall, a
 377 good agreement is indicated by a square shape, with corners on the diagonal identity line and a large
 378 amount of dark gray filling the area.

379

380 Reliability of semi-quantitative tests

381 The agreement of results provided by the two independent readers is shown in Table 5,
 382 overall we could observe a high percentage of concordant results for all tests (from 76.8%
 383 to 88.3%). More precisely, BTNX Tri Line, ALLTEST dipstick and BTNX Quad Line tests
 384 showed the best proportion of concordant readings, above 85%, and a simple kappa >0.8.

385

386

387

388 **Table 5. Between-reader agreement for the semi-quantitative tests.**

Index test	Kappa (simple) [95%CI]	Kappa (weighted*) [95%CI]	Concordant (n/N) [95%CI]	Discordant (n/N) [95%CI]
BTNX Tri Line	0.831 [0.778, 0.884]	0.880 [0.843, 0.917]	88.3% (249/282) [84.0%, 91.8%]	11.7% (33/282) [8.2%, 16.0%]
ALLTEST Dipstick	0.818 [0.764, 0.873]	0.907 [0.877, 0.937]	88.0% (263/299) [83.7%, 91.4%]	12.0% (36/299) [8.6%, 16.3%]
BTNX Quad Line	0.811 [0.758, 0.863]	0.889 [0.856, 0.921]	86.0% (258/300) [81.6%, 89.7%]	14.0% (42/300) [10.3%, 18.4%]
BIOTEST	0.752 [0.693, 0.810]	0.868 [0.835, 0.901]	82.0% (246/300) [77.2%, 86.2%]	18.0% (54/300) [13.8%, 22.8%]
ALLTEST Cassette (four categories)	0.734 [0.673, 0.796]	0.861 [0.825, 0.896]	82.3% (247/300) [77.5%, 86.5%]	17.7% (53/300) [13.5%, 22. 5%]
NADAL Cassette	0.729 [0.669, 0.789]	0.835 [0.796, 0.873]	79.7% (239/300) [74.7%, 84.1%]	20.3% (61/300) [15.9%, 25.3%]
ACTIM CRP	0.725 [0.663, 0.787]	0.848 [0.811, 0.885]	81.7% (245/300) [76.8%, 85.9%]	18.3% (55/300) [14.1%, 23.2%]
NADAL Dipstick	0.674 [0.609, 0.739]	0.817 [0.779, 0.856]	77.3% (232/300) [72.2%, 81.9%]	22.7% (68/300) [18.1%, 27.8%]
ALLTEST Cassette (three categories)	0.599 [0.516, 0.682]	0.650 [0.575, 0.724]	76.8% (228/297) [71.5%, 81.5%]	23.2% (69/297) [18.5%, 28.5%]

389

390 Simple kappa, linear weighted kappa, and the percentages of concordant and discordant results are shown,
 391 including the 95% CI. The tests are ranked according to the simple kappa results. For each test, 300
 392 measurements (100 samples in triplicate) were evaluated, except for the ALLTEST dipstick (1 of 300 was
 393 not evaluable) and the ALLTEST cassette (3 of 300 were not evaluable).

394 **Quantitative and semi-quantitative tests**

395 **Method comparison of binary test results**

396 To allow comparison across all the tests and hence aid the selection for different use
 397 cases, one cut-off value was chosen (10 mg/L) that could be found across all of the
 398 selected tests (quantitative and semi-quantitative) (Fig 5). Compared with semi-
 399 quantitative methods, the PPA (positive percent of agreement) and NPA (negative

400 percent of agreement) for the quantitative tests were more accurate at the selected cut-
401 off; in fact, all PPAs were more than 90% and all NPAs were 100%, with the exception of
402 the EPITHOD 616 test (NPA = 75%). For the semi-quantitative tests, the PPAs were
403 similarly high to those seen with the quantitative tests (except 76.4% for the NADAL
404 Dipstick test), but lower NPAs were observed (two tests, 50%–60%; two tests, 60%–80%;
405 and three tests, 80%–90%), while only two tests showed an NPA >90% (NADAL Dipstick
406 and BIOTEST). This means that for a use case requiring a cut-off of 10 mg/L, a
407 quantitative test is recommended.

408
409 **Fig 5. PPA and NPA for semi-quantitative and quantitative tests (cut-off: 10 mg/L).** For a cut-off of 10
410 mg/L, which was available for all quantitative and semi-quantitative methods, the binary test results were
411 assessed; thus, the positive and negative percent agreement were estimated, together with their 95%-
412 Pearson Clopper CI.

413

414 **Discussion**

415 Here, we report the results of a method comparison analysis, in which eight quantitative
416 and nine semi-quantitative tests used to measure CRP levels were evaluated for their
417 analytical performance, by comparing them with a known reference method following
418 standard guidelines [38]. To the best of our knowledge, this is the first study where this
419 number of POC tests for CRP have been compared and the results made publicly
420 available to assist developers, users, and procurers. The experiments were designed to
421 cover the expected clinical range of CRP values, with a particular view toward the use

422 case of guiding antibiotic prescribing and other triage decisions at the POC [28] [20] for a
423 broad range of CRP values, in line with the distribution spectrum of CRP values measured
424 in clinical samples. Overall, the quantitative tests showed a satisfactory performance, with
425 the QuikRead go and Spinit tests (and, to a lesser extent, the INCLIX and GS 1200 tests)
426 displaying better agreement with the reference method than the other quantitative tests.
427 For most of the semi-quantitative tests, the percentage of agreement with the reference
428 method varied according to the CRP-level category being examined. Notably, tests with
429 three test bands (the equivalent of four categories) showed a lower percentage of
430 agreement for the 10–40 and 40–80 mg/L categories (from 7% to 83%), while a higher
431 percentage of agreement was observed for the CRP categories <10 (from 54% to 99%)
432 and >80 mg/L (from 48% to 100%). For BTNX Tri Line, on the contrary, a high percentage
433 agreement was observed for the intermediate categories 10–30 and 30–80 mg/L (86 %
434 and 80% respectively), while the category > 80 mg/L showed only 27% of agreement.
435 The current results highlight that, depending on the relevant CRP cut-off needed for the
436 clinical use-case, the utility of the different tests may vary.

437 The (binary) diagnostic accuracy for CRP-measurements at a cut-off 10 mg/L was also
438 explored, as this is a common cut-off used, across tests and use cases, like tuberculosis
439 and pneumonia [39][21]. While differences existed, overall, the findings suggested that
440 all of the quantitative tests could be used for the cut-off value of 10 mg/L and, more in
441 general, for a broad clinical relevant range, while none of the semi-quantitative tests
442 performed well, and a more cut-off/range-specific selection will be needed. The latter is a
443 critical finding, considering the current push for decentralized testing to inform antibiotic
444 use, TB triage, or other clinical interventions at the first point of contact in resource-limited

445 settings [21][40]. While simple lateral flow-type semi-quantitative tests are likely
446 considered easier to perform, cheaper and closer to existing target product profiles [16],
447 our finding highlights that quantitative tests might have a broader utility across a wider
448 range of CRP categories and hence use cases.

449 Our findings align with those of a previous study conducted by Minnaard and colleagues,
450 in which the analytical performance of QuikRead go was evaluated [25]. Although we
451 cannot directly compare our results, as in their BA analysis they calculated the mean
452 differences for three ranges of CRP (<20, 20–100, and >100 mg/L), their study showed a
453 good agreement with the reference method for QuikRead go. In contrast, an evaluation
454 conducted by Brouwer and colleagues [26], also using QuikRead go, revealed a
455 significant underestimation of the CRP value compared with the reference method (slope
456 = 0.85), while for our study the lower values provided by QuikRead go were deemed
457 acceptable (slope =0.96). In accordance with our results, their evaluation of the
458 performance of the semi-quantitative test ACTIM also showed an insufficient correlation
459 with the reference method. Overestimation as well as underestimation of CRP values has
460 been observed [26].

461 Although we aimed to follow the standardized procedures as outlined by the CLSI
462 [37][30], there were some limitations to our study. Two different reference methods were
463 used (C-reactive protein high sensitive ELISA from IBL International and Cobas 8000
464 Modular analyzer from Roche Diagnostics International AG), but we could confirm
465 equivalence within the prespecified limits, hence it should not be assumed that this
466 caused any issues. Regarding the diagnostic accuracy at 10 mg/L, few samples were
467 used in the range <10 mg/L for the quantitative tests. This is because the focus of this

468 study was a method comparison analysis, and samples were selected to cover a broad
469 range of CRP values.

470 In summary, we set out to provide pragmatic data to users, procurement agencies,
471 laboratories, implementers, and ministries of health, which can help to inform access to
472 and roll-out of CRP testing tools for a variety of use cases, outside of central reference
473 laboratories. CRP appears on the Essential Diagnostics List produced by WHO and those
474 of individual countries [41][42] and is recognized to be a simple (albeit imperfect) tool to
475 complement clinical assessments to guide antibiotic treatment [43] or conduct triage for
476 infectious diseases [21]. Therefore, it is now of critical importance to continuously assess
477 relevant tools and identify diagnostic tests that can ensure quality data are being
478 generated and used for patient care at all levels of the health system.

479 **Acknowledgments**

480 The authors wish to thank the participants who provided the samples and Adam Bodley
481 for medical writing assistance and the editorial support.

482 **References**

- 483 1. Black S, Kushner I, Samols D. C-reactive Protein. *J Biol Chem.*
484 2004;279(47):48487–90.
- 485 2. Andreeva E, Melbye H. Usefulness of C-reactive protein testing in acute cough /
486 respiratory tract infection : an open cluster-randomized clinical trial with C-reactive
487 protein testing in the intervention group. 2014;15(1):1–9.
- 488 3. Sulis G, Adam P, Nafade V, Gore G, Daniels B, Daftary A, Das J, Gandra S, Pai M.
489 Antibiotic prescription practices in primary care in low- And middle-income
490 countries: A systematic review and meta-analysis. *PLoS Med.* 2020 Jun 1;17(6).
- 491 4. Mcgowan DR, Sims HM, Zia K, Uheba M, Shaikh IA. The value of biochemical
492 markers in predicting a perforation in acute appendicitis 2013;83:79–83.
- 493 5. Rebnord IK, Hunskaar S, Gjesdal S, Hetlevik Ø. Point-of-care testing with CRP in

- 494 primary care : a registry-based observational study from Norway. *BMC Fam Pract.*
495 2015;1–8.
- 496 6. Eccles S, Pincus C. Diagnosis and management of community and hospital
497 acquired pneumonia in adults : summary of NICE. 2014;6722(December):1–5.
- 498 7. Käser SA, Fankhauser G, Glauser PM, Toia D, Maurer CA. Diagnostic value of
499 inflammation markers in predicting perforation in acute sigmoid diverticulitis. *World*
500 *J Surg.* 2010 Nov;34(11):2717–22.
- 501 8. Wu HP, Lin CY, Chang CF, Chang YJ, Huang CY. Predictive value of C-reactive
502 protein at different cutoff levels in acute appendicitis. *Am J Emerg Med.*
503 2005;23(4):449–53.
- 504 9. Omraninava M, Javan S, Mahmodi G. Precision of C-reactive protein in the
505 diagnosis of acute appendicitis. *Soc Determinants Health.* 2022;8(1):1–6.
- 506 10. Kapasi AJ, Dittrich S, González IJ, Rodwell TC. Host Biomarkers for Distinguishing
507 Bacterial from Non-Bacterial Causes of Acute Febrile Illness : A Comprehensive
508 Review. *PLoS One.* 2016;11(8): e0160278. doi:10.1371/journal.pone.0160278
- 509 11. Simon L, Gauvin F, Amre DK, Saint-louis P, Lacroix J. Serum Procalcitonin and C-
510 Reactive Protein Levels as Markers of Bacterial Infection : A Systematic Review
511 and Meta-analysis. *CID.* 2004;8064.
- 512 12. Peng C, Tian C, Zhang Y, Yang X, Feng Y, Fan H. C-Reactive Protein Levels
513 Predict Bacterial Exacerbation in Patients With Chronic Obstructive Pulmonary
514 Disease. *Am J Med Sci.* 2013;345:190_194.
- 515 13. Korppi M, Kröger L. C-reactive protein in viral and bacterial respiratory infection in
516 children. *Scand J Infect Dis.* 1993;25(2):207–13.
- 517 14. Verheij TJM, Prins JM, Salomé PhL, Bindels PJ, Ponsioen BP, Sachs APE.
518 NHGStandaard Acut hoesten. *Huisarts Wet.* 2011;54(68–92).
- 519 15. Chandna A, Osborn J, Bassat Q, Bell D, Burza S, D’Acremont V, Fernandez-
520 Carballo BL, Kain KC, Mayxay M, Wiens M, Dittrich S. Anticipating the future:
521 prognostic tools as a complementary strategy to improve care for patients with
522 febrile illnesses in resource-limited settings. *BMJ Glob Heal.* 2021 Jul;6(7).
- 523 16. Dittrich S, Tadesse BT, Moussy F, Chua A, Zorzet A, Tängdén T, Dolinger DL,
524 Page AL, Crump JA, D’Acremont V, Bassat Q, Lubell Y, Newton PN, Heinrich N,
525 Rodwell TJ, González IJ. Target product profile for a diagnostic assay to
526 differentiate between bacterial and non-bacterial infections and reduce
527 antimicrobial overuse in resource-limited settings: An expert consensus. *PLoS*
528 *One.* 2016;11(8):1–12.
- 529 17. AMR Collaborators. Global burden of bacterial antimicrobial resistance in 2019 : a
530 systematic analysis. *Lancet* 2022; 399: 629–55
- 531 18. Llor C, Bjerrum L. Antimicrobial resistance: Risk associated with antibiotic overuse
532 and initiatives to reduce the problem. *Ther Adv Drug Saf.* 2014;5(6):229–41.
- 533 19. Giacomini E, Perrone V, Alessandrini D, Paoli D, Nappi C, Esposti LD. Evidence of
534 antibiotic resistance from population-based studies: A narrative review. *Infect Drug*
535 *Resist.* 2021;14:849–58.
- 536 20. WHO consolidated guidelines on tuberculosis. Module 2: screening – systematic
537 screening for tuberculosis disease. Geneva: World Health Organization; 2021.
538 Licence: CC BY-NC-SA 3.0 IGO.
- 539 21. Samuels THA, Wyss R, Ongarello S, Moore DAJ, Schumacher SG, Denkinge CM.

- 540 Evaluation of the diagnostic performance of laboratory-based c-reactive protein as
541 a triage test for active pulmonary tuberculosis. *PLoS One*. 2021 Jul 1;16(7 July).
- 542 22. Braga F, Panteghini M. Derivation of performance specifications for uncertainty of
543 serum C-reactive protein measurement according to the Milan model 3 (state of the
544 art). *Clin Chem Lab Med*. 2020;58(11):E263–5.
- 545 23. The European Commission. Regulation (EU) 2017/746 of the European parliament
546 and of the council on in vitro diagnostic medical devices. *Off J Eur Union*.
547 2017;5(5):117–76.
- 548 24. Desirable Specifications for Total Error , Imprecision , and Bias , derived from Intra-
549 and Inter-Individual Biologic Variation. 2012;54(Sup 219):414–9.
- 550 25. Minnaard MC, Van De Pol AC, Broekhuizen BDL, Verheij TJM, Hopstaken RM, Van
551 Delft S, Kooijman-Buiting AMJ, De Groot JAH, De Wit NJ. Analytical performance,
552 agreement and user-friendliness of five C-reactive protein point-of-care tests.
553 *Scand J Clin Lab Invest*. 2013 Dec;73(8):627–34.
- 554 26. Brouwer N, van Pelt J. Validation and evaluation of eight commercially available
555 point of care CRP methods. *Clin Chim Acta*. 2015 Jan 5;439:195–201.
- 556 27. Lykkegaard J, Olsen JK, Sydenham RV. C-reactive protein cut-offs used for acute
557 respiratory infections in Danish general practice. 2021;1–9.
- 558 28. Escadafal C, Geis S, Siqueira AM, Agnandji ST, Shimelis T, Tadesse BT, Loembé
559 MM, Harris V, Leticia Fernandez-Carballo B, Macé A, Ongarello S, Rodriguez W,
560 Dittrich S. Bacterial versus non-bacterial infections: A methodology to support use-
561 case-driven product development of diagnostics. *BMJ Glob Heal*. 2020;5(10).
- 562 29. FIND BIOBANK. Available from: <https://www.finddx.org/biobank-services/>
- 563 30. CLSI. Measurement Procedure Comparison and Bias Estimation Using Patient
564 Samples; Approved Guideline-Third Edition. CLSI Doc EP09-A3 Wayne, PA Clin
565 Lab Stand Institute; 20 13. 2013;
- 566 31. Landis JR, Koch GG. The measurement of observer agreement for categorical
567 data. *Biometrics*. 1977 Mar;33(1):159–74.
- 568 32. Douglas GA. Practical statistics for medical research. *Stat Med*. 1991 Oct
569 1;10(10):1635–6.
- 570 33. Bablok W. Ein neues biometrisches Verfahren zur Überprüfung der Gleichheit von
571 Meßwerten von zwei analytischen Methoden: Anwendung von linearen
572 Regressionsverfahren bei Methodenvergleichsstudien in der Klinischen Chemie,
573 Teil I. *Clin Chem Lab Med*. 1983;21(11):709–20.
- 574 34. Martin Bland J, Altman D. Statistical methods for assessing agreement between
575 two methods of clinical measurement. *Lancet*. 1986 Feb 8;327(8476):307–10.
- 576 35. Barnhart HX, Haber MJ, Lin LI. An overview on assessing agreement with
577 continuous measurements. *J Biopharm Stat* . 2007;17(4):529—569.
- 578 36. Bangdiwala SI, Shankar V. The agreement chart. *BMC Med Res Methodol*. 2013
579 Jul;13:97.
- 580 37. Garrett PE. User Protocol for Evaluation of Qualitative Test Performance; Approved
581 Guideline - Second Edition. Vol. 28, CLSI/NCCLS document EP12-A2. 2008.
- 582 38. Jan S. Krouwer PD, Daniel W. Tholen MS, Carl C. Garber PD, Henk M.J.
583 Goldschmidt PD, Martin Harris Kroll MD, Kristian Linnet, M.D. PD, Kristen Meier
584 PD, Max Robinowitz MD, Kennedy JW. EP9-A2 Method Comparison and Bias
585 Estimation Using Patient Samples. Vol. 22. 2002. 54 p.

- 586 39. Şahin F, Yildiz P. Distinctive biochemical changes in pulmonary tuberculosis and
587 pneumonia. Arch Med Sci 2013; 9, 4: 656–661. doi: 10.5114/aoms.2013.34403
- 588 40. Wilson D, Badri M, Maartens G. Performance of Serum C-Reactive Protein as a
589 Screening Test for Smear-Negative Tuberculosis in an Ambulatory High HIV
590 Prevalence Population. PLoS One. 2011 Jan 10;6(1):e15248.
- 591 41. The selection and use of essential in vitro diagnostics: report of the third meeting
592 of the WHO Strategic Advisory Group of Experts on In Vitro Diagnostics, 2020
593 (including the third WHO model list of essential in vitro diagnostics). Geneva: World
594 Health Organization; 2021 (WHO Technical Report Series, No. 1031). Licence: CC
595 BY-NC-SA 3.0 IGO.
- 596 42. Indian Council of Medical Research. National Essential Diagnostics List. 2019;
- 597 43. Biomarkers for acute febrile illness at the point-of-care in low-resource settings.
598 2020. Available from: [https://www.finddx.org/wp-content/uploads/2021/05/Meeting-](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)
599 [report_Biomarkers-for-acute-febrile-illness-at-the-point-of-care-in-low-resource-](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)
600 [settings.pdf](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)

604 **Supporting information captions**

605 **S1 Fig. Practical evaluation template**

606

607 **S2 Fig. Bland–Altman plot for the stability of reference measurements.** A) excluding values <10 mg/L;

608 B) values from 0 to 10 mg/L. Gray solid line: bias, grey dotted line: 90% confidence interval band. Gray

609 dash line: allowable range. Symbol star: outlier not included in analysis.

610

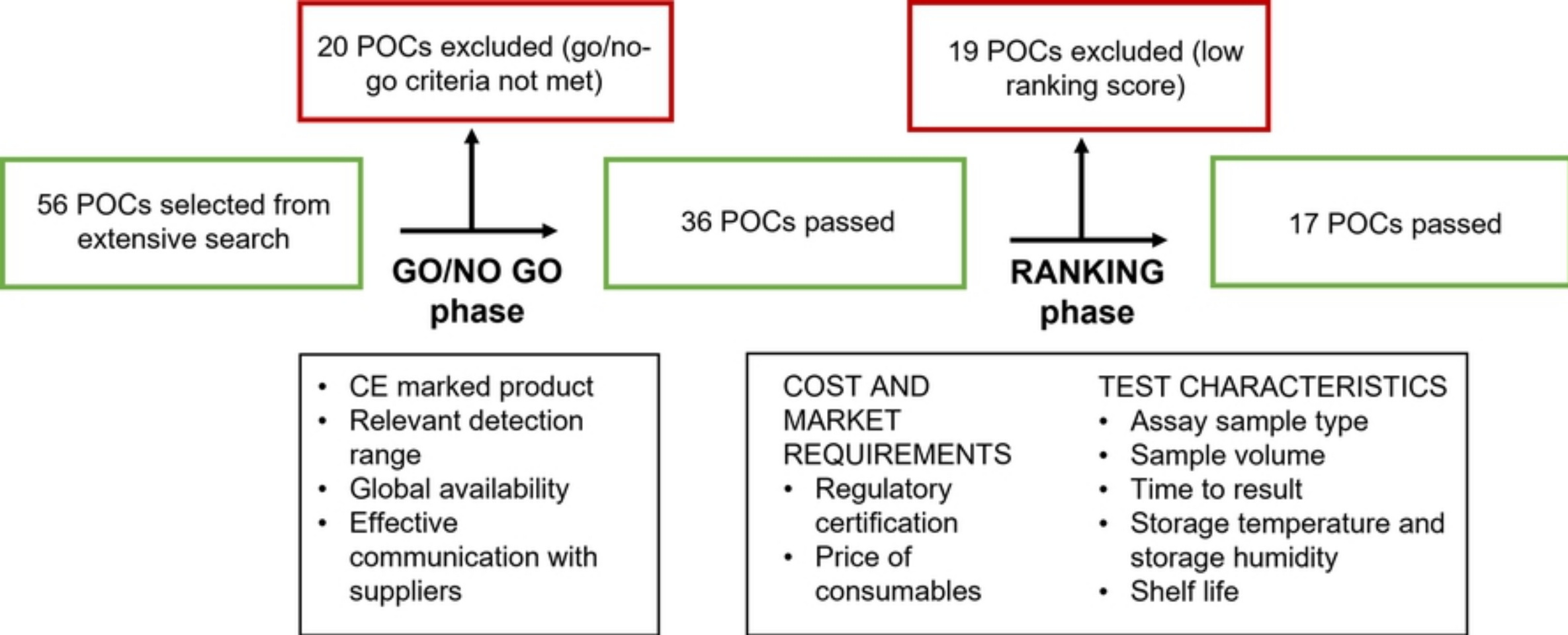


Fig 1

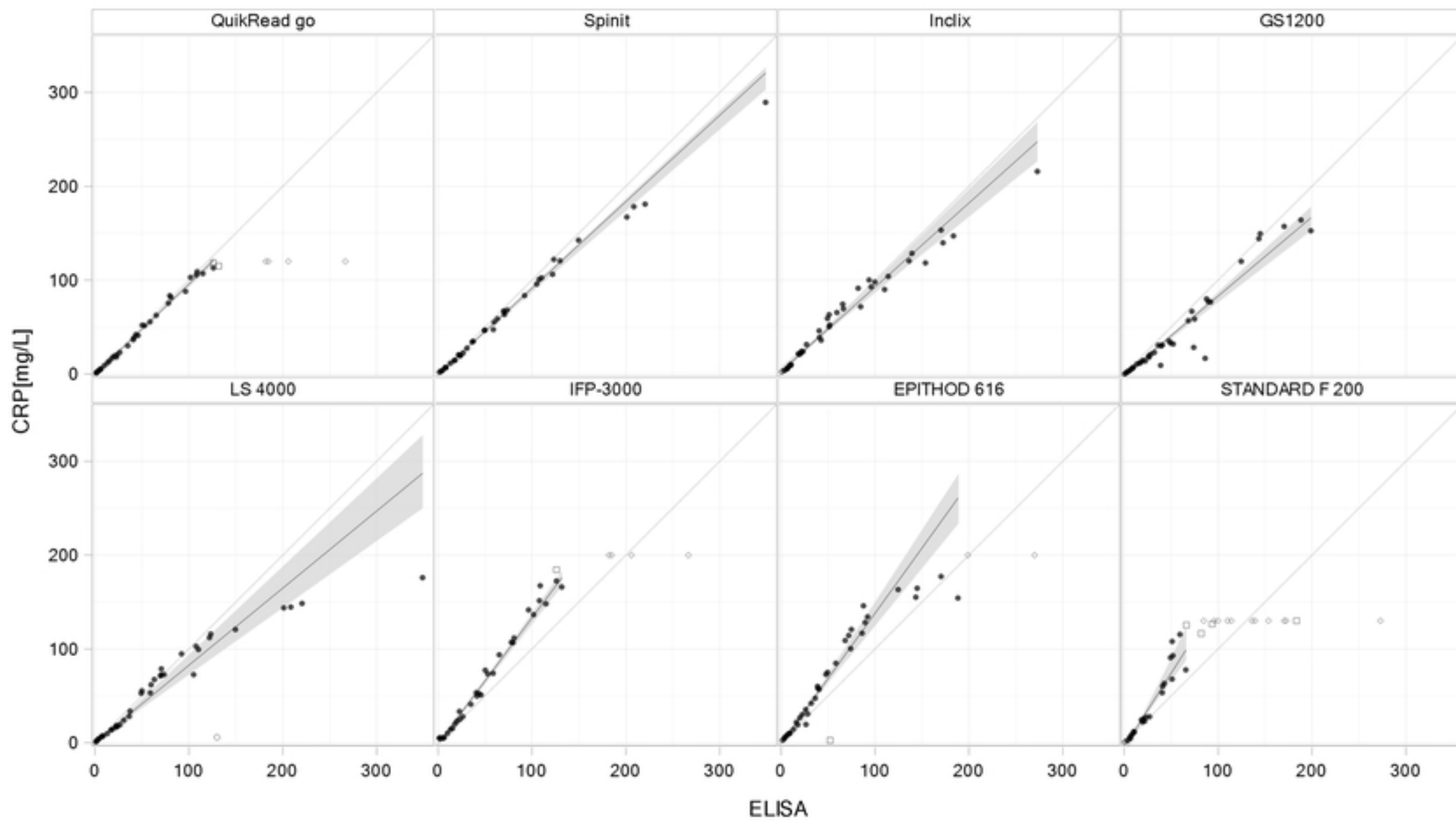


Fig 2

BIAS estimation - Difference plots

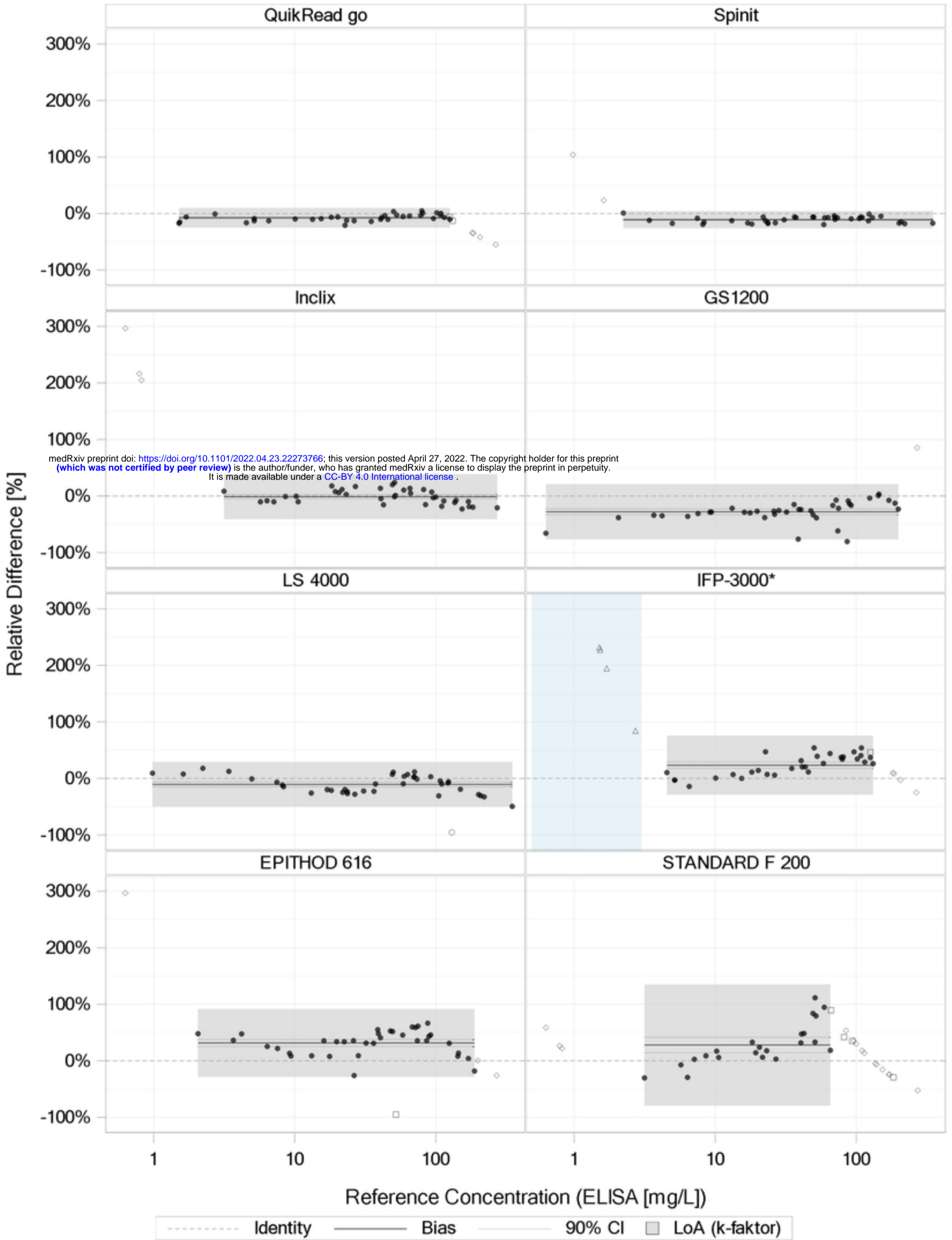


Fig 3

'Diagnostic Accuracy' for Cut-off: '10'

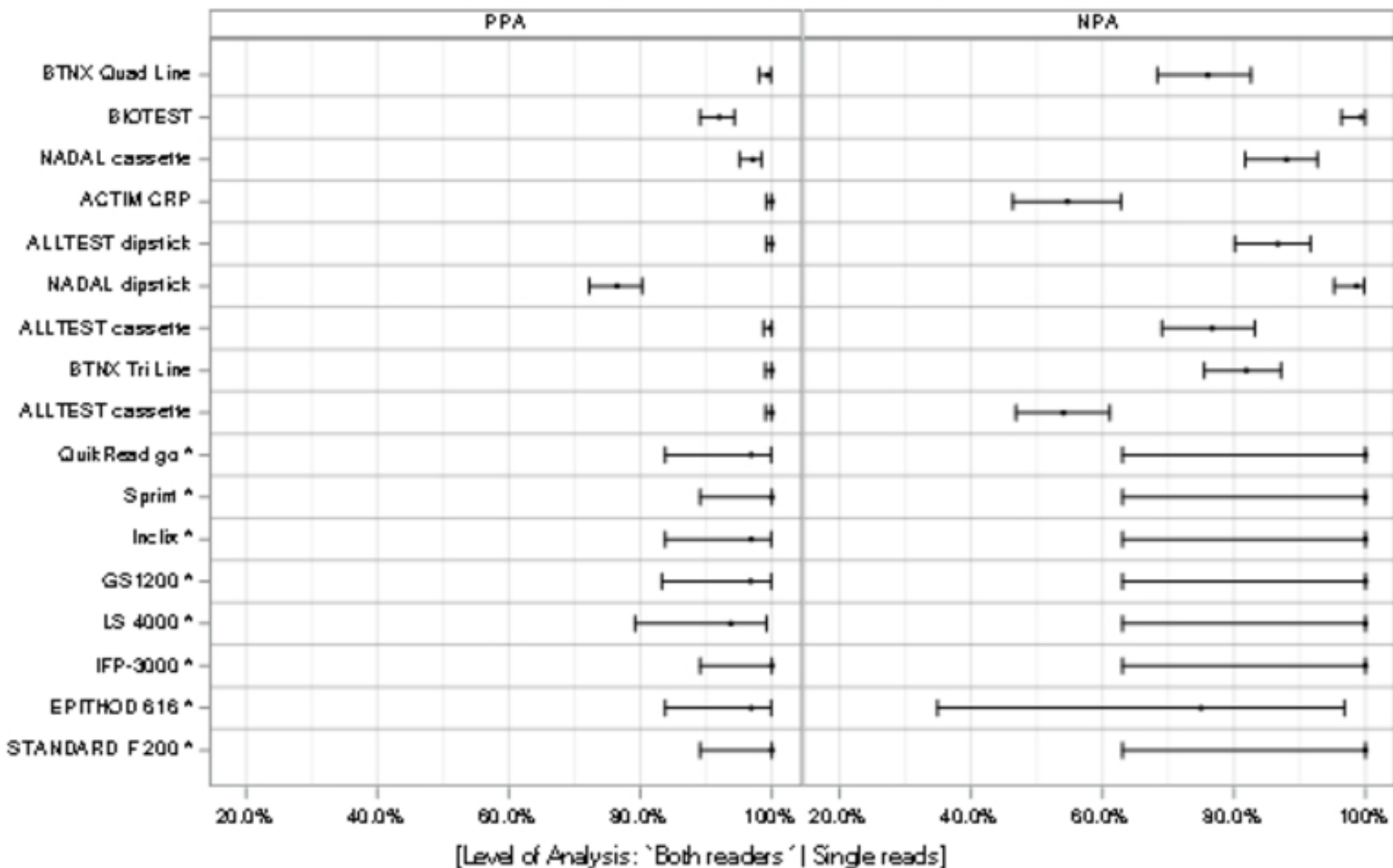


Fig 5

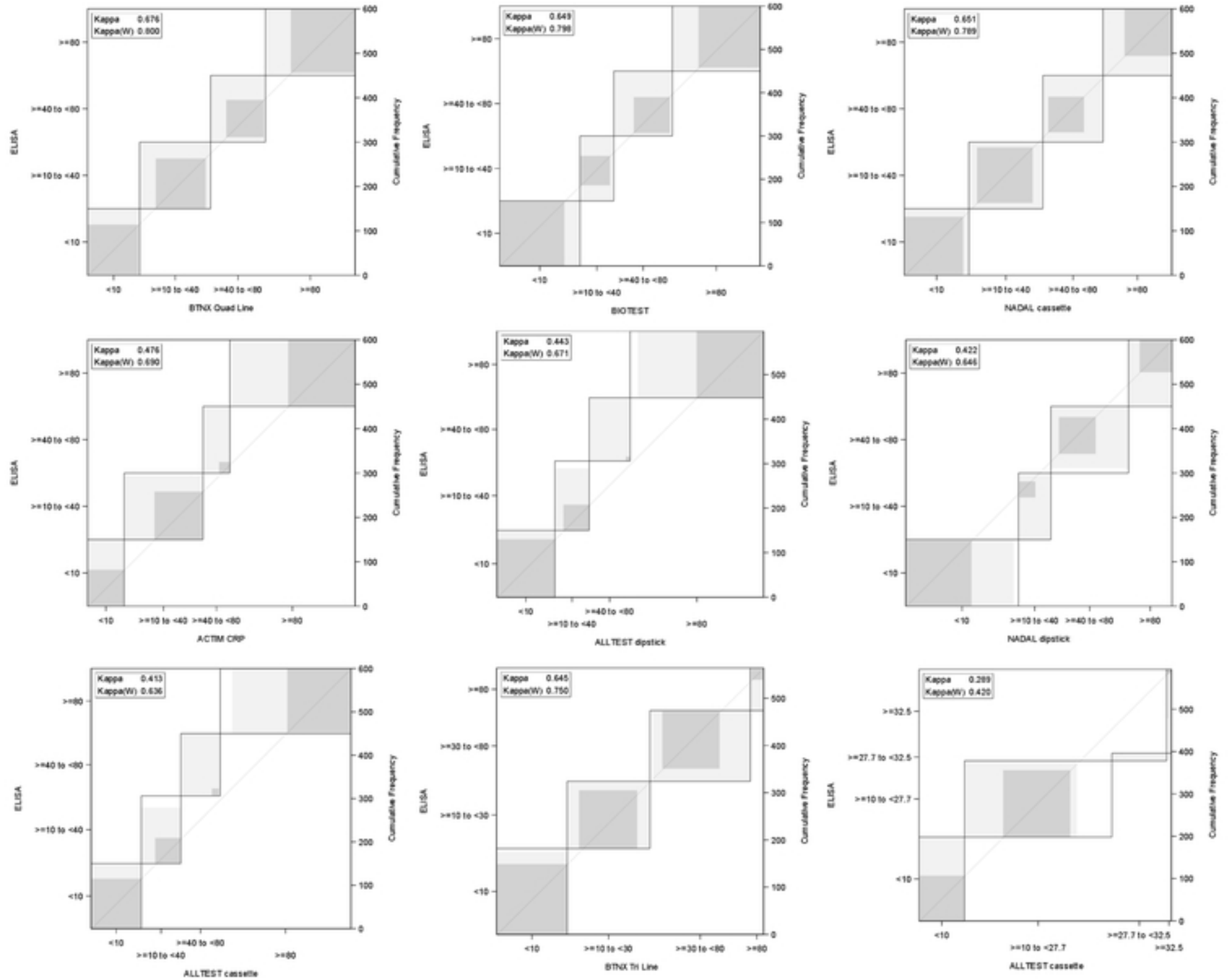


Fig 4