Analytical
Methods

rsc.li/methods

ROYAL SOCIETY
OF CHEMISTRY

PAPER
Emanuel Carrilho *et al.*
Improved assessment of accuracy and performance indicators in
paper-based ELISA

Check for updates

Improved assessment of accuracy and performance indicators in paper-based ELISA†

Thiago Mazzu-Nascimento, [ab] Giorgio Gianini Morbioli,[abc]
Luis Aparecido Milan, [d] Diego Furtado Silva,[e] Fabiana Cristina Donofrio,[f]
Carlos Alberto Mestriner[g] and Emanuel Carrilho [*ab]

Paper-based devices are an excellent match for low-cost point-of-care testing (POCT) tools. Their user-friendliness, portability, and short time of analysis, coupled with ease of local manufacture make these devices the best option for inexpensive diagnostic testing tools. However, despite all their positive features, these low-cost diagnostic devices must present good performance indicators, such as sensitivity, specificity, and accuracy. We developed and validated a paper-based ELISA for toxoplasmosis diagnosis through the detection of *Toxoplasma gondii* immunoglobulin G (IgG) antibodies in 100 human serum samples. From among the different ways to define the cut-off value, we chose Youden's *J* index (cut-off = 21.73 A.U.), which presented a higher sensitivity value. Our paper-based assay presented a sensitivity of 0.96, a specificity of 0.87, and a gray zone comprising 16 samples (±15% of the cut-off value, with 3 false positive outputs). The accuracy of the test was estimated by using ROC curves (AUC = 0.97). We also created a macro in Microsoft Excel® to estimate the accuracy of the test (m-Accuracy) based on a non-parametric method, which evidenced a value = 0.88, which classifies our test as moderately to highly accurate. We also provide the m-Accuracy macro for download and the paper-based microplate designs for printing, in order to collaborate with the scientific community and facilitate further studies using this platform. The improvement of these diagnostic tools can bring this technology for those who need it, contributing to population health and well-being.

Introduction

Enzyme-linked immunosorbent assay (ELISA) is one of the most used assays for diagnostic tests. This immunoassay makes use of antigens or antibodies previously immobilized on the surface of a support, and reports the analyte detection by using specific antibodies labeled with enzymes. This versatile technique allows for distinct experiments, such as capture, sandwich or competitive assays.[1]

[a]*Instituto de Química de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense, 400, 13566-590 São Carlos, SP, Brazil. E-mail: emanuel@iqsc.usp.br*

[b]*Instituto Nacional de Ciência e Tecnologia de Bioanalítica-INCTBio, 13083-970 Campinas, SP, Brazil*

[c]*School of Chemistry and Biochemistry, Georgia Institute of Technology, 30332 Atlanta, GA, USA*

[d]*Departamento de Estatística, Universidade Federal de São Carlos, Rod. Washington Luís km 235, 13565-905 São Carlos, SP, Brazil*

[e]*Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 13566-590 São Carlos, SP, Brazil*

[f]*Instituto de Ciências da Saúde, Universidade Federal de Mato Grosso, 78557-267 Sinop, MT, Brazil*

[g]*Wama Produtos para Laboratório Ltda, 13560-971 São Carlos, SP, Brazil*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7ay00505a

Paper-based ELISA (p-ELISA) has the same versatility as conventional immunoenzymatic assays,[2] but it presents several advantages over the traditional analytical tools, such as low-cost, portability, low sample and reagent consumption, user-friendliness and analytical flexibility, besides a reduced waste generation.[2–4] Colorimetric paper-based outcomes can be digitalized using a flatbed scanner or simply photographed using a cellphone camera, and the digital image can be analyzed later offsite by a specialist or processed in real-time by a smartphone application.[3–9] p-ELISA has the potential to be implemented in developing countries and needy regions to diagnose neglected diseases, such as malaria,[10] dengue fever,[11] syphilis,[12] and HIV[13] which demonstrates the applicability and versatility of POCT devices such as ours.[5,8,14–16]

Much is said about the advantages related to the costs of these devices,[17] but besides the advantage of their low cost, these POCT tools bring an uncertainty related to their output, being necessary to establish parameters to evaluate their usefulness. The most common approaches involve the assessment of the traditional figures of merit of the method, which includes the limits of detection (LODs),[18] quantification (LOQs), dynamic linear range and analytical sensitivity, or a direct graphical comparison between the reference method (gold-standard) and the new method (the paper-based test, in this

case), fitting a model to observe if there is a correlation between the methods.[19] However, the graphical analysis by itself is not capable of assessing the real performance of the new test. Moreover, a simple analytical curve and the determination of the conventional figures of merit of the new method (such as LOD and LOQ) will be dealing with negative outcomes,[20] which are not relevant for clinical applications.

Thus, to determinate a safety margin to the outcomes of such low-cost testing devices, it is necessary to assess their performance indicators,[20] such as sensitivity (proportion of diseased people with a positive outcome), specificity (proportion of healthy people with a negative outcome)[21] and the accuracy (proportion of correct classifications).[22] Moreover, a critical step for the establishment of a new diagnostic test is the determination of the cut-off value, which will define the threshold value between healthy and diseased patients.[20,22]

Here we discuss these parameters: first, we developed a paper-based ELISA for detection of immunoglobulin G anti-*Toxoplasma gondii*, comparing the results with those of an ELISA benchtop assay (the gold-standard test). Then, we performed a statistical assessment of the performance of p-ELISA using ROC curves (sensitivity, specificity, and accuracy), besides comparing different methods to define the cut-off value of the test and uncertainty zone of the test (gray zone). Moreover, we bring for the first time a different approach to estimate the accuracy of the test using a macro in Microsoft Excel® (m-Accuracy), based on a non-parametric method[23] (available for downloading – ESI†).

Experimental

Chemicals and materials

Chromatography paper Whatman® no. 1; Qualitative filter paper 80GR J. PROLAB; Wax Printer Xerox® Phaser 8569; hot plate; GE Image Scanner III; Adobe Photoshop® CS5; Nestlé® MOLICO skimmed milk powder; Tween-20 detergent; protein extract from *Toxoplasma gondii* (WAMA Products, Ltd., São Carlos, Brazil); sodium carbonate (Synth®); sodium bicarbonate (Synth®); sodium chloride (Synth®); potassium chloride (Synth®); sodium phosphate dibasic (Synth®); potassium phosphate monobasic (Synth®); anti-human IgG-peroxidase antibody produced in goats (Sigma-Aldrich®); 3,3′,5,5′-tetramethylbenzidine (TMB) liquid substrate, supersensitive, for ELISA ready to use solution (Sigma-Aldrich®); TMB stop buffer (ScyTek Laboratories®); toxoplasmosis IgG ELISA kit (Katal Biotech Commercial Ind., Ltd); human serum samples (Laboratório Médico Dr Maricondi, Ltd). All chemicals were used as received, without any further purification.

Fabrication of paper-based microplates

Paper-based microplates were designed as the conventional 96-well plates, suitable to be used with multichannel pipettes. Sheets of Whatman no. 1 chromatography paper were cut into A4 size and inserted into the Xerox Phaser 8560 wax printer. The patterned papers were heated in a heat press during 2 minutes, at 150 °C, in order to melt the wax. The wax patterns delimited the microzones where the reactions could take place, due to hydrophobic characteristics of wax.[24]

Paper-based indirect immunoassay protocol

Sensitization. The sensitization solution was prepared from a protein extract solution (2.5 mg mL^{-1}), originated from the sonication of *T. gondii* sediment diluted in carbonate–bicarbonate buffer (0.6 mol L^{-1}, pH 9.6), in a 1 : 10 ratio. Each microzone was sensitized with 5 μL of the sensitization solution. After application of the reagents, the paper plate was allowed to dry at room temperature for 30 min. The spotted areas were washed twice with PBS-T buffer (0.01 mol L^{-1}, pH 7.4) with 0.075% (v/v) Tween-20 detergent, for removal of non-adsorbed proteins. An auxiliary filter paper positioned under the p-ELISA plate removed the excess of washing buffer.

Blocking. The application of blocking solution is necessary to block unoccupied sites during the sensitization step, thereby preventing nonspecific adsorption reactions, which result in false-positive outputs. A solution containing 10% (m V^{-1}) of skimmed milk powder and 0.075% (v/v) Tween-20 detergent in PBS buffer was used for the blocking step. The volume of 5 μL of blocking solution was spotted in each microzone, and the device was dried at room temperature. The microzones were washed twice.

Sample application. One hundred serum samples from patients with suspected toxoplasmosis (among immunocompromised patients and pregnant women) were kindly donated by Laboratório Médico Dr Maricondi Ltda (São Carlos, SP – Brazil). A volume of 5 μL of samples, at a 1 : 20 dilution in PBS-T solution with 1% (m V^{-1}) of skimmed milk, was added to each microzone. The paper was allowed to dry for 30 min at room temperature, and then the spots were washed 5 times with PBS-T washing solution, with the excess liquid being removed with an auxiliary filter paper or paper towel. The washing step removes antibodies not bound to the *T. gondii* antigens, avoiding false-positive outcomes.

Readout. In each zone, we added 5 μL of a human anti-IgG solution conjugated to peroxidase enzyme, at 1 : 1000 dilution in PBS-T solution with 1% (m V^{-1}) of skimmed milk powder. After 20 s, the spots were washed 10 times with PBS-T washing solution, removing the excess liquid with an auxiliary filter paper. After washing, we added 5 μL of a ready-to-use redox indicator solution of 3,3′,5,5′-tetramethylbenzidine (TMB) liquid substrate, forming a blue color on the paper zone. The colorimetric reaction was stopped after 2 min, with the addition of 5 μL TMB stop buffer, which is composed of an acidic solution, such as hydrochloric acid, changing the formed dye color from blue to yellow, due to the protonation of the oxidized form of the redox indicator. The devices were dried at room temperature during 30 min, and then digitalized using a flatbed scanner. The images were analyzed using the Adobe Photoshop® software, converting the digitalized test image to the CMYK color mode, and analyzing the mean pixel intensity at the yellow channel using the histogram tool. Fig. 1 presents a scheme of all steps involved in paper-based immunoassays, from manufacture of the platform to data acquisition.
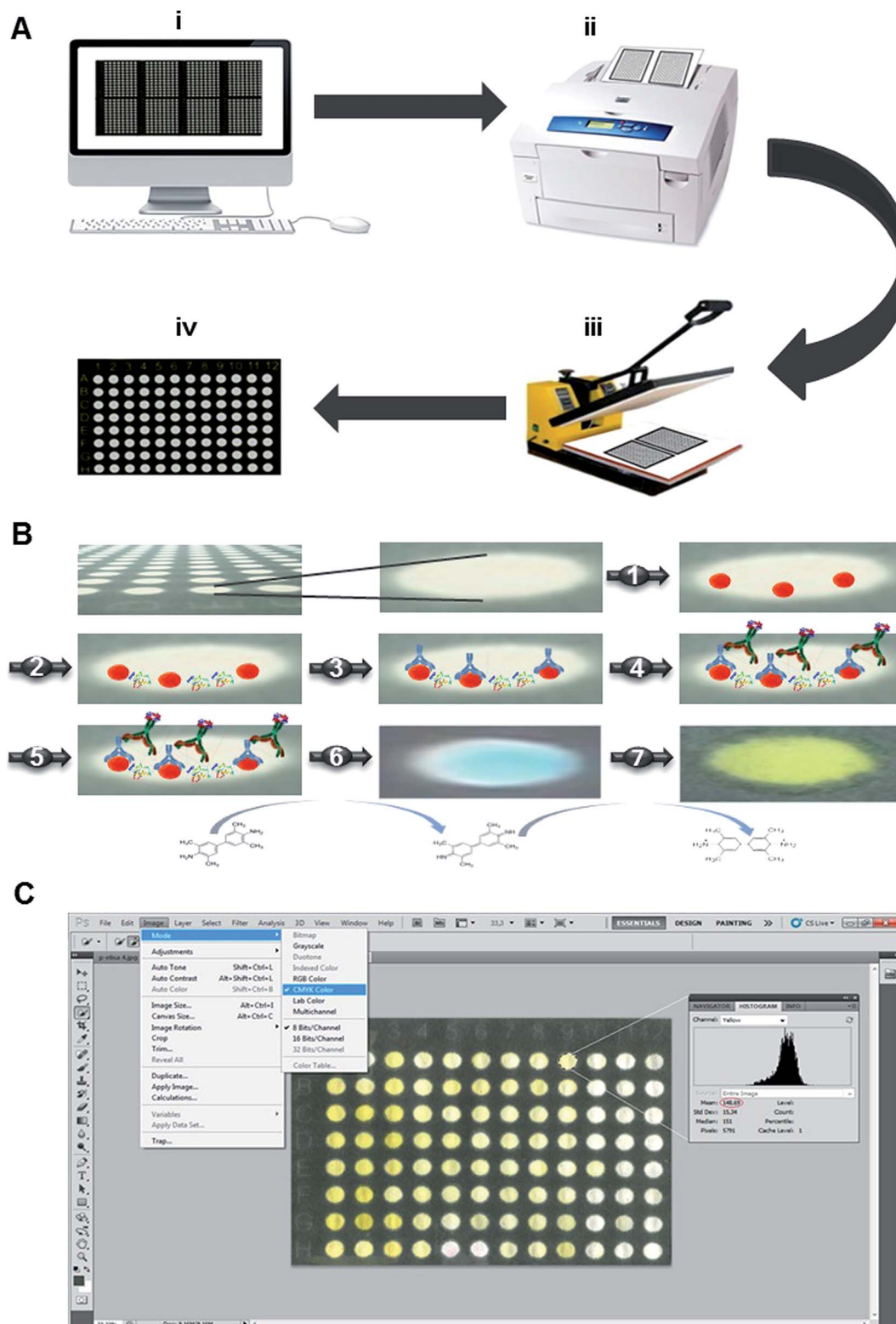
Fig. 1 General scheme of the steps involved in p-ELISA assay. (A) Steps involved in the manufacture of p-ELISA plates by wax printing: (i) layout design; (ii) paper patterning by wax printing; (iii) wax reflow by heating; (iv) ready-to-use p-ELISA plate. (B) Protocol for performing an immu-noassay on a p-ELISA plate: (1) sensitization step with *T. gondii* antigens; (2) blocking step with milk proteins; (3) serum sample addition; (4) human anti-IgG labeled with peroxidase addition step; (5) TMB substrate solution addition step; (6) revealing step; (7) TMB stop buffer addition step. (C) Digitalized image analysis on Adobe Photoshop®.

Sample analysis by a commercial ELISA kit and comparison of results

The samples (human serum) were also subjected to analysis by a commercially available toxoplasmosis ELISA kit. This test uses recombinant antigen of *T. gondii*, and the procedure was followed according to the manufacturer's instructions. The commercial microplates were subjected to the blocking step, and the samples were diluted at 1 : 20 ratio with the diluent solution. The spectrophotometric readings were performed using a microplate reader at 450 nm. The spectrophotometric results were compared to those obtained by the paper-based scanner readout.

Statistical analysis

We have compared two methods to determine the cut-off value for the paper-based assay. The first method makes use of a graph of sensitivity (Se) and specificity (Sp) of the test as a function of cut-off value, which determines the cut-off value by the intersection point between both curves.[25] The second method makes use of Youden's *J*-index: a graph of (sensitivity + specificity − 1) as a function of cut-off value, which determines the cut-off value by the maximum value in the curve.[26,27]

The sensitivity, specificity, and accuracy of the test were calculated using the area under the curve (AUC) of the receiver operating characteristic (ROC) curves, generated by the free software environment for statistical computing and graphics (R), the R-package "pROC".[28]

Additionally, we have created a macro in Microsoft Excel® to estimate the accuracy of the test using a non-parametric method, based on the work of Obuchowski *et al.*[23] (eqn (1)).

$$\theta = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} w(p_{it}, p_{js}) \sigma(i,j) \tag{1}$$

where *n* is the number of available results in the study, p_{it} is the result of the *i*-th patient tested with the gold standard t and therefore, p_{js} is the result of the *j*-th patient's gold standard. The function $\sigma(i,j)$ ensures that the weight *w* is only assigned if $j \neq i$, *i.e.*, if $j = i$, $\sigma(i,j) = 0$, otherwise $\sigma(i,j) = 1$. Then, the weights *w* are assigned as follows:

- $w = 1$, case t > s and $p_{it} > p_{js}$ or t < s e $p_{it} < p_{js}$;
- $w = 0.5$, case t = s and $p_{it} = p_{js}$;
- $w = 0$, otherwise.

Results and discussion

Comparison between p-ELISA and the gold standard ELISA

Enzyme-linked immunosorbent assays are considered the gold standard of clinical analysis, due to their high sensitivity (low number of false negatives) and specificity (low number of false positives). This immunoassay makes use of antigens or antibodies previously immobilized on the surface of a support, and reports the analyte detection by using specific antibodies labeled with enzymes. This versatile technique allows for distinct experiments, such as capture, sandwich or competitive assays. The limitations associated with ELISA are related to the need for a microplate reader, which has an approximate cost of US $ 20 000 (ref. 2) for an entry-level model, and the relatively long time needed for each step.

p-ELISA devices are being used for detection of auto-immune antibodies,[19] neuropeptides Y,[29] as well as HIV antigens,[2] showing versatility in clinical applications. The advantages of p-ELISA usage arise due to the paper characteristics, such as capillarity, thinness, lightweight and a large surface area, in comparison with traditional plastic substrates, which facilitates sample absorption and speeds up solvent dryness. These enzyme-immunoassays on a paper platform allow all steps to be performed at room temperature without the use of sophisticated equipment. Additionally, the p-ELISA devices can be manufactured locally in the laboratory meeting the needs of developing countries and needy regions for monitoring neglected diseases.

p-ELISA can make use of conventional technologies to digitalize the images, such as a flatbed scanner or a camera of a cellphone, and a software to analyze images (such as the freeware ImageJ – available at: https://imagej.nih.gov/ij/), with a total approximate cost of $ 100.[2] With the continuous studies in the area, paper-based analytical devices should provide as good results as the conventional tests do, with the benefits of low-cost, simplicity and user-friendliness.

The wax printing technology to fabricate paper-based devices is the simplest and most versatile way to create distinct patterns on paper, presenting great efficiency in the creation of hydrophobic barriers to contain reactions inside the testing zones (Fig. 2A and B), which avoids cross-contamination.

Considering the volume spent on carrying out the assays, the p-ELISA presents great advantage over the commercial ELISA tests. While p-ELISA uses 5 μL per step and per spot, conventional ELISA requires a minimum of 100 μL per step and per spot (Fig. S1 – ESI†). The low reagent and sample consumption is an attractive feature to neglected disease diagnosis due to the overall low-cost, besides generating less biological waste.

Regarding time requirements for carrying out the assays, the paper-based method also presents advantages over conventional tests. p-ELISA requires 20 min for each step, and 20 s incubation with secondary antibody (anti-human IgG), thus requiring around 60 min to complete the whole process. ELISA testing, on the other hand, requires over 300 min to complete the assay (Fig. S2 – ESI†), which is five times longer than this low-cost alternative.

In terms of costs, a p-ELISA microplate costs only US $ 0.10, against US $ 5.00 for a conventional ELISA plate (Fig. 2C). Thus, considering individual assays, our p-ELISA for toxoplasmosis diagnosis presented a cost of ∼US $ 0.34 per assay, against ∼US $ 0.92 per assay for the conventional ELISA assay (Fig. 2C). When we analyze costs per step, p-ELISA devices presented smaller costs than the conventional ELISA mainly in sensitization and revealing steps, resulting in a total cost difference of ∼$ 55.00. Moreover, the high cost associated with the ELISA microplate reader results in a larger price discrepancy between the conventional immunological technique and the paper-based assays.

Performance assessment of paper-based ELISA

It is important to emphasize that our discussion is focused more on the clinical aspects of the method rather than an
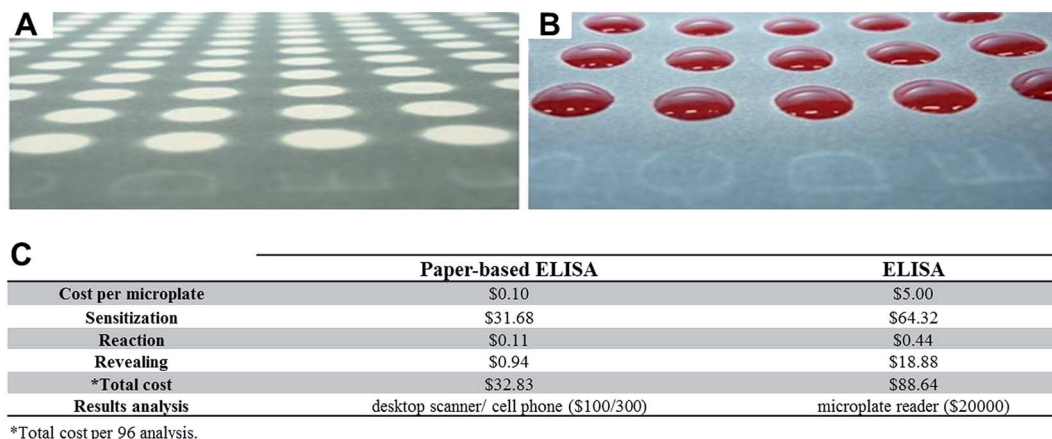
Fig. 2 Comparison between p-ELISA and ELISA techniques. (A) Ready-to-use paper-based device: reaction microzones (in white) separated by hydrophobic barriers (black). (B) Demonstration of liquid containment in reaction microzones, avoiding liquid overflow and cross-contamination. (C) Cost comparison between p-ELISA and conventional ELISA by the assay step.

analytical approach, and we are willing to bring this discussion to the analytical chemistry community. This means that the terms sensitivity and specificity do not correspond to the slope of a calibration curve nor to the ability to detect one species in the presence of others, respectively, but to the number of false negatives and false positives that the test provides. It also means that an analytical curve and the definition of the figures of merit of the method are not mandatory for comparison of assays, provided that a gold standard is one of the objects of study. Moreover, the validation of immunoassays requires the definition of the cut-off value, sensitivity, specificity, accuracy and gray zone, parameters which are not provided by an analytical curve.

The cut-off value is used as the threshold to differentiate sick from healthy patients.[27,30,31] The gold-standard ELISA bench kit used in this study for toxoplasmosis diagnosis presented a cut-off value at an absorbance of 1.0 and a confidence interval of ±10% (described by the manufacturer in the diagnostic kit label). This means that samples presenting an absorbance value below 0.9 were classified as negative, above 1.1 were classified as positive, and the rest situated between these values (0.9 < Abs < 1.1) were classified as uncertain (gray zone).

One hundred human serum samples were tested using the gold-standard ELISA bench kit, in which 62 samples were classified as positive, 33 as negative and 5 were classified as uncertain.

During the development of a new diagnostic test the determination of the cut-off value is necessary. A common approach is to plot a graph of sensitivity (Se) and specificity (Sp) as a function of cut-off, and the intersection point between both curves is chosen as the cut-off value.[25] By using this method, we have obtained a cut-off of 24.74 A.U. (arbitrary units from the mean pixel intensity of the yellow channel), with a specificity = sensitivity = 0.87 (Fig. 3A).

A second method involves a plot of Youden's J-index (where J = specificity + sensitivity − 1, and ranges from 0 to 1) as a function of cut-off, being the highest point of the curve chosen as the cut-off value of the test.[26,27] By using this method, we have obtained a cut-off of 21.73 A.U., with sensitivity = 0.96 and specificity = 0.87 (Fig. 3B).

It is relevant to underline that in Fig. 3A there is a region in between 19 to 24 A.U. in which the specificity of the test remains unchanged, while the sensitivity presents a sharp drop. It would be possible to choose an arbitrary point in this
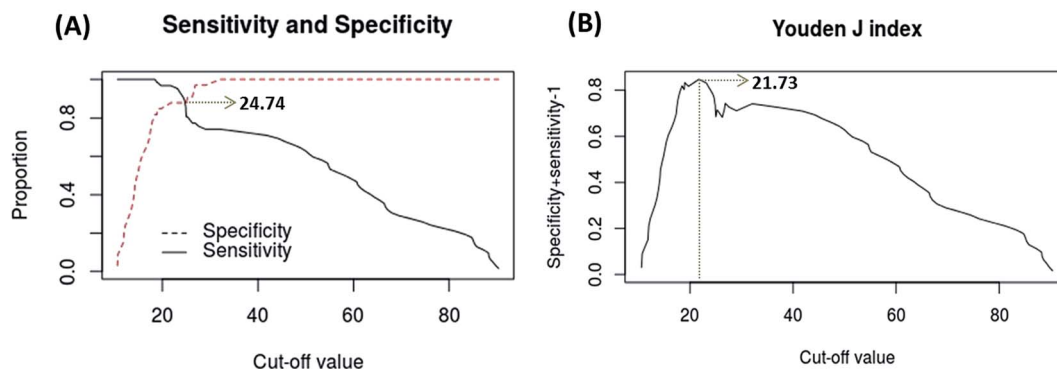

Fig. 3 Definition of the cut-off value of the test (A) Sensitivity (Se) and specificity (Sp) plot as a function of the cut-off value, with a cut-off value = 24.74 A.U., Se = Sp = 0.87. (B) Youden's J-index as a function of the cut-off value, with a cut-off value = 21.73 A.U., Se = 0.96, Sp = 0.87.

region in which both sensitivity and specificity are as high as possible. However, it is not advisable to define an arbitrary cut-off value for a diagnostic test. When Youden's *J*-index is used instead, the cut-off value is defined as the highest value for the combination of the variables (exactly in the region in between 19 to 24 A.U.), with the advantage of being a defined point (non-subjective decision), so we have chosen this non-arbitrary method to define the cut-off value for this diagnostic test.

The cut-off value cannot be used directly as a threshold value to separate diseased from health patients: even the gold-standard ELISA bench kit presents an uncertainty zone (0.9 < Abs < 1.1, in the present case) in which it is not possible to define patient's status.[32,33] Then, it is necessary to create

a three-zone partition to provide a confidence interval also for the new diagnostic test: diseased, non-diseased, and an inconclusive outcome region in between, known as the gray zone.[32] Analysis of the positive likelihood ratio (LR+) and negative likelihood ratio (LR−) can be used to create this zone,[32,33] as well as numerical factors on standard deviation (SD), or even a tolerance percentage on the cut-off value.[34–36] Different tolerance percentages on the cut-off value (10%, 12%, 15% or 20%) have showed suitability for the establishment of a gray zone in paper-based immunoassays for detection of tumor markers.[20] We have opted for the later approach, in an attempt of eliminating the largest number of samples classified as false positive or false negative (Fig. 4B–F).
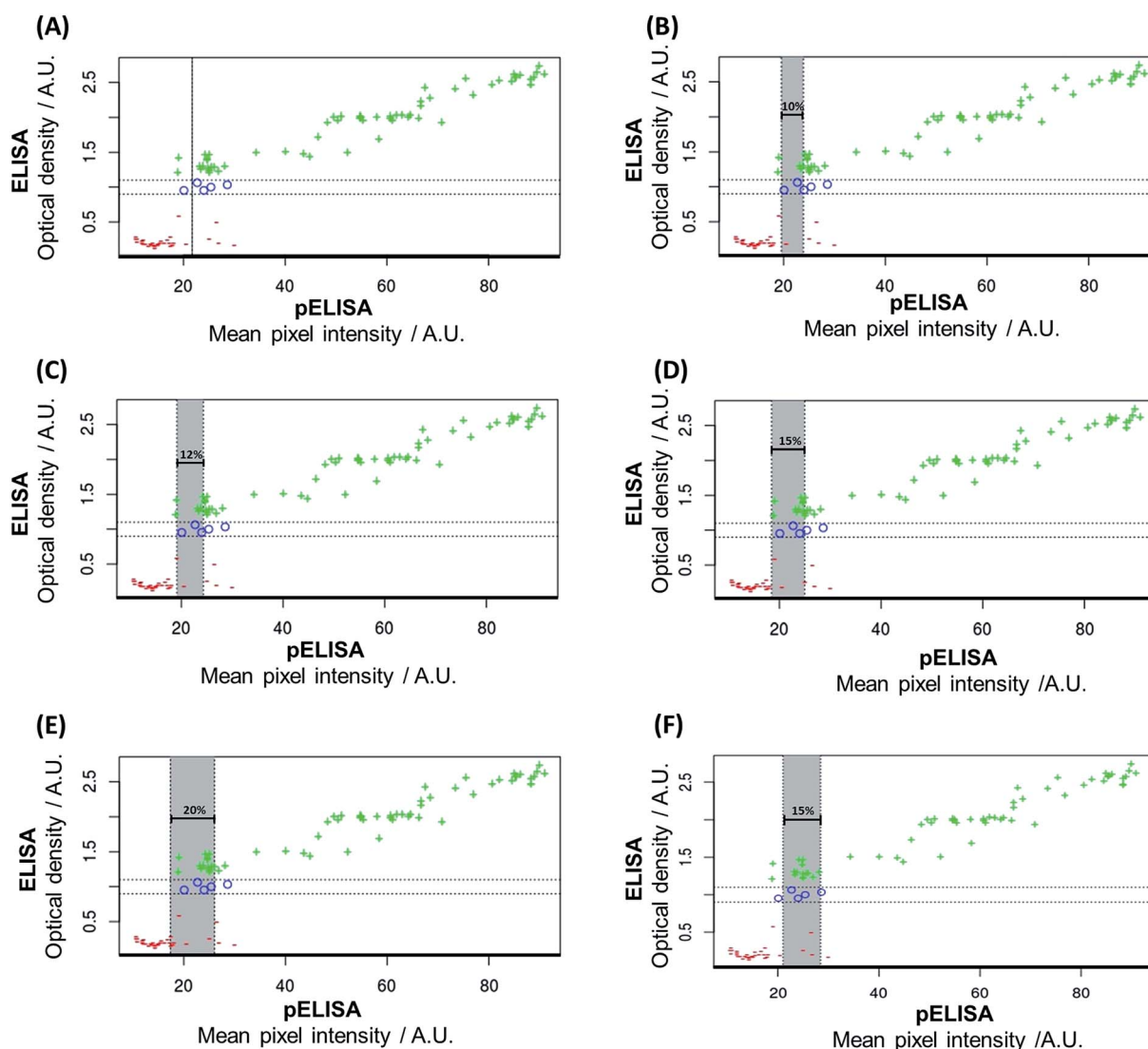

Fig. 4 Correlation between the gold-standard ELISA bench kit output (optical density) × p-ELISA output (mean pixel intensity in the yellow channel). The gray zone was defined using a tolerance percentage on the cut-off value. (A) Cut-off value defined by Youden's *J*-index (mean pixel intensity of 21.73 A.U.). (B) Tolerance percentage of ±10% on 21.73 A.U. (C) Tolerance percentage of ±12% on 21.73 A.U. (D) Tolerance percentage of ±15% on 21.73 A.U. (E) Tolerance percentage of ±20% on 21.73 A.U. (F) Cut-off value defined by the graph of sensitivity (Se) and specificity (Sp) as a function of cut-off (mean pixel intensity of 24.74 A.U.) with a tolerance percentage of ±15% on the cut-off value. + = positive outcome defined by the gold-standard ELISA bench kit; − = negative outcome defined by the gold-standard ELISA bench kit; o = uncertain outcomes (gold-standard ELISA bench kit gray zone).

From Fig. 4A it is observed that when no tolerance percentage is used there are 2 false negative and 4 false positive outcomes misclassified by the p-ELISA test. When the tolerance percentage is set to $\pm 10\%$ of the cut-off value (Fig. 4B), the gray zone encompasses 6 samples (but misses 2 false-negatives and 4 false-positives). When the tolerance percentage is set to $\pm 12\%$ of the cut-off value (Fig. 4C), the gray zone encompasses 8 samples (still missing 2 false-negatives and 4 false-positives). These percentages ($\pm 10\%$ and $\pm 12\%$ of the cut-off value) were not able to eliminate the false negative outcomes or to decrease the amount of false positives.

However, when the tolerance percentage is set to $\pm 15\%$ of the cut-off value (Fig. 4D), the gray zone encompasses 16 samples and removes all false negative results, besides reducing the number of samples classified as false positives to 3. A further increase in the tolerance percentage (20% of the cut-off value, Fig. 4E) presents no advantage over 15%, once it encompasses 23 samples (23% of all tested samples) and misses 3 false-positive outcomes. We have evaluated the best tolerance percentage ($\pm 15\%$) on the cut-off value determined by the first method (24.74 A.U.), which resulted in a gray zone encompassing 20 samples, but presenting a false-positive and 2 false-negative outcomes (Fig. 4F). These results have demonstrated that the use of the cut-off value determined by Youden's *J*-index and the tolerance percentage of $\pm 15\%$ of the cut-off value presented the best choice for this assay.

Paper-based ELISA protocols still require optimizations, because of their intrinsic multistep requirements that heavily depend on the operator skills (washing, blocking, and sensing steps).[1,20] Even though lateral flow immunoassays[14,37,38] and other bioassays carried out in low-cost platforms[39,40] do not present such requirements, being faster and easier to use, pELISA testing tools overcome these testing tools because they present advantages such as a great multiplexing capacity and improved quantitative outputs,[20] and their low-cost also allows for broad-range screening tests in needy regions. These advantages are attainable only for tests with a high sensitivity (low number of false negative outputs),[41] as those demonstrated by our device after the definition of its tolerance percentage. Patients with a positive outcome in a screening test are then referred to a health center for further testing. In the case of a false positive output for toxoplasmosis, for example, the clinical center would perform more specific tests in order to confirm the absence of the disease, which does not affect the patient's health. However, when tests present false negative outcomes the individuals do not receive adequate treatment, which allows for disease progression, increased treatment costs and lower chances of success. These scenarios justify a tolerance percentage of $\pm 15\%$ of the cut-off value due to the complete elimination of false negative outputs.

The Receiver Operating Characteristic (ROC) curve is one of the most used tools for the measurement of accuracy of diagnostic tests.[23,25,41] The ROC curve is generated by plotting the sensitivity of the test against ($1 -$ specificity),[41] and it requires the classification of the samples by the gold-standard in a dichotomous outcome (1 for positive and 0 for negative).[23,25] There is a comparison between the classification made by the
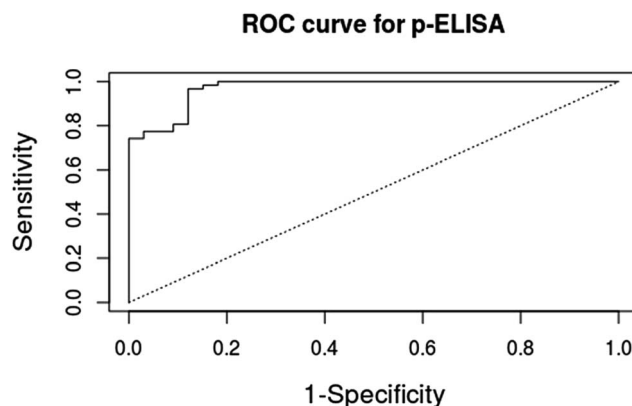

Fig. 5 ROC curve for the p-ELISA point-of-care testing tool. The gold standard used to classify the samples was a commercial toxoplasmosis ELISA bench kit. 95 samples were used to generate this curve.

gold standard and the outputs of the new diagnostic test (p-ELISA).[23] However, there are cases in which the gold standard provides a continuous outcome instead of dichotomous outcome, and the common practice is to force the dichotomization of the results, which requires the exclusion of the samples classified in the gray zone area of the gold standard (5 samples in the present case). We used the R-package pROC[28] to obtain the ROC curve for the toxoplasmosis p-ELISA testing of the 95 remaining samples, as shown in Fig. 5.

The accuracy was calculated using the area under the curve (AUC) of Fig. 5. The maximum accuracy value is one, which represents a perfect test, and our paper-based immunoassay toxoplasmosis diagnosis presented an AUC = 0.97, defined as a highly precise diagnostic test, close to the maximum accuracy.[23] It is relevant to say that the forced dichotomization of the test with the subsequent exclusion of the samples classified in the gray zone area of the gold standard tends to overestimate the calculated accuracy of the diagnostic assay being evaluated. It is necessary then to compute the samples in the gray zone of the gold standard in order to obtain a more realistic estimate of the accuracy of the new diagnostic test.[23]

Improved accuracy measurements

In order to obtain a more reliable accuracy estimate using all the samples it is necessary to use a non-parametric test, as that proposed by Obuchowski *et al.* (eqn (1)).[23] By using such a method, we have created a macro in Microsoft Excel® to calculate the accuracy of a new diagnostic method (m-Accuracy), and we tested it for our p-ELISA.

The m-Accuracy method compares each sample output with all other sample outputs in both the gold-standard and the new method, analyzing if the variation that occurs in each sample pair in the gold standard method also occurs in the paper-based ELISA. Different weights are assigned according to the event that occurs. In other words, a value of 1 is assigned to a specific pair when the sample with a higher outcome value in the gold-standard also presents a higher value in the new method. If there is no difference in values between a specific pair of
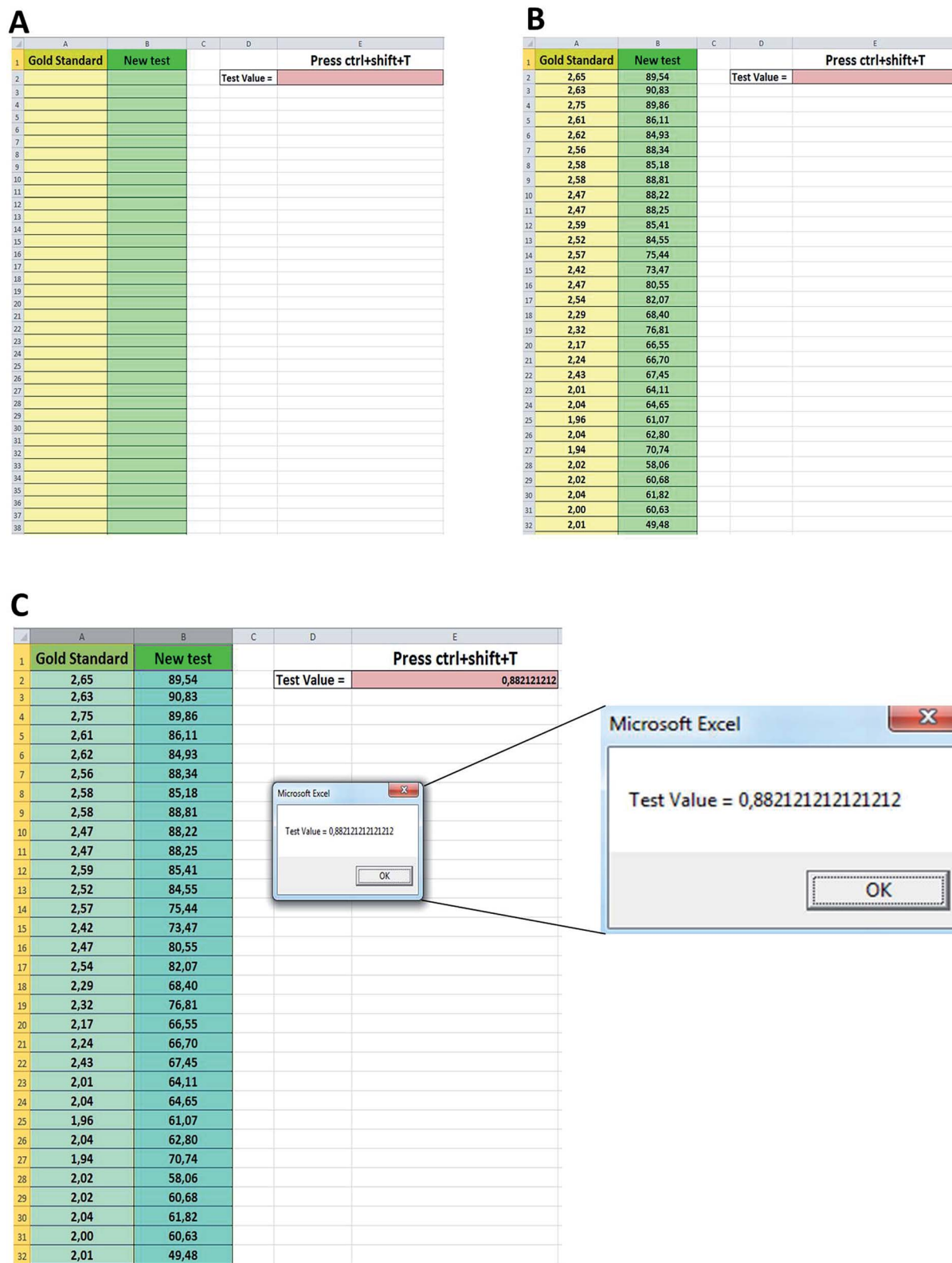
Fig. 6 Schematic diagram showing all the steps to calculate the accuracy of a new diagnostic test by the m-Accuracy tool. (A) Initial macro screen. (B) Data input: gold standard method in the first column, new method in the second column. (C) Selection of the two columns containing the values of the results, and the ctrl + shift + T command pressed, providing an accuracy of 0.88.

samples, then a value of 0.5 is assigned to that pair. If neither of these events occurs, then the model interprets that as a "misleading outcome" and a value of 0 is assigned to that pair. Thus, to estimate the accuracy of the test these values are summed up and then divided by the number of pairs compared. The analysis is performed by a single command (ctrl + shift + T) and the functioning of the m-Accuracy tool is very straightforward. There are two columns in the macro: in the first column the values obtained by the gold standard test are inputted (in this case 100 results (mean of triplicates)), while in the second column the values obtained by the new diagnostic test are inputted (for the same samples, in the same order). The ROC curve provided an overestimated accuracy of 0.97, against a more reliable value of accuracy of 0.88 provided by the m-Accuracy tool, which classifies the test as moderately to highly accurate.[23] Fig. 6 illustrates the operation of m-Accuracy in a tutorial format.

It is relevant to notice that the use of the m-Accuracy tool becomes very attractive since all samples are used in the accuracy computation, including those within the gray zone of the gold standard, by comparing the results obtained by the gold standard method and the new diagnostic test. Due to the ease of use and more reliable results provided by this tool we provide m-Accuracy for download as part of our ESI.†

Conclusions

There are several advantages of paper-based ELISA assays in comparison with conventional immunological tests, such as: (i) ease of local manufacture; (ii) small consumption of reagents and samples; (iii) short time of analysis; (iv) do not require an expensive microplate reader for data readout, which can be performed by a cell phone camera; and (v) all stages are performed at room temperature, which dispenses the use of refrigeration. All these characteristics allow for the use of this technology in remote locations with scarce resources, which is useful for the clinical management of patients, becoming a great ally of telemedicine. What we evidence here is the need for the determination of the safety margins for these low-cost diagnostic tools. More than that, we want to present here how the evaluation of the performance of a new low-cost diagnostic test should be performed, including the creation of a simple tool to estimate its accuracy. Thus, by creating a p-ELISA for a neglected disease (toxoplasmosis diagnosis using IgG antibodies against *T. gondii* present in the sample), we performed a statistical evaluation of this new low-cost testing tool, calculating a cut-off value (21.73 A.U.), sensitivity (0.96), specificity (0.87) and gray zone definition (±15% of the cut-off value, encompassing 16 samples, with 3 false positive outputs), besides an estimate of accuracy by using ROC curves (AUC = 0.97). When we used the non-parametric tool to estimate the accuracy of the test by a macro (m-Accuracy), we have obtained a value = 0.88, which is more reliable than the accuracy obtained just by the ROC curve. The m-Accuracy macro appears as a very attractive tool that can become a trend for accuracy estimation in diagnostic test validation, being able to reach other areas of chemistry. We envision that the use of these tools can improve investigation in the low-cost diagnostics field, which can bring this technology for those who need it.

References

1 R. M. Lequin, *Clin. Chem.*, 2005, **51**, 2415–2418.
2 C.-M. Cheng, A. W. Martinez, J. Gong, C. R. Mace, S. T. Phillips, E. Carrilho, K. A. Mirica and G. M. Whitesides, *Angew. Chem., Int. Ed. Engl.*, 2010, **49**, 4771–4774.
3 E. Carrilho, S. T. Phillips, S. J. Vella, A. W. Martinez and G. M. Whitesides, *Anal. Chem.*, 2009, **81**, 5990–5998.
4 A. W. Martinez, S. T. Phillips, B. J. Wiley, M. Gupta and G. M. Whitesides, *Lab Chip*, 2008, **8**, 2146–2150.
5 C. D. Chin, Y. K. Cheung, T. Laksanasopin, M. M. Modena, S. Y. Chin, A. A. Sridhara, D. Steinmiller, V. Linder, J. Mushingantahe, G. Umviligihozo, E. Karita, L. Mwambarangwe, S. L. Braunstein, J. van de Wijgert, R. Sahabo, J. E. Justman, W. El-Sadr and S. K. Sia, *Clin. Chem.*, 2013, **59**, 629–640.
6 A. W. Martinez, S. T. Phillips, E. Carrilho, S. W. Thomas, H. Sindi and G. M. Whitesides, *Anal. Chem.*, 2008, **80**, 3699–3707.
7 M. Webster and V. S. Kumar, *Clin. Chem.*, 2012, **58**, 956–958.
8 C. D. Chin, T. Laksanasopin, Y. K. Cheung, D. Steinmiller, V. Linder, H. Parsa, J. Wang, H. Moore, R. Rouse, G. Umviligihozo, E. Karita, L. Mwambarangwe, S. L. Braunstein, J. van de Wijgert, R. Sahabo, J. E. Justman, W. El-Sadr and S. K. Sia, *Nat. Med.*, 2011, **17**, 1015–1019.
9 N. L. Ruiz, V. F. Curto, M. M. Erenas, F. Benito-, D. Diamond, A. José, P. López and L. F. Capitan-vallvey, *Anal. Chem.*, 2014, **86**, 9554–9562.
10 S. Kim, S. Nhem, D. Dourng and D. Ménard, *Malar. J.*, 2015, **14**, 114.
11 V. C. Gan, L.-K. Tan, D. C. Lye, K.-Y. Pok, S.-Q. Mok, R. C.-R. Chua, Y.-S. Leo and L.-C. Ng, *PLoS One*, 2014, **9**, e90037.
12 D. C. Mabey, K. A. Sollis, H. A. Kelly, A. S. Benzaken, E. Bitarakwate, J. Changalucha, X.-S. Chen, Y.-P. Yin, P. J. Garcia, S. Strasser, N. Chintu, T. Pang, F. Terris-Prestholt, S. Sweeney and R. W. Peeling, *PLoS Med.*, 2012, **9**, e1001233.

13 H. Shafiee, S. Wang, F. Inci, M. Toy, T. J. Henrich, D. R. Kuritzkes and U. Demirci, *Annu. Rev. Med.*, 2015, **66**, 387–405.

14 J. Hu, S. Wang, L. Wang, F. Li, B. Pingguan-Murphy, T. J. Lu and F. Xu, *Biosens. Bioelectron.*, 2014, **54**, 585–597.

15 G. M. Whitesides, *Clin. Chem.*, 2013, **59**, 589–591.

16 T. Mazzu-Nascimento, P. A. G. C. Leão, J. R. Catai, G. G. Morbioli and E. Carrilho, *Anal. Methods*, 2016, **8**, 7312–7318.

17 D. M. Cate, J. A. Adkins, J. Mettakoonpitak and C. S. Henry, *Anal. Chem.*, 2014, **87**, 19–41.

18 W. Liu, Y. Guo, M. Zhao, H. Li and Z. Zhang, *Anal. Chem.*, 2015, **87**, 7951–7957.

19 C. Hsu, H. Huang, W. Chen, W. Nishie, H. Ujiie, K. Natsuga, S. Fan, H. Wang, J. Y. Lee, W. Tsai, H. Shimizu and C. Cheng, *Anal. Chem.*, 2014, **86**, 4605–4610.

20 T. Mazzu-Nascimento, G. G. Morbioli, L. A. Milan, F. C. Donofrio, C. A. Mestriner and E. Carrilho, *Anal. Chim. Acta*, 2016, **950**, 156–161.

21 A. K. Akobeng, *Acta Paediatr.*, 2007, **96**, 338–341.

22 A. K. Akobeng, *Acta Paediatr.*, 2007, **96**, 644–647.

23 N. A. Obuchowski, M. L. Lieber and F. H. Wians, *Clin. Chem.*, 2004, **50**, 1118–1125.

24 E. Carrilho, A. W. Martinez and G. M. Whitesides, *Anal. Chem.*, 2009, **81**, 7091–7095.

25 M. Greiner, D. Pfeiffer and R. D. Smith, *Prev. Vet. Med.*, 2000, **45**, 23–41.

26 W. J. Youden, *Cancer*, 1950, **3**, 32–35.

27 E. F. Schisterman, N. J. Perkins, A. Liu and H. Bondell, *Epidemiology*, 2005, **16**, 73–81.

28 E. R. DeLong, D. M. DeLong and D. L. Clarke-Pearson, *Biometrics*, 1988, **44**, 837–845.

29 R. C. Murdock, L. Shen, D. K. Griffin, N. Kelley-Loughnane, I. Papautsky and J. A. Hagen, *Anal. Chem.*, 2013, **85**, 11634–11642.

30 H. Xu, J. Lohr and M. Greiner, *J. Immunol. Methods*, 1997, **208**, 61–64.

31 M. Greiner, D. Sohr and P. Göbel, *J. Immunol. Methods*, 1995, **185**, 123–132.

32 J. Coste and J. Pouchot, *Int. J. Epidemiol.*, 2003, **32**, 304–313.

33 J. Coste, P. Jourdain and J. Pouchot, *Clin. Chem.*, 2006, **52**, 2229–2235.

34 W. H. Kim, J. H. Lee, E. Kim, G. Kim, H. J. Kim and H. W. Lim, *Thorac. Cardiovasc. Surg.*, 2016, **64**, 281–289.

35 G. Icardi, F. Ansaldi, B. M. Bruzzone, P. Durando, S. Lee, C. D. E. Luigi and P. Crovari, *J. Clin. Microbiol.*, 2001, **39**, 3110–3114.

36 J. C. P. Muñoz, L. G. Oliveira, A. Carvalho Braga, I. M. Trevisol and P. M. Roehe, *Pesqui. Vet. Bras.*, 1999, **19**, 123–127.

37 D. Quesada-González and A. Merkoçi, *Biosens. Bioelectron.*, 2015, **73**, 47–63.

38 E. M. Linares, L. T. Kubota, J. Michaelis and S. Thalhammer, *J. Immunol. Methods*, 2012, **375**, 264–270.

39 A. W. Martinez, S. T. Phillips, M. J. Butte and G. M. Whitesides, *Angew. Chem., Int. Ed.*, 2007, **46**, 1318–1320.

40 W. K. Tomazelli Coltro, C. M. Cheng, E. Carrilho and D. P. de Jesus, *Electrophoresis*, 2014, **35**, 2309–2324.

41 A. K. Akobeng, *Acta Paediatr.*, 2007, **96**, 644–647.