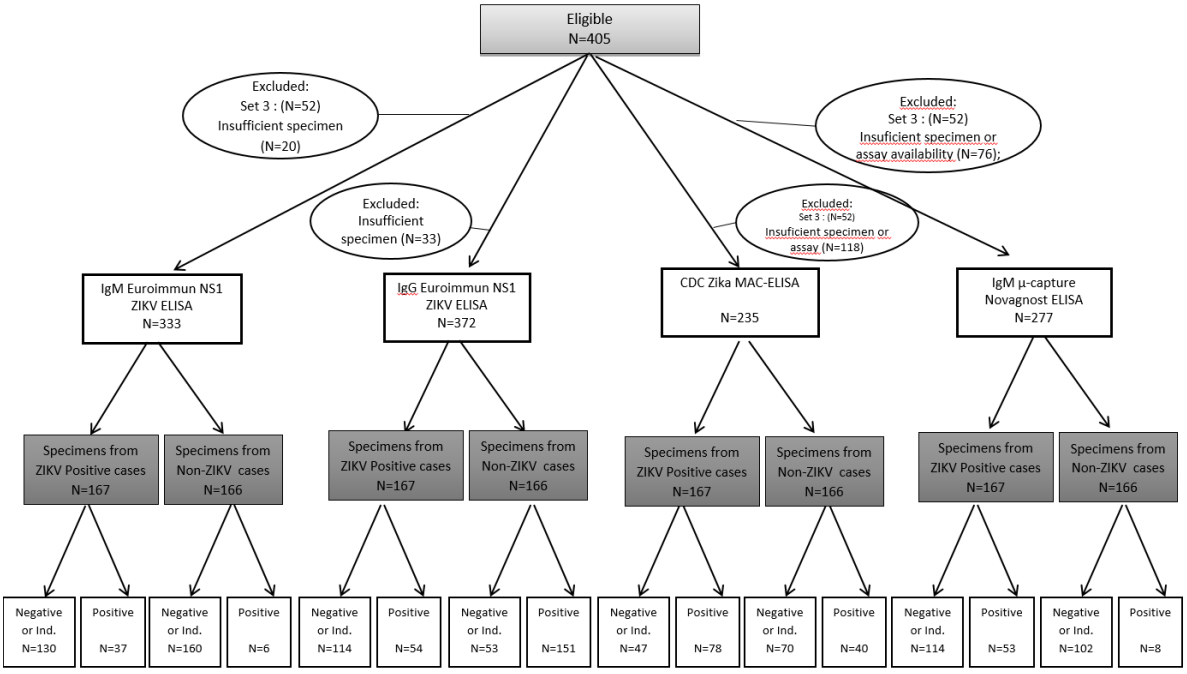
Supplementary material:

Flow Chart S1: IgM and IgG ZIKV ELISA Assays evaluated include: IgM and IgG Euroimmun NS1 ZIKV assay, CDC Zika MAC-ELISA and IgM μ-capture Novagnost ELISA.

ZIKV Positive cases have clinical presentation of ZIKV (rash-fever symptoms) and positive detection of ZIKV RNA by RT-PCR. Non-ZIKV cases or Controls are specimens collected in or prior to 2013, before the arrival of ZIKV in Rio de Janeiro. Ind- indeterminate result in assay, following manufacturer's instructions.



Flow Chart S2: IgM and IgG DENV Panbio ELISA and Plaque Reduction Neutralization Testing (PRNT) assays evaluated.

ZIKV Positive cases have clinical presentation of ZIKV (rash-fever symptoms) and positive detection of ZIKV RNA by RT-PCR. Non-ZIKV cases or controls are specimens collected in or

prior to 2013, before the arrival of ZIKV in Rio de Janeiro. Ind- indeterminate result in assay, following manufacturer's instructions.

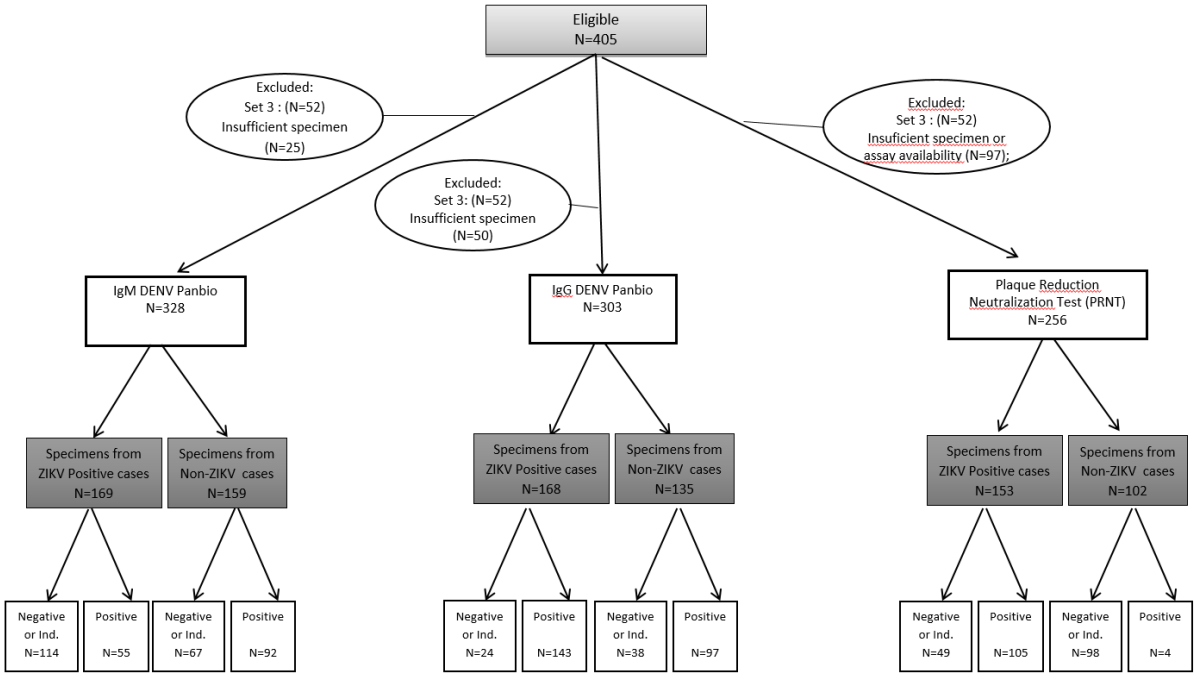


Table S1 Breakdown of serial collections of the ZIKV study population. First serum collection is PCR positive (Set1; ZIKV Panel).

ZIKA Panel	ZIKV Positive Patients	Total Samples	Year of Collection
1 Collection	5	5	2015-2016
2 Collections	55	110	2015-2016
3 Collections	7	18	2015-2016
>4 Collections	4	33	2015-2016
TOTAL	72	169	2015-2016

Table S2. Sensitivity and specificity for the IgM and IgG anti-DENV antibody ELISAs focusing on DENV detection.

	IgM DENV Panbio ELISA			IgG DENV Panbio ELISA		
	Tested (n)	% Sens. (95%CI)	% Spec. (95% CI)	Tested (n)	% Sens. (95%CI)	% Spec. (95% CI)
Zika Positive	169		67.5(60.4-74.5)	168		14.4(9.1-19.7)
Zika (1-6 Days)	79		77.2(70.2-84.3)	78		16.7(11.3-22.0)
Zika (≥7-13 Days)	41		48.8(41.7-55.8)	41		14.6(9.3-20.0)
Zika (≥14 Days)	49		67.3(60.3-74.4)	49		10.2(4.9-15.5)
Zika (≥7 Days)	90		58.9(51.8-66.0)	90		12.2(6.9-17.5)
DENV (all)	90	83.3(75.6-91)		88	81.8(73.8-89.9)	
DENV1	21	95.2(86.1-100)		21	52.4(31-73.7)	
DENV2	17	100(100-100)		16	100(100-100)	
DENV3	21	95.2(86.1-100)		20	85(69.4-100)	
DENV4	31	83.3(75.6-91)		31	90.3(79.9-100)	
Overall (≥7 days)	**249**	83.3(75.6-91)	58.9(51.8-66.0)	**225**	81.8(73.8-89.9)	12.2(6.9-17.5)
PPV (≥7 days)		67(60.9-72.5)			47.7(44.6-50.8)	
NPV (≥7 days)		77.9(68.3-85.3)			40.7(25.3-58.3)	
Accuracy (≥7 days)		71.1(63.9-77.6)			46.6(39.1-54.2)	

Table S2 - The sensitivity and specificity of the IgM and IgG Panbio DENV commercial ELISA with the DENV panel (Set 2) and the Control ZIKV group (Set 1). DENV samples were from individuals with a clinical presentation characteristic of DENV and with a positive DENV PCR result and they were collected in the following years: DENV1 (2010 and 2011), DENV2 (2008, 2010, 2011), DENV3 (2002, 2007, 2008) and DENV4 (2012 and 2013). Specificity values were calculated for each assay based on the Set 1 (individuals with confirmed ZIKV PCR positivity). Data from DENV-positive cases served only for determining the sensitivity and was not used for the specificity calculation. Overall sensitivity, PPV, NPV and accuracy were calculated with the DENV Positive samples.

Note: Sens., Sensitivity; Spec., Specificity; CI, Coefficient Interval; PPV, Positive Predictive Value; NPV, Negative Predictive Value; ZIKV, Zika virus; DENV, Dengue virus; Days, number of days the sample was collected after symptom onset. Indeterminate results were considered negative for the calculation. DENV (all) includes all DENV samples (DENV 1-4).

Table S3. Sensitivity, specificity and highest likelihood ratio of various cutoff values for IgG NS1 Euroimmun ZIKV ELISA as based on the ROC analysis performed in Figure 6. *Current cutoff suggested by the manufacturer; ** Suggested cutoff with higher likelihood ratio. CI: Coefficient Interval

CUTOFF	Sensitivity%	95% CI	Specificity%	95% CI	Likelihood ratio
> 0.5400	92.22	84,63% to 96,82%	65.13	56,99% to 72,67%	2.645
> 0.8350	91.11	83,23% to 96,08%	70.39	62,46% to 77,52%	3.078
> 0.9450	90	81,86% to 95,32%	73.03	65,24% to 79,90%	3.337
> 1.005	88.89	80,51% to 94,54%	75.66	68,04% to 82,25%	3.652
> 1.025	88.89	80,51% to 94,54%	76.32	68,75% to 82,83%	3.753
> 1.095*	**86.67**	**77,87% to 92,92%**	**77.63**	**70,17% to 83,98%**	**3.875**
> 1.220	83.33	74,01% to 90,36%	78.29	70,88% to 84,56%	3.838
> 1.295	82.22	72,74% to 89,48%	78.29	70,88% to 84,56%	3.787
> 1.375	82.22	72,74% to 89,48%	80.26	73,04% to 86,27%	4.166
> 1.500**	**78.89**	**69,01% to 86,79%**	**82.24**	**75,22% to 87,96%**	**4.441**
> 1.650	73.33	62,97% to 82,11%	82.89	75,95% to 88,51%	4.287
> 1.835	70	59,43% to 79,21%	85.53	78,91% to 90,70%	4.836
> 2.040	65.56	54,80% to 75,26%	87.5	81,17% to 92,30%	5.244
> 2.285	62.22	51,38% to 72,23%	89.47	83,47% to 93,86%	5.911
> 2.495	58.89	48,02% to 69,16%	91.45	85,82% to 95,37%	6.885
> 2.835	54.44	43,60% to 64,98%	92.76	87,42% to 96,33%	7.523
> 3.275	50	39,27% to 60,73%	94.74	89,89% to 97,70%	9.5
> 4.200	43.33	32,92% to 54,20%	96.71	92,49% to 98,92%	13.17

Checklist S1. Standards for the Reporting of Diagnostic accuracy studies (STARD)

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	3
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	4,5
	4	Study objectives and hypotheses	5
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	5, 6
Participants	**6**	Eligibility criteria	6, 7
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	6, 7
	8	Where and when potentially eligible participants were identified (setting, location and dates)	6, 7
	9	Whether participants formed a consecutive, random or convenience series	6, 7
Test methods	**10a**	Index test, in sufficient detail to allow replication	7, 8
	10b	Reference standard, in sufficient detail to allow replication	8
	11	Rationale for choosing the reference standard (if alternatives exist)	6, 7
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	7,8
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	8
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	7
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	7
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	8
	15	How indeterminate index test or reference standard results were handled	8
	16	How missing data on the index test and reference standard were handled	9
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	N/A
	18	Intended sample size and how it was determined	N/A
RESULTS			
Participants	**19**	Flow of participants, using a diagram	Flowchart S1 and Flowchart S2
	20	Baseline demographic and clinical characteristics of participants	10, Table 1
	21a	Distribution of severity of disease in those with the target condition	N/A.
	21b	Distribution of alternative diagnoses in those without the target condition	10, Table 1

	22	Time interval and any clinical interventions between index test and reference standard	6, Table 1
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	9, Table 2, Table 3
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	9, 10, 11, Table2, Table 3
	25	Any adverse events from performing the index test or the reference standard	N/A as only the MAC-CDC ELISA assay was used for reporting results as part of local standard of care
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	*14,15*
	27	Implications for practice, including the intended use and clinical role of the index test	14, 15, 16
OTHER INFORMATION			
	28	Registration number and name of registry	N/A
	29	Where the full study protocol can be accessed	N/A
	30	Sources of funding and other support; role of funders	17

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.