Markus A. Thaler, Andreas Bietenbeck, Meng-Xin Yin, Udo Steigerwald, Andrew B. Holmes, Edelgard Lindhoff-Last and Peter B. Luppa*

Evaluation of antiphospholipid antibody assays using latent class analysis to address the lack of a reference standard

DOI 10.1515/cclm-2016-0116
Received February 12, 2016; accepted April 14, 2016

Abstract

Background: Method evaluation of new assays for the detection of antiphospholipid antibodies (aPL) such as anti-cardiolipin (aCL) or anti-β2-glycoprotein I (aβ2-GPI) is challenging, as no internationally accepted reference material is available yet. Besides a lack of standardization, unacceptable inter-laboratory comparability of established tests is regularly observed. Owing to the absence of a commonly accepted reference standard, the evaluation of two research surface plasmon resonance (SPR) biosensor assays was performed using statistical methods from latent class analysis (LCA).

Methods: aCL and aβ2-GPI IgG and IgM were measured in sera from 63 antiphospholipid syndrome patients, fulfilling the Sydney criteria, and in 34 healthy controls with four commercial assays. LCA was performed on the results and sera were assigned to the antibody-positive or antibody-negative group. Sera were subsequently evaluated in the SPR assays for aCL and aβ2-GPI. Optimal cutoffs and diagnostic performances of the research systems were established employing the LCA-derived gold standard.

Results: With area under the curve results of 0.96 and 0.89 for the detection of aCL and aβ2-GPI, the research SPR assays discriminated well between antibody-positive and antibody-negative sera. Their sensitivities and specificities were comparable to the investigated commercial immunoassays.

Conclusions: SPR assays are a suitable tool for the detection of aCL and aβ2-GPI with diagnostic performances not different from currently available commercial tests. LCA enabled the calculation of sensitivities and specificities for aPL assays in absence of a reference standard.

Keywords: antiphospholipid antibodies; biosensor; latent class analysis; method evaluation; surface plasmon resonance.

Introduction

According to the currently employed Sydney classification [1], positive titers of anti-cardiolipin antibodies (aCL) IgG or IgM and of anti-β2-glycoprotein I antibodies (aβ2-GPI) IgG or IgM constitute laboratory criteria for the diagnosis of antiphospholipid syndrome (APS) [2]. Typically, these autoantibodies are assessed in immunoassays [3]. To extend the diagnostic tools, our group established surface plasmon resonance (SPR) biosensor assays for the measurement of aCL and aβ2-GPI [4, 5]. Upon preparing the in-depth evaluation of diagnostic performances, we encountered the methodological question every developer of antiphospholipid antibody (aPL) tests inevitably faces: against which standard or method should the new test be evaluated?

Currently available methods lack standardization, which may explain the poor inter-laboratory comparability of both aCL and aβ2-GPI assays. This is irrespective of whether the results are reported as numerical values or in a semiquantitative or qualitative manner [6–9]. Substantial efforts have been undertaken to improve standardization. Attempts to establish an internationally accepted reference material resulted in the so-called Harris standards (mixtures of high positive and normal sera [10])

***Corresponding author: Peter B. Luppa,** Institut für Klinische Chemie und Pathobiochemie, Klinikum rechts der Isar der Technischen Universität München, Ismaninger Strasse 22, 81675 München, Germany, Phone: +49-89-4140-4759, Fax: +49-89-4140-4875, E-mail: p.luppa@tum.de
Markus A. Thaler and Andreas Bietenbeck: Institut für Klinische Chemie und Pathobiochemie, Klinikum rechts der Isar der Technischen Universität München, München, Germany
Meng-Xin Yin and Andrew B. Holmes: School of Chemistry, Bio21 Institute, University of Melbourne, Parkville, Australia
Udo Steigerwald: Zentrallabor, Zentrum Innere Medizin – A4, Universitätsklinikum Würzburg, Würzburg, Germany
Edelgard Lindhoff-Last: Coagulation Research Center CCB (Cardioangiologisches Centrum Bethanien), Frankfurt am Main, Germany

and monoclonal antibodies such as HCAL (IgG) [11] and EY2C9 (IgM) [12]. Neither of these has been fully accepted in clinical practice. Beyond that, two candidate aβ2-GPI reference materials have been suggested recently: a polyclonal reference consisting of pooled, affinity-purified IgG and IgM from two APS patients and a monoclonal reference, essentially being protein G-purified HCAL [13, 14]. They are, however, not available yet and certainly require further laborious validation. Finally, recommendations on aPL testing were published repeatedly [15–17] in an elaborate, but unfortunately only limited successful attempt to provide a common frame for researchers, manufacturers, and end-users.

Considering the disproportion of effort invested and progress made, resignation on standardization of aPL testing can be observed [18, 19]. Extreme clonal heterogeneity of aPL [20] may actually render assay standardization impossible [21]. Nevertheless, many new technologies for aPL detection emerged in the past years [22]. Irrespective of discussions on the proper reference material and standardization, laboratories and clinicians demand performance evaluations of these emerging tests.

Social sciences have been familiar with the concept of latent class analysis (LCA) since the 1960s [23]. In recent years, LCA also entered medical research [24]. A joint probabilistic model is assumed for the relationship between observed multivariate categorical data and an unobserved distinct class variable. The parameters of this model can be estimated by maximizing the likelihood function. Currently, the applications of LCA in laboratory medicine are primarily focused on the subfield of microbiology. Diagnostic strategies for autoimmune diseases, meanwhile, also suggest themselves as excellent candidate for LCA. This statistical approach has been proven as a versatile tool to assess diagnostic accuracy in situations without "gold standard", but numerous imperfect tests [25, 26]. Additionally, autoimmune disorders are commonly diagnosed dependent on the fulfillment of a minimum number of predefined (clinical and serological) criteria. Yet, surprisingly, LCA has only been used for analysis in a limited number of autoimmunity trials [27, 28].

In the original form, LCA relates the observed outcomes of k tests $\{T_1, \ldots, T_k\}$ on each of i=1, …, n subjects to an unobserved or latent binary variable D. In this work, we use D to model the presence or absence of APS antibodies. $P(T_k|D=1)$ and $P(T_k|D=0)$ are the probabilities that outcome T_k is observed, given the presence or the absence of the latent variable. Given the prevalence $p=P(D=1)$, the combined probability of an outcome T_k is $pP(T_k|D=1)+(1-p)P(T_k|D=0)$. Let the true-positive rate $tp_k=P(T_{ik}=1|D=1)$ and the false-positive rate (fpr) $fp_k=P(T_{ik}=1|D=0)$ be the parameters of the tests $\theta=\{(tp_k, fp_k), k=1, \ldots, K\}$. If the test variables are statistically independent given D, the classic LCA model can be expressed as the likelihood function

$$L(\theta, p)=\prod_{i=1}^{n}\left\{p\prod_{k=1}^{K}P(T_{ik}|D=1)+(1-p)\prod_{k=1}^{K}P(T_{ik}|D=0)\right\}.$$

With a minimum of k=3 different tests, the parameters can subsequently be estimated by maximizing $L(\theta, p)$ [26]. Now the not directly observable state of each sample (in our case antibody positive or negative) that corresponds most likely with the observed outcomes is assigned to variable D. Based on this information characteristics of new tests can be estimated. The basic concept of LCA is depicted step by step in a small and clear sample data set in the Supplemental Data.

The aim of our trial is to demonstrate that the statistical concept of LCA offers a feasible approach to address the absence of reference materials and standardization problems in aPL testing. Exemplified with the SPR biosensor research assays mentioned above, we show how diagnostic performance can be statistically evaluated without a commonly agreed on reference standard.

Materials and methods

Patients

Sera from 63 APS patients were collected in the University Hospitals of Würzburg, Frankfurt, and the TU München. All patients satisfied the Sydney criteria [1] as determined by the respective local work-up. The patient group consisted of 46 women and 17 men with a mean age of 44 years. Forty-two were classified as primary and 18 as secondary APS, whereas 3 could not be assigned clearly to the primary or secondary syndrome groups. The main clinical manifestations were thromboembolic events in 54 (37 venous, 13 arterial, 4 venous and arterial), pregnancy complications in 4 and a combination of both in the remaining 5 subjects. Twenty-five women and nine men with a mean age of 41 years were recruited from the laboratory staff of the University Hospital of the TU München to represent the 34 healthy controls. Patient and control group did not differ significantly with respect to age and sex.

The Institutional Ethics Committees of all three involved hospitals approved the study. Written informed consent was received from all study participants. No financial compensation was provided.

Commercial aPL assays

aCL IgG, aCL IgM, aβ2-GPI IgG, and aβ2-GPI IgM were determined in all sera with the following commercial assays: Alegria (ALE) from Orgentec (Mainz, Germany), ACL AcuStar (ACU) from Instrumentation Laboratory (Kirchheim, Germany), UniCAP 250 (UNI) from Thermo Fisher Scientific (Freiburg, Germany), and AESKULISA (AES)

from Aesku.diagnostics (Wendelsheim, Germany). For detection of aβ2-GPI, human β2-GPI was employed as antigen in all commercial assays, whereas the human (ALE, ACU, AES) as well as the bovine protein (UNI) were applied for aCL detection.

SPR aPL assays

A BIAcore X device (GE Healthcare, Freiburg, Germany) was used for all SPR experiments. Preparation of the aβ2-GPI- and the aCL-surface and serum measurements were performed as described previously [4, 5] with slight modifications for purposes of harmonization.

Human β2-GPI (Scipac, Sittingbourne, UK) or an amine-functionalized cardiolipin derivative synthesized as described in [29] were immobilized on the specific flow cell (FC2) and human transferrin (Sigma-Aldrich, Taufkirchen, Germany) on the reference flow cell (FC1). Binding of diluted serum was plotted in a sensorgram as the signal difference FC2–FC1 and quantified in arbitrary resonance units (RU). Noteworthy, the running buffer for the aCL chip contained fetal calf serum as β2-GPI source to create the generally accepted main target antigen of aCL, i.e. β2-GPI complexed with the anionic cardiolipin.

Additional details on the preparation of the SPR chips, serum measurements, and biochemical properties of the biosensor assays are given in previous publications [4, 5] and in the Supplemental Data.

Statistics

The basic application of LCA requires the observed categorical outcomes to be statistically independent from one another, given the true unobserved variable. Knowledge of one test result must not give any information about other test results from subjects with the same true disease status. For most diagnostic tests detecting similar biological properties this assumption does not hold true [30]. Therefore, a random effects model that relaxes this assumption was employed [31]. All calculations were performed using the software "R", version 3.1 (R Foundation for Statistical Computing, Vienna, Austria) and "randomLCA" version 1.0-1 [32].

LCA was conducted using the results from the commercial assays and the manufacturers' cutoff recommendations (specific IgG or IgM present or absent). Antibody subclasses were evaluated separately. Since the SPR biosensor does not distinguish between Ig subclasses, the constructed gold standard considered a sample as containing antibodies if one of the LCAs yielded a positive result (Figure 1).

The sensorgrams of the SPR biosensors were analyzed for biospecific interaction features allowing a robust discrimination between samples with and without aPL. Maximum RU as well as median and maximum slope of the binding curve were determined. The constructed gold standard was used to determine the area under the curve (AUC) for each feature.

Additional analysis of SPR data was conducted using Microsoft Excel and the Analyse-it statistics software (Analyse-it Software, Leeds, UK). Receiver operating characteristic (ROC) analysis was performed and optimal cutoff values were determined via maximum Youden-index. Sensitivity and specificity data were calculated and compared in the McNemar test for the paired nominal assay data.

Results

Patient classification in the commercial assays

When applying the manufacturers' cutoffs the following numbers of APS patients tested positive in one, two, three or four of the commercial assays: 5, 10, 12, and 30 for aCL IgG; 5, 4, 3, and 12 for aCL IgM; 8, 8, 22, and 14 for aβ2-GPI IgG; 0, 0, 2, and 21 for aβ2-GPI IgM. Controls were almost exclusively found to be negative for all four auto-antibodies in all four assays. Exceptions were merely borderline positive and encompassed one subject positive in the ACU assays for aCL and aβ2-GPI IgG, another one positive in the ACU aβ2-GPI IgG, and a third one positive for aβ2-GPI IgM with the AES system.

Numbers and percentages of samples classified correctly and falsely according to the manufacturers' cutoffs with respect to results of LCA are depicted in Table 1. Between 73% and 100% were classified correctly. ACU exhibits a higher fpr (4%–10%) than false-negative rate (fnr) (0%–7%). Only with aβ2-GPI IgM does the ACU classify all samples correctly. In contrast, UNI displays fnr values of 4%–17%, considerably greater than fpr values of 0%–4% for the detection of aCL. UNI assigns all sera to the correct group with respect to aβ2-GPI, resulting in low fpr and fnr for the presence of any aPL (3% and 1%, respectively). The ALE tests show a higher fpr than fnr for aCL IgM, presence of aCL and aβ2-GPI IgG (1%–8% vs. 0%–4%) and the reverse behavior for presence of aβ2-GPI and of a combination of aβ2-GPI and aCL (fnr 2%–4% vs. fpr 1%). No misclassifications were observed with the ALE for aβ2-GPI IgG and similar to the two assays discussed above for aβ2-GPI IgM. AES proved to be the only assay to classify all samples correctly for aCL IgM. However, AES exhibited higher fpr (6%) than fnr (1%–2%) for detection of aCL IgG, presence of aCL and presence of any aPL along with higher fnr (2%–27%) than fpr (0%–1%) for aβ2-GPI IgG and IgM and presence of aβ2-GPI.

Cutoff values for the SPR assays

The binding behaviors of sera in the SPR assays showed distinct differences according to presence or absence of aPL, as concluded from LCA (Figure 2). Sera classified to contain aCL exhibit maximum bindings to the aCL surface ranging from 263 to 4809 RU. Sera without aCL showed considerably less binding between 148 RU and 543 RU. In the aβ2-GPI SPR assay, aβ2-GPI-positive sera bind with

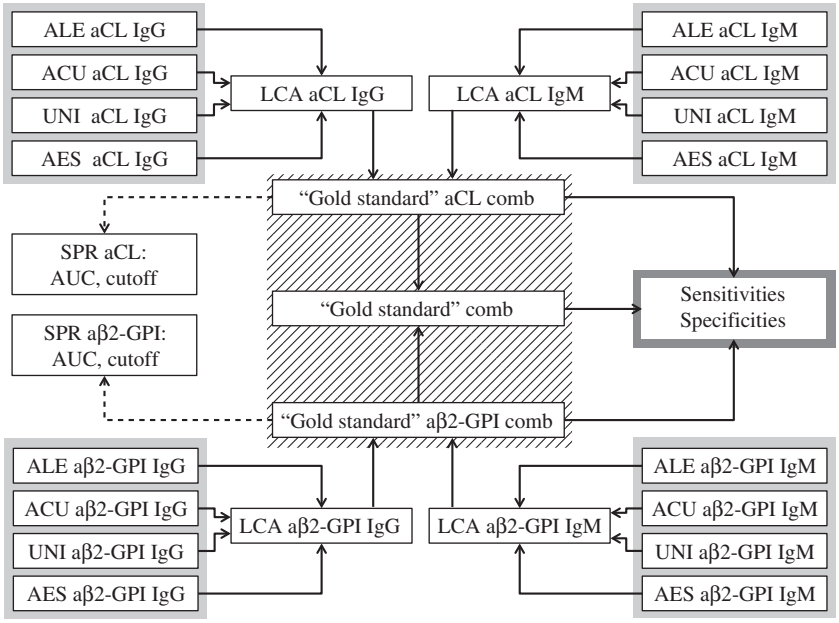

Figure 1: Construction of the "gold standards".
Process of constructing the three different "gold standards" aCL comb (for presence of aCL), aβ2-GPI comb (for presence of aβ2-GPI), and comb (for presence of any aPL) from LCA and of the subsequent calculation of cutoff values for the SPR assays and of sensitivities and specificities for all investigated assays.

Table 1: Correctly and falsely classified samples.

	aCL IgG		aCL IgM		aCL comb		aβ2-GPI IgG		aβ2-GPI IgM		aβ2-GPI comb		comb	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
ALE														
LCA correct	97	100.0	90	92.8	96	99.0	85	87.6	97	100.0	92	94.8	94	96.9
LCA false positive	0	0.0	7	7.2	1	1.0	8	8.2	0	0.0	1	1.0	1	1.0
LCA false negative	0	0.0	0	0.0	0	0.0	4	4.1	0	0.0	4	4.1	2	2.1
ACU														
LCA correct	81	83.5	92	94.8	85	87.6	87	89.7	97	100.0	90	92.8	86	88.7
LCA false positive	9	9.3	4	4.1	10	10.3	10	10.3	0	0.0	7	7.2	9	9.3
LCA false negative	7	7.2	1	1.0	2	2.1	0	0.0	0	0.0	0	0.0	2	2.1
UNI														
LCA correct	77	79.4	93	95.9	83	85.6	97	100.0	97	100.0	97	100.0	93	95.9
LCA false positive	4	4.1	0	0.0	4	4.1	0	0.0	0	0.0	0	0.0	3	3.1
LCA false negative	16	16.5	4	4.1	10	10.3	0	0.0	0	0.0	0	0.0	1	1.0
AES														
LCA correct	89	91.8	97	100.0	90	92.8	71	73.2	94	96.9	74	76.3	90	92.8
LCA false positive	6	6.2	0	0.0	6	6.2	0	0.0	1	1.0	1	1.0	6	6.2
LCA false negative	2	2.1	0	0.0	1	1.0	26	26.8	2	2.1	22	22.7	1	1.0
SPR														
LCA correct	–	–	–	–	90	93.8	–	–	–	–	85	87.6	89	92.7
LCA false positive	–	–	–	–	0	0.0	–	–	–	–	2	2.1	2	2.1
LCA false negative	–	–	–	–	6	6.3	–	–	–	–	10	10.3	5	5.2

Numbers and percentages of correctly and falsely classified samples (according to LCA results) when applying the manufacturers cutoff for the ALE, ACU, UNI, and AES and the LCA-derived cutoff for the SPR assay. aCL comb, presence of aCL; aβ2-GPI comb, presence of aβ2-GPI; comb, presence of any aPL.

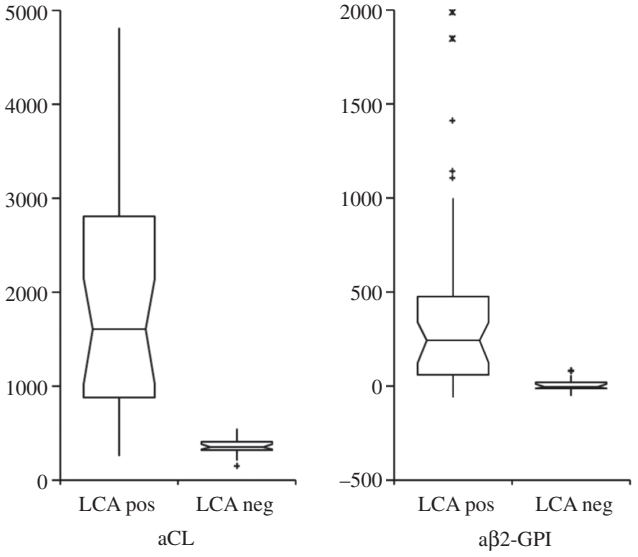

Figure 2: Box and whisker plots of the SPR assay results.
Box and whisker plots of the maximum binding in the aCL and the aβ2-GPI SPR assay for the investigated sera. Sera were assigned to the respective antibody positive or negative group according to the results of LCA (y-axis: maximum binding [Δ RU]).

maxima of up to 1987 RU, whereas the aβ2-GPI-negative binding maxima do not exceed 83 RU.

Antigen binding behaviors of samples did not differ between patients suffering from arterial or venous thromboses or obstetric complications. Maximum binding proved to be the best indicator for presence of the respective antibody.

ROC analysis yielded an AUC of 0.96 for the aCL and of 0.89 for the aβ2-GPI surface (Figure 3). Both AUCs were significantly higher than 0.5 ($p<0.0001$). Maximum Youden indices were determined at 543 RU and 56 RU for the detection of aCL and aβ2-GPI, respectively. Patients with higher maximum binding were assigned to the positive group for the respective antibody in the SPR biosensor assays.

Applying the determined cutoff values, the SPR assays allocate 94%, 88%, and 93% of the samples correctly regarding the presence or absence of aCL, aβ2-GPI, or any aPL. The fnr are clearly higher than the fpr (5%–10% vs. 0%–2%, Table 1).

Comparison of sensitivities and specificities

Sensitivities and specificities of all five investigated assays, as compared to LCA as "gold standard", are summarized in Table 2. p-Values for their comparison are given in Table 3.

Among the commercial systems, significantly lower sensitivities were calculated for the UNI and AES system (0.769 and 0.551) regarding detection of aCL and aβ2-GPI, respectively. No significant differences could be demonstrated in the sensitivities for the presence of any aPL. Concerning the specificities, the ACU system with values from 0.792 to 0.854 performed significantly worse than the ALE for the detection of aCL, aβ2-GPI, and of any aPL, and also than the UNI system for aβ2-GPI.

Specificities of the SPR assays for aCL, aβ2-GPI, and any aPL were 1.000, 0.958, and 0.956, respectively, and significantly higher than those of the ACU for aCL (0.792) and any aPL (0.804) and than that of the AES for aCL (0.875). In the remaining cases, the research sensors specificities did not differ significantly from the commercial assays. The formers' sensitivity for aCL of 0.878 proved to

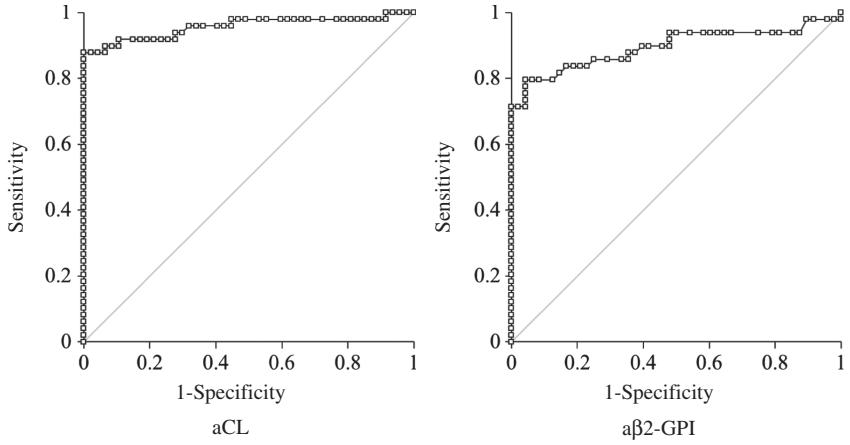

Figure 3: ROC analysis of the SPR assay results.
ROC curves for aCL and aβ2-GPI SPR assays when using LCA results as gold standard.

Table 2: Sensitivities and specificities of the investigated assays.

"Gold standard"	aCL.comb (from LCA)			aβ2-GPI.comb (from LCA)			comb (from LCA)		
	aCL	Lower 95% CI	Upper 95% CI	aβ2-GPI	Lower 95% CI	Upper 95% CI	aCL+aβ2-GPI	Lower 95% CI	Upper 95% CI
Sensitivities									
ALE	1.000	0.927	1.000	0.918	0.804	0.977	0.961	0.865	0.995
ACU	0.959	0.860	0.995	1.000	0.927	1.000	0.961	0.865	0.995
UNI	0.796	0.657	0.898	1.000	0.927	1.000	0.980	0.896	1.000
AES	0.980	0.891	0.999	0.551	0.402	0.693	0.980	0.896	1.000
SPR	0.878	0.752	0.954	0.796	0.657	0.898	0.902	0.786	0.967
Specificities									
ALE	0.979	0.889	0.999	0.979	0.889	0.999	0.978	0.885	0.999
ACU	0.792	0.650	0.895	0.854	0.722	0.939	0.804	0.661	0.906
UNI	0.917	0.800	0.977	1.000	0.926	1.000	0.935	0.821	0.986
AES	0.875	0.748	0.953	0.979	0.889	0.999	0.870	0.737	0.951
SPR	1.000	0.925	1.000	0.958	0.857	0.995	0.956	0.849	0.995

Results of LCA served as "gold standard". Classification of sera according to the manufacturers' information (ALE, ACU, UNI, AES) or the determined cutoff values (SPR assay). CI, confidence interval; aCL comb, presence of aCL; aβ2-GPI comb, presence of aβ2-GPI; comb, presence of any aPL.

Table 3: Comparison of sensitivities and specificities.

	ALE	ACU	UNI	AES	SPR
aCL					
ALE		0.5000	0.0020[a]	1.0000	0.0313[a]
ACU	0.0039[a]		0.0215[a]	1.0000	0.2891
UNI	0.3750	0.0703		0.0039[a]	0.3877
AES	0.0625	0.2188	0.6875		0.0625
SPR	1.0000	0.0020[a]	0.1250	0.0313[a]	
aβ2-GPI					
ALE		0.1250	0.1250	<0.0001[a]	0.1094
ACU	0.0313[a]		1.0000	<0.0001[a]	0.0020[a]
UNI	1.0000	0.0156[a]		<0.0001[a]	0.0020[a]
AES	1.0000	0.0703	1.0000		0.0118[a]
SPR	1.0000	0.1250	0.5000	1.0000	
aCL+aβ2-GPI					
ALE		1.0000	1.0000	1.0000	0.2500
ACU	0.0078[a]		1.0000	1.0000	0.4531
UNI	0.6250	0.0703		1.0000	0.2188
AES	0.0625	0.4531	0.4531		0.1250
SPR	1.0000	0.0391[a]	1.0000	0.2188	

p-Values for the comparison of the sensitivities (no shading) and specificities (gray shading) of the investigated assays. [a]Values with $p<0.05$.

be significantly lower than the ALE (1.000), but not significantly different from the ACU, UNI and AES (0.959, 0.796, and 0.980, respectively). For the detection of aβ2-GPI, a sensitivity of 0.796 was calculated for the SPR sensor, which was significantly lower than two (ACU and UNI, both 1.000), not significantly different from one (ALE, 0.918) and significantly higher than one of the commercial systems (AES, 0.551). With 0.902, the SPR assays sensitivity for the presence of any aPL did not differ significantly from any of the other assays.

Discussion

A novel approach to evaluate the diagnostic performance of aPL assays by use of LCA has been shown. LCA was employed to deduce a supposed gold standard from the measurement results of four commercial aPL assays to assess performance data of two SPR biosensor assays for aCL and aβ2-GPI.

The research SPR assays discriminated well between antibody-positive and antibody-negative sera with AUCs of 0.96 and 0.89 for the detection of aCL and aβ2-GPI. Binding maxima for the aCL chip were higher than for the aβ2-GPI chip as a result of differing serum dilutions and antigen densities each being optimized for the specific surface. SPR biosensors detect mass changes at the chip surface. We demonstrated in previous works that our surface design enables the selective detection of immunoglobulins [33]. The current assay setup, however, does not allow to distinguish among IgG, IgM, and IgA. Hence, the final comparison of sensitivities and specificities must be performed independently of the antibody isotype. Generally, the diagnostic performances of the investigated commercial assays were fairly comparable. The UNI and AES systems identify aCL and aβ2-GPI less sensitively, respectively. Furthermore, the ACU system has a slight weakness with respect to the specific detection of aPL. Specificities of the SPR sensors were non-inferior to all and possibly superior to some

commercial assays. With respect to sensitivity, the biosensor performed equally or better than at least two of the commercial assays. Hence, the diagnostic performances of the biosensor assays are comparable to commercially available systems. Remarkably, β2-GPI originating from different species (human, bovine) in the various aCL assays as potential source of variability did not impair comparability of the investigated systems. Furthermore, the ability to detect IgA aPL did not adversely affect the biosensor systems diagnostic performances. This may be due to the investigated patient cohort solely encompassing subjects with well-established diagnoses of APS, where IgA aPL should play a minor role as compared to IgG or IgM.

The SPR assays enabled reproducible measurements of aPL for more than 15 measurement cycles with excellent intra- and inter-assay variations [4, 5]. The sensorgrams additionally yield insights into the binding kinetics of each individual serum – a type of data not obtainable with common aPL immunoassay designs. Previous studies demonstrated that aβ2-GPI IgG avidity for β2-GPI critically determines the pathogenic potential [34, 35]. Even if we were not able to confirm aforementioned data, SPR assays readily offer themselves as suitable and convenient tool for aPL research. Potential areas of application encompass untangling the complex relationships of aPL avidity and clinical significance. Furthermore, they may enable a better characterization and possibly subclassification of APS patients with respect to clinical presentation and disease severity (as discussed, e.g. in [36, 37]).

Evaluation of the diagnostic performance of aPL tests is a challenge developers have been facing ever since introduction of these assays into patient care. The common solution to the lack of a reference method and of definitive reference material, was to select an already established in-house [22, 38, 39] or commercial [40] aPL assay as alternative reference standard. Alternatively, clinical features of APS served as "surrogate" reference standards [41–44]. Other authors explored both approaches in the same work [45] or only gave vague information on the data, used to assign individuals to the APS or control group [46, 47]. Whereas the latter should be avoided for reasons of documentation and reproducibility, both of the previously mentioned alternative reference standards are far from ideal. Definitive assessment whether the aPL assay used as comparative method measures the analyte accurately or whether the assay to be evaluated is closer to the truth, is impossible. Thromboses and pregnancy complications can be caused by a plethora of other disease conditions, limiting their utility as reference standard.

In the absence of an alternative reference standard, several other approaches were proposed [48]. One possibility is the use of a composite reference standard: several variables are combined according to predefined rules to assign subjects to the patient or control group. With respect to APS, one might consider using the Sydney classification [1]. This classification, however, includes positivity of aPL as criteria itself and therefore is not suitable. Another possible approach would be to gather an expert panel, in which a group of specialists decides on the presence or absence of disease in each individual patient. Collection of a complete medical history, results of clinical examinations, laboratory and other tests, and ideally clinical follow-up is a laborious task. Additionally, different views within the panel group might emerge. We therefore decided to use LCA to evaluate the diagnostic performance of our research SPR aPL assays.

To the best of our knowledge, the presented work constitutes the first application of LCA to APS and aPL testing. LCA assigns the observed test results to unknown, i.e. latent classes of subjects. It is important to keep in mind, that in case of the presented work, not "APS" and "healthy", but "antibody-positive" and "antibody-negative" represent the latent classes. Only inferences on the diagnostic performances of the assays are valid. Conclusions on the clinical significance cannot be drawn. Our work should not be considered as novel suggestion for patient and risk classification (such as the concept of single, double, and triple positivity in aPL assays [49]), but rather as prototypic approach for the evaluation of the analytical performance of aPL assays. One pivotal prerequisite of LCA is that the classifying test results must only be influenced by the latent classes. Non-consideration of confounding classes can result in invalid conclusions. For instance, ethnical disparities in the studied population at the collection sites or different treatment of the samples are not likely, but cannot be excluded and may influence antibody measurements. LCA defines presence of a disease condition as a statistical-based entity [48]. Relevance of the results for daily practice may seem blurred and clinical interpretation challenging. At the same time, LCA provides an objective assessment of the strengths of relationships of variables with statistically well-defined methods. For detection of aPL, a pure clinical definition is not applicable. An unequivocal biochemical definition does not currently and may even never exist [21]. Hence, advantages of LCA by far outweigh its disadvantages. LCA yields the best possible statistical estimation of the target condition (i.e. antibody-positive vs. antibody-negative) for the observed pattern of aPL measurement results when validity of the different methods cannot be conclusively assessed. Possible field of application arising from our work encompasses the LCA as tool for comparing already

existing as well as evaluating new aPL tests. As such, this may help to improve the insufficient comparability of aPL assays. Without doubt, a reference standard would be the ideal tool for these purposes. We therefore have to emphasize that our work certainly was not undertaken to deny the need for an aPL reference standard. It was, however, intended to bring an innovative approach into the discussion which may help to improve the current unsatisfactory situation until a well-characterized, stable, and generally accepted reference standard is available.

In summary, we used LCA to evaluate the diagnostic performance of two research SPR aPL assays. We demonstrated that the SPR assays are suitable for the detection of aCL and aβ2-GPI and non-inferior to common commercial systems. Beyond that, LCA as a statistically sound concept allowed us to bypass discussions about the proper reference method and material and to calculate sensitivity and specificity of the new assays. Future work might refine and extend the presented data by using locally established reference ranges, considering quantitative results, including a higher number of immunoassays, or by application to other autoimmune diseases.

Acknowledgments: We thank Ms. Anita Schreiegg for excellent technical assistance, Dr. Evangeline Thaler for reviewing the manuscript, and the Australian Research Council (ARC) Discovery Project Grant (A.B.H., MXY DP1094497).
Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Miyakis S, Lockshin MD, Atsumi T, Branch DW, Brey RL, Cervera R, et al. International consensus statement on an update of the classification criteria for definite antiphospholipid syndrome (APS). J Thromb Haemost 2006;4:295–306.
2. Hughes GR. Thrombosis, abortion, cerebral disease, and the lupus anticoagulant. Br Med J (Clin Res Ed) 1983;287:1088–9.
3. Devreese K, Hoylaerts MF. Challenges in the diagnosis of the antiphospholipid syndrome. Clin Chem 2010;56:930–40.
4. Metzger J, von Landenberg P, Kehrel M, Buhl A, Lackner KJ, Luppa PB. Biosensor analysis of beta2-glycoprotein I-reactive autoantibodies: evidence for isotype-specific binding and differentiation of pathogenic from infection-induced antibodies. Clin Chem 2007;53:1137–43.
5. Schlichtiger A, Baier C, Yin MX, Holmes AB, Maruyama M, Strasser R, et al. Covalent attachment of functionalized cardiolipin on a biosensor gold surface allows repetitive measurements of anticardiolipin antibodies in serum. Anal Bioanal Chem 2013;405:275–85.
6. Favaloro EJ, Silvestrini R, Mohammed A. Clinical utility of anticardiolipin antibody assays: high inter-laboratory variation and limited consensus by participants of external quality assurance programs signals a cautious approach. Pathology 1999;31:142–7.
7. Favaloro EJ, Wheatland L, Jovanovich S, Roberts-Thomson P, Wong RC. Internal quality control and external quality assurance in testing for antiphospholipid antibodies: part I – Anticardiolipin and anti-beta2-glycoprotein I antibodies. Semin Thromb Hemost 2012;38:390–403.
8. Tincani A, Allegri F, Sanmarco M, Cinquini M, Taglietti M, Balestrieri G, et al. Anticardiolipin antibody assay: a methodological analysis for a better consensus in routine determinations – a cooperative project of the European Antiphospholipid Forum. Thromb Haemost 2001;86:575–83.
9. Reber G, Schousboe I, Tincani A, Sanmarco M, Kveder T, de Moerloose P, et al. Inter-laboratory variability of anti-beta2-glycoprotein I measurement. A collaborative study in the frame of the European Forum on Antiphospholipid Antibodies Standardization Group. Thromb Haemost 2002;88:66–73.
10. Harris EN, Gharavi AE, Patel SP, Hughes GR. Evaluation of the anti-cardiolipin antibody test: report of an international workshop held 4 April 1986. Clin Exp Immunol 1987;68:215–22.
11. Ichikawa K, Tsutsumi A, Atsumi T, Matsuura E, Kobayashi S, Hughes GR, et al. A chimeric antibody with the human gamma1 constant region as a putative standard for assays to detect IgG beta2-glycoprotein I-dependent anticardiolipin and anti-beta2-glycoprotein I antibodies. Arthritis Rheum 1999;42:2461–70.
12. Ichikawa K, Khamashta MA, Koike T, Matsuura E, Hughes GR. β2-Glycoprotein I reactivity of monoclonal anticardiolipin antibodies from patients with the antiphospholipid syndrome. Arthritis Rheum 1994;37:1453–61.
13. Willis R, Grossi C, Orietta Borghi M, Martos-Sevilla G, Zegers I, Sheldon J, et al. International standards for IgG and IgM anti-beta2glycoprotein antibody measurement. Lupus 2014;23:1317–9.
14. Meroni PL, Biggioggero M, Pierangeli SS, Sheldon J, Zegers I, Borghi MO. Standardization of autoantibody testing: a paradigm for serology in rheumatic diseases. Nat Rev Rheumatol 2014;10:35–43.
15. Pierangeli SS, Favaloro EJ, Lakos G, Meroni PL, Tincani A, Wong RC, et al. Standards and reference materials for the anticardiolipin and anti-beta2glycoprotein I assays: a report of recommendations from the APL Task Force at the 13th International Congress on Antiphospholipid Antibodies. Clin Chim Acta 2012;413:358–60.
16. Lakos G, Favaloro EJ, Harris EN, Meroni PL, Tincani A, Wong RC, et al. International consensus guidelines on anticardiolipin and anti-beta2-glycoprotein I testing: report from the 13th International Congress on Antiphospholipid Antibodies. Arthritis Rheum 2012;64:1–10.
17. Bertolaccini ML, Amengual O, Andreoli L, Atsumi T, Chighizola CB, Forastiero R, et al. 14th International Congress on Antiphospholipid Antibodies Task Force. Report on antiphospholipid syndrome laboratory diagnostics and trends. Autoimmun Rev 2014;13:917–30.

18. de Groot PG, Derksen RH, de Laat B. Twenty-two years of failure to set up undisputed assays to detect patients with the antiphospholipid syndrome. Semin Thromb Hemost 2008;34:347–55.
19. Devreese KM. Standardization of antiphospholipid antibody assays. Where do we stand? Lupus 2012;21:718–21.
20. Lieby P, Soley A, Levallois H, Hugel B, Freyssinet JM, Cerutti M, et al. The clonal analysis of anticardiolipin antibodies in a single patient with primary antiphospholipid syndrome reveals an extreme antibody heterogeneity. Blood 2001;97:3820–8.
21. Reber G, Boehlen F, de Moerloose P. Technical aspects in laboratory testing for antiphospholipid antibodies: is standardization an impossible dream? Semin Thromb Hemost 2008;34:340–6.
22. Andreoli L, Rizzini S, Allegri F, Meroni P, Tincani A. Are the current attempts at standardization of antiphospholipid antibodies still useful? Emerging technologies signal a shift in direction. Semin Thromb Hemost 2008;34:356–60.
23. Lazarsfeld PF, Henry NW. Latent structure analysis. Boston: Houghton Mill, 1968.
24. Formann AK, Kohlmann T. Latent class analysis in medical research. Stat Methods Med Res 1996;5:179–211.
25. Goetghebeur E, Liinev J, Boelaert M, Van der Stuyft P. Diagnostic test analyses in search of their gold standard: latent class analyses with random effects. Stat Methods Med Res 2000;9:231–48.
26. Pepe MS, Janes H. Insights into latent class analysis of diagnostic test performance. Biostatistics 2007;8:474–84.
27. Taylor WJ, Marchesoni A, Arreghini M, Sokoll K, Helliwell PS. A comparison of the performance characteristics of classification criteria for the diagnosis of psoriatic arthritis. Semin Arthritis Rheum 2004;34:575–84.
28. Grayson PC, Maksimowicz-McKinnon K, Clark TM, Tomasson G, Cuthbertson D, Carette S, et al. Distribution of arterial lesions in Takayasu's arteritis and giant cell arteritis. Ann Rheum Dis 2012;71:1329–34.
29. Johns MK, Yin MX, Conway SJ, Robinson DE, Wong LS, Bamert R, et al. Synthesis and biological evaluation of a novel cardiolipin affinity matrix. Org Biomol Chem 2009;7:3691–7.
30. Menten J, Boelaert M, Lesaffre E. Bayesian latent class models with conditionally dependent diagnostic tests: a case study. Stat Med 2008;27:4469–88.
31. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. Biometrics 1996;52:797–810.
32. Beath K. randomLCA: Random effects latent class analysis. R package version 1.0-1 2011 [cited 2014]. Available from: http://CRAN.R-project.org/package=randomLCA.
33. Müller C, Schlichtiger A, Balling G, Steigerwald U, Luppa PB, Thaler M. Standardized antigen preparation to achieve comparability of anti-beta2-glycoprotein I assays. Thromb Res 2010;126:e102–9.
34. Cucnik S, Kveder T, Ulcova-Gallova Z, Swadzba J, Musial J, Valesini G, et al. The avidity of anti-beta2-glycoprotein I antibodies in patients with or without antiphospholipid syndrome: a collaborative study in the frame of the European forum on antiphospholipid antibodies. Lupus 2011;20:1166–71.
35. Cucnik S, Kveder T, Artenjak A, Ulcova Gallova Z, Swadzba J, Musial J, et al. Avidity of anti-beta2-glycoprotein I antibodies in patients with antiphospholipid syndrome. Lupus 2012;21:764–5.
36. Harris EN, Pierangeli SS. Primary, secondary, and catastrophic antiphospholipid syndrome: what's in a name? Semin Thromb Hemost 2008;34:219–26.
37. Favaloro EJ, Wong RC. The antiphospholipid syndrome: a large elephant with many parts or an elusive chameleon disguised by many colours? Auto Immun Highlights 2010;1:5–14.
38. de Moerloose P, Reber G, Musial J, Arnout J. Analytical and clinical performance of a new, automated assay panel for the diagnosis of antiphospholipid syndrome. J Thromb Haemost 2010;8:1540–6.
39. Forastiero R, Papalardo E, Watkins M, Nguyen H, Quirbach C, Jaskal K, et al. Evaluation of different immunoassays for the detection of antiphospholipid antibodies: report of a wet workshop during the 13th International Congress on Antiphospholipid Antibodies. Clin Chim Acta 2014;428:99–105.
40. Noubouossie D, Valsamis J, Corazza F, Rozen L, Debaugnies F, Demulder A. An automated chemiluminescence immunoassay may detect mostly relevant IgG anticardiolipin antibodies according to revised Sydney criteria. Acta Clin Belg 2012;67:184–9.
41. Vikerfors A, Johansson AB, Gustafsson JT, Jonsen A, Leonard D, Zickert A, et al. Clinical manifestations and anti-phospholipid antibodies in 712 patients with systemic lupus erythematosus: evaluation of two diagnostic assays. Rheumatology (Oxford) 2013;52:501–9.
42. Decavele AS, Schouwers S, Devreese KM. Evaluation of three commercial ELISA kits for anticardiolipin and anti-beta2-glycoprotein I antibodies in the laboratory diagnosis of the antiphospholipid syndrome. Int J Lab Hematol 2011;33:97–108.
43. Chung Y, Kim JE, Lim HS, Kim HK. Clinical performance of anticardiolipin and antibeta2 glycoprotein I antibodies using a new automated chemiluminescent assay: superior thrombotic prediction of combined results measured by two different methods. Blood Coagul Fibrinolysis 2014;25:10–5.
44. Li R, Daguzan M, Vandermijnsbrugge F, Gyling M, Cantinieaux B. Both IgG and IgM anti-beta2 glycoprotein I antibodies assays are clinically useful to the antiphospholipid syndrome diagnosis. Acta Clin Belg 2014;69:433–8.
45. Persijn L, Decavele AS, Schouwers S, Devreese K. Evaluation of a new set of automated chemiluminescense assays for anticardiolipin and anti-beta2-glycoprotein I antibodies in the laboratory diagnosis of the antiphospholipid syndrome. Thromb Res 2011;128:565–9.
46. Van Hoecke F, Persijn L, Decavele AS, Devreese K. Performance of two new, automated chemiluminescence assay panels for anticardiolipin and anti-beta2-glycoprotein I antibodies in the laboratory diagnosis of the antiphospholipid syndrome. Int J Lab Hematol 2012;34:630–40.
47. Villalta D, Alessio MG, Tampoia M, Da Re A, Stella S, Da Re M, et al. Accuracy of the first fully automated method for anticardiolipin and anti-beta2 glycoprotein I antibody detection for the diagnosis of antiphospholipid syndrome. Ann NY Acad Sci 2009;1173:21–7.
48. Reitsma JB, Rutjes AW, Khan KS, Coomarasamy A, Bossuyt PM. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. J Clin Epidemiol 2009;62:797–806.
49. Forastiero R. Multiple antiphospholipid antibodies positivity and antiphospholipid syndrome criteria re-evaluation. Lupus 2014;23:1252–4.

Supplemental Material: The online version of this article (DOI: 10.1515/cclm-2016-0116) offers supplementary material, available to authorized users.