BMJ Open

Evaluation of seven recombinant VCA-IgA ELISA kits for the diagnosis of nasopharyngeal carcinoma in China: a case-control trial

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-013211
Article Type:	Research
Date Submitted by the Author:	27-Jun-2016
Complete List of Authors:	Gao, Rui; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Wang, Lin; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Nasopharyngeal Carcinoma Zhang, Lifang; Guangzhou Women and Children\'s Medical Center, Division of Birth Cohort Study Ye, Yanfang; Sun Yat-Sen University, Sun Yat-sen Memorial Hospital Xie, Shanghang; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention Du, Jinlin; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Chen, Suihong; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Guo, Jie; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Yang, Mengjie ; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Lin, Chuyang; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Liu, Qing; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention Cao, Sumei; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Diagnostics, Epidemiology, Research methods
Keywords:	Nasopharyngeal carcinoma, Epstein-Barr virus, VCA-IgA, Diagnostic effect, Screening

Evaluation of seven recombinant VCA-IgA ELISA kits for the diagnosis of nasopharyngeal carcinoma in China: a case-control trial

Rui Gao[1,2*], Lin Wang[3*], Li-Fang Zhang[4] ,Yan-Fang Ye[5], Shang-Hang Xie[1], Jin-Lin Du[1,2], Sui-Hong Chen[1,2], Jie Guo[1,2], Meng-Jie Yang[1,2], Chu-yang Lin[1,2], Qing liu[1#], Su-Mei Cao[1#]

[1]Department of Cancer Prevention, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Guangzhou, 510060, P.R. China;
[2]School of Public Health, Sun Yat-Sen University, Guangzhou, 510060, P.R. China;
[3]Department of Nasopharyngeal Carcinoma, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Guangzhou, 510060, P.R. China;
[4]Division of Birth Cohort Study, Guangzhou Women and Children's Medical Center, Guangzhou, 510623, P.R. China;
[5]Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, 510120, P.R. China.
*These authors contributed equally to this work.

[#]Correspondence to
Prof. Su-Mei Cao, caosm@sysucc.org.cn;
Prof. Qing Liu, liuqing@sysucc.org.cn

Authors' E-mail:
Rui Gao: gaor5@mail2.sysu.edu.cn Lin Wang: wangl1@sysucc.org.cn
Li-Fang Zhang: lifanglydia@hotmail.com Yan-Fang Ye: yeyanfang1@gmail.com
Shang-Hang Xie: xieshh@sysucc.org.cn Jin-Lin Du: dujl3@mail2.sysu.edu.cn
Sui-Hong Chen: chensh43@mail2.sysu.edu.cn Jie Guo: guojiegina@126.com
Meng-Jie Yang: yangmj@sysucc.org.cn Chu-yang Lin: linchy23@mail2.sysu.edu.cn

Preferred reviewers:
1. Fang Fang, Fang.Fang@ki.se, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
2. Liming Liang, lliang@hsph.harvard.edu, Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA, US.
3. Qing Chen, qingchen@fimmu.com, Department of Epidemiology, School of Public Health and Tropical Medicine, Southern Medical University, Guangzhou, China

Key words: Nasopharyngeal carcinoma, Epstein-Barr virus, VCA-IgA, Diagnostic effect, Screening

Word count: 2695

ABSTRACT

Objective: Seven recombinant VCA-IgA ELISA kits are widely used in China, but their diagnostic effects have not been evaluated. In this study, we evaluated whether the diagnostic effects of these kits are similar to those of the standard kit (EUROIMMUN, Lübeck, Germany).

Methods: A diagnostic case-control trial was conducted, with 200 cases of nasopharyngeal carcinoma (NPC) and 200 controls from NPC-endemic areas in southern China. The areas under the curve (AUCs), the sensitivities and the specificities of testing kits were compared with those of the standard kit. The test–retest reliability of each kit was determined by intraclass correlation coefficient (ICC). Their diagnostic accuracy in combination with EBNA1-IgA was also evaluated in logistic models.

Results: Three testing kits – KSB, BB and HA – showed diagnostic accuracy equal to that of the standard kit, with good performance in the AUCs (0.926~0.945), and no significant differences in sensitivity were found between early- and advanced-stage NPCs. ICCs exceeded 0.8. Three new logistic regression models were built, and the AUCs of these models (0.961~0.977) were better than those of the individual VCA-IgA kits. All new models had diagnostic accuracy equal to that of the standard kit. New cutoff values of these three kits and their corresponding combinations for the early detection of and screening for NPC were defined.

Conclusions: Three recombinant VCA-IgA kits – KSB, BB and HA – had diagnostic effects equal to those of the standard kit, and, in combination with EBNA1-IgA in logistic regression models, can be used in future screening for NPC.

Strengths and limitations of this study

- This is the first study to carry out a comprehensive evaluation of recombinant commercial diagnostic VCA-IgA (ELISA) kits in China, and new logistic models combining VCA-IgA with EBNA1-IgA were established.
- New cutoff values for VCA-IgA kits and their corresponding combinations were defined for the early detection of and screening for NPC.
- All cases and controls were from NPC-endemic areas of southern China, and thus these results might not be applicable to other populations.
- Cutoff values for NPC screening by means of the new models described in this study must be verified in prospective mass screening.

INTRODUCTION

Nasopharyngeal carcinoma (NPC) is a common form of squamous-cell carcinoma in southern China and southeastern Asia. The annual incidence rate of NPC in southern China can reach 25 per 100,000, which is about 25-fold higher than in the rest of the world.[1–4] NPC is a complex disease caused by a combination of Epstein-Barr virus (EBV), chronic infection, the environment and host genes in a multi-step process of carcinogenesis, but until now there have been no effective preventive measures.[5–9] Long-term survival rates differ substantially between patients with advanced- (stages III and IV) and early- (stages I and II) stage NPC. Four-year survival rates of

early-stage NPC patients are 96.7% compared with 67.1% for those with advanced-stage NPC.[5] Mass screening has become the most practical method for improved early detection in and overall prognosis of NPC patients in the endemic areas.[10 11]

Serum antibodies against EBV-related antigens, especially immunoglobulin A (IgA) against viral capsid antigen (VCA-IgA), early antigen (EA-IgA), EBV nuclear antigen 1(EBNA1-IgA) and so on, remain elevated for an average of 38 months in the preclinical phase,[9–14] and serological tests for these markers are simple and inexpensive.[15–18] Therefore, since the 1970s, these tests have been used as screening markers for NPC in endemic areas. In our previous study, we evaluated the diagnostic performance of seven commercial EBV-related antibodies by enzyme-linked immunosorbent assay (ELISA) and found EBNA1-IgA (Zhongshan Biotech, China) and VCA-IgA (EUROIMMUN, Lübeck, Germany) ELISA to be the top two seromarkers, with AUCs of 0.95 (95%CI, 0.93–0.97) and 0.94 (95%CI, 0.92–0.97), respectively. We further verified that the combination of VCA-IgA and EBNA1-IgA outperformed any individual EBV seromarkers, with AUC up to 0.97 (95%CI, 0.96, 0.99).[15–17] Thus, since 2011, the combination of VCA-IgA and EBNA1-IgA has been recommended as the standard tool for NPC screening in China.[18]

The EBV capsid antigen (VCA) is a late protein produced in the EBV lytic infection period. VCA contains a batch of capsid proteins, such as VCA-p18 (BFRF3), VCA-p23 (BLRF2), gp125/110 (BALF4) and so on, which have unique immune dominants and virus-specific antigenic domains. These domains contain several small peptide regions (epitopes) which can be combined to form a powerful diagnostic reagent for VCA-IgA.[19 20] The capsid proteins in the EUROIMMUN kit[21] were extracted from the pyrolysis products of human B lymphocytes (P3HR1 cell line) infected by EBV and contained a combined native capsid protein of EBV. In addition, several kinds of commercial VCA-IgA kits based on recombinant peptides have been developed in China and are presently widely used for the early detection of and screening for NPC. However, their diagnostic performance for NPC alone and in combination with EBNA1-IgA has not been evaluated. In this study, we evaluated whether the effects of the NPC-diagnostic kits are comparable with those of the standard VCA-IgA kit and can be substituted for it. If so, we will further explore the combination diagnostic strategy with EBNA1-IgA for the early detection of and mass screening for NPC.

METHODS

Study population

Serum specimens were continuously collected from 200 patients hospitalised with NPC in the Sun Yat-sen University Cancer Center (SYSUCC) from January 2013 to June 2013. These cases were histologically confirmed by biopsy, and the clinical stages were classified according to the 2009 Union for International Cancer Control (UICC) criteria, including 33 patients with early-stage NPC (stages I, II) and

167 with advanced-stage NPC (stages III, IV). The inclusion criteria included being between 30 and 59 years of age and residing in one of the six high-endemic provinces of southern China (Guangdong, Guangxi, Jiangxi, Hunan, Fujian or Hainan Province). Other information, including demographic data, smoking, drinking histories and family history of NPC, was collected by the physician in charge. All serum samples were collected before treatment.

The 200 healthy controls were randomly selected from among healthy people who participated in physical examinations at the Sihui Cancer Center (Sihui City, Guangdong Province, China) from July 2013 to September 2013 and were frequency-matched with cases by age (5-year age groups) and gender. All participants completed a short questionnaire to record demographic data, smoking, drinking histories and family history of NPC and donated 3 mL of blood.

This study was approved by the Clinical Research Ethics Committee of the SYSUCC (YB2015-029-01), and all participants provided written informed consent.

Detection of serological EBV antibodies

Serum and buffy coat were separated less than 4 hours after collection and stored at -80°C before being tested. None of the specimens was hemolytic or repeatedly frozen more than twice. Seven recombinant VCA-IgA kits, the standard VCA-IgA kit (EUROIMMUN) and the standard EBNA1-IgA kit (Zhongshan) were tested (table 1).

Table 1 Product information for eight brands of VCA-IgA kits and the EBNA1-IgA kit

Abbreviation for kits	Manufacturer
VCA-IgA	
KSB	Shenzhen Kang Sheng Bao Bio-Technology Co., Ltd.
BNV	Bioneovan Co., Ltd.
GBI	Beijing BGI-GBI Biotech Co., Ltd.
BB	Beijing Beier Bioengineering Co., Ltd.
HA	Shenzhen HuianBioscitech Co., Ltd.
HK	Shen Zhen HuaKang Co., Ltd.
ZS	ZhongShan Biotech Co., Ltd.
EUROIMMUN	EUROIMMUN Medizinische Labordiagnostika AG
EBNA1-IgA	ZhongShan Biotech Co., Ltd.

All samples were renumbered and tested blindly by one technician according to the manufacturers' instructions. Levels of antibodies were assessed by photometric measurement, which provided optical density (OD) values. Reference ODs (rOD) were obtained according to manufacturers' instructions by dividing OD values by a reference control. To investigate the test-retest reliability of each kit, 10% serum samples (40 samples) were randomly chosen and retested.

Statistical analysis

Demographic characteristics and NPC risk factors between cases and controls were compared by chi-squared tests. The cutoff value of each single kit was defined with the largest Youden Indices (sensitivity+specificity-1) chosen from each receiver operating characteristic (ROC). The diagnostic efficacy of each kit was evaluated by AUC, and non-inferiority tests based on the bootstrap approach were performed to determine whether the AUCs of these recombinant testing kits were inferior to that of the standard kit (Δ=0.05).[22–25] The sensitivity and specificity of each kit were calculated, and their 95% confidence intervals (CIs) were estimated by the methods of Simel and colleagues.[26] Differences in sensitivities and specificities of the testing kits compared with those of the standard kit and differences in sensitivities between early- and advanced-stage NPC with each kit were compared by chi-squared tests. Intraclass correlation coefficients (ICC) were performed to determine test–retest reliability.

Binary unconditional logistic regressions were used to establish formulae for VCA-IgA and EBNA1-IgA. The diagnostic efficacy of each formula was evaluated by sensitivity, specificity and AUC, compared with the standard formula, *Logit P* = -3.934 + 2.203 VCA-IgA + 4.797 EBNA1-IgA. The cutoff p-value in the corresponding logistic regression for distinguishing between NPC cases and controls was defined with the largest Youden Index chosen from each ROC. Two minimally acceptable false-positive rates (1-Specificity), 3% and 7%, were used empirically to establish the cutoff p-values for classifying different NPC risk subgroups.[16 17]

The non-inferiority tests were one-sided, and p>0.05 was considered to be non-inferior. Other tests were two-sided, and p<0.05 was considered to be statistically significant. Data were analysed by SAS9.2 and SPSS16.0 software.

RESULTS

Baseline information

Baseline information on gender, age, smoking, drinking and NPC family history was comparable between cases and controls, and no statistically significant differences were found between them. Further, there were no statistically significant differences for these items between early- and advanced-stage cases (table 2).

Table 2 Demographic characteristics of NPC cases and controls

Categories	NPC Cases (N_1=200) No. (%)				Controls (N_2=200) No. (%)	p[5]
	Early stage (n=33)	Advanced stage (n=167)	p[4]	Total		
Gender			0.472			0.417
Male	27 (81.8)	127 (76.0)		154 (77.0)	147 (73.5)	
Female	6 (18.2)	40 (24.0)		46 (23.0)	53 (26.5)	

Age (years)			0.299		0.785
30~	6 (18.2)	47 (28.1)		53 (26.5)	47 (23.5)
40~	13 (39.4)	70 (41.9)		83 (41.5)	87 (43.5)
50~	14 (42.4)	50 (29.9)		64 (32.0)	66 (33.0)
Smoking[1]			0.857		0.746
Yes	11 (33.3)	53 (31.7)		64 (32.0)	61 (30.5)
No	22 (66.7)	114 (68.3)		136 (68.0)	139 (69.5)
Drinking[2]			0.641		0.494
Yes	6 (18.2)	25 (15.0)		31 (15.5)	27 (13.5)
No	27 (81.8)	142 (85.0)		169 (84.5)	173 (86.5)
NPC family history[3]			0.732		0.224
Yes	3 (9.1)	13 (7.8)		16 (8.0)	10 (5.0)
No	30 (90.9)	154 (92.2)		184 (92.0)	190 (95.0)

[1]'Smoking' refers to people who smoked more than one cigarette every three days within half a year and included current and former smokers.[2]'Drinking' refers to people who consumed alcoholic beverages every week within half a year and included current and former drinkers. [3]'NPC family history' refers to people whose parents, children and siblings have or did have NPC. [4]Differences in early- and advanced-stage NPC were compared by chi-squared tests. *p<0.05 was considered as significant. [5]Differences in NPC Cases and Controls were compared by chi-squared tests. *p<0.05 was considered as significant.

The diagnostic accuracies of eight brands of VCA-IgA kits and the EBNA1-IgA kit

Table 3 shows that the AUCs of four kits – KSB, BB, BNV and HA – were as high as that of the standard VCA-IgA kit (AUC, 0.942; 95%CI, 0.920-0.964). The AUCs, in order, were 0.945 for KSB (95%CI 0.925-0.966), 0.940 for BB (95%CI 0.916-0.964), 0.936 for BNV (95%CI 0.911-0.961) and 0.926 for HA (95%CI 0.900-0.953). Also, no differences in sensitivities and specificities were found between these four kits and the standard kit. In addition, the AUCs of GBI, HK and ZS were lower than that of the standard kit. Furthermore, no significant differences were found between early- and advanced-stage NPC in the sensitivities of six kits (p>0.05), except for BNV (p=0.044).

Table 3 The diagnostic accuracies of eight brands of VCA-IgA kits and the EBNA1-IgA kit in distinguishing between NPC cases and controls

Kits	Cutoff values[1]	Sensitivity (%)			Specificity (%)	AUC (95%CI)	p[5]
		Early stage (95%CI)	Advanced stage (95%CI)	Average (95%CI)[3]	Control (95% CI)[4]		
VCA-IgA							
BB	0.58	75.8 (69.8-81.7)	88.6 (84.2-93.0)	86.5 (81.8-91.2)	92.0 (88.2-95.8)	0.940 (0.916-0.964)	0.002

BNV	0.923	72.7 (66.6-78.9)	88.0 (83.5-92.5)*	86.0 (81.2-90.8)	93.5 (90.1-96.9)	0.936 (0.911-0.961)	0.003
GBI	0.825	72.7 (66.6-78.9)	76.6 (70.8-82.5)	76.0 (70.1-81.9)*	92.0 (88.2-95.8)	0.899 (0.868-0.930)	0.341*
HA	0.884	93.9 (90.6-97.2	88.0 (83.5-92.5)	89.0 (84.7-93.3)	86.0 (81.2-90.8)	0.926 (0.900-0.953)	0.012
HK	1.218	81.8 (76.5-87.2)	83.2 (78.1-88.4)	83.0 (77.8-88.2)	89.5 (85.3-93.7)	0.913 (0.884-0.942)	0.075*
KSB	0.283	100.0 (100.0-100.0)	86.8 (82.1-91.5)	89.0 (84.7-93.3)	87.5 (82.9-92.1)	0.945 (0.925-0.966)	0.000
ZS	0.418	75.8 (69.8-81.7)	74.3 (68.2-80.3)	74.5 (68.5-80.5)*	87.5 (82.9-92.1)	0.868 (0.831-0.904)	0.878*
EUROIMMUN	1.561	87.9 (83.4-92.4)	85.6 (80.8-90.5)	86.0 (81.2-90.8)	90.0 (85.8-94.2)	0.942 (0.921-0.964)	
EBNA1-IgA	1.203	93.9 (90.6-97.2)	86.2 (81.5-91.0)	87.5 (82.9-92.1)	92.5 (88.8-96.2)	0.956 (0.937-0.975)	0.000

[1]Cutoff value for NPC diagnosis was defined as the value with the largest Youden Index chosen from each ROC. [2]Differences in the sensitivities of early- and advanced-stage NPC were compared by chi-squared tests. *p<0.05 was considered as significant. [3]Differences in the sensitivities of EUROIMMUN and other kits were compared by chi-squared tests. *p<0.05 was considered as significant. [4]Differences in the specificities of EUROIMMUN and other kits were compared by chi-squared tests. *p<0.05 was considered as significant. [5]p values were estimated by non-inferiority tests based on the bootstrap approach for AUC between EUROIMMUN and other kits.*p>0.05 was considered as significant.

The test–retest reliabilities of eight brands of VCA-IgA kits and the EBNA1-IgA kit

Ten percent serum samples (40 samples) were randomly chosen and retested for calculation of the ICC of each brand of kit, VCA-IgA or EBNA1-IgA. The test-retest reliabilities of all kits were excellent (>0.75) according to Fleiss's classification[27] (table 4).

Table 4 The test–retest reliabilities of eight brands of VCA-IgA kits and the EBNA1-IgA kit

Kits	ICC	95%CI
VCA-IgA		
BB	0.990	0.980-0.994
BNV	0.982	0.967-0.991
GBI	0.964	0.933-0.981
HA	0.975	0.952-0.987
HK	0.876	0.764-0.935
KSB	0.823	0.666-0.906
ZS	0.978	0.958-0.988

EUROIMMUN	0.913	0.830-0.955
EBNA1-IgA	0.981	0.964-0.990

The diagnostic accuracies of the combinations of VCA-IgA and EBNA1-IgA with logistic models

We chose for testing three VCA-IgA kits with high AUCs, no differences in diagnoses for early- and later-stage NPC and excellent test-retest reliabilities, and then combined each with the EBNA1-IgA kit by logistic models. Three new logistic regression models were established:

LogitP = -3.2323 + 0.8060VCA-IgA (BB) + 1.1044 EBNA1-IgA

LogitP = -2.7591 + 0.6380VCA-IgA (HA) + 1.0620EBNA1-IgA

LogitP = -2.6039 + 0.5312VCA-IgA (KSB) + 1.1673 EBNA1-IgA

In all these models, both VCA-IgA and EBNA1-IgA were statistically significant independent predictors of NPC risk ($p<0.05$), and the AUC of each combination was statistically significantly larger than that of each single VCA-IgA ($p<0.05$). The AUC of KSB increased from 0.945 (95%CI 0.925-0.966) to 0.964 (95%CI 0.947-0.981); BB increased from 0.940 (95%CI 0.916-0.964) to 0.977 (95%CI 0.963-0.991); and HA increased from 0.926 (95%CI 0.900-0.953) to 0.961 (95%CI 0.943-0.979) (figure 1).

Table 5 shows the diagnostic accuracies of the three new combinations and the standard combination [*Logit P*=-3.934+2.203VCA-IgA (EUROIMMUN) +4.797EBNA1-IgA] in distinguishing between NPC cases and controls. The AUCs of the three new combinations were as high as that of the standard combination (AUC, 0.970; 95%CI, 0.956-0.985) ($p<0.05$). Furthermore, no statistically significant difference was found in the sensitivity of each combination between early- and advanced-stage NPC ($p>0.05$).

We used two minimally acceptable false-positive rates (1-Specificity) of 3% and 7% to define the high-risk and medium-risk cutoff values for the new combinations. The corresponding logistic regression p-values were 0.707 and 0.232 for BB, 0.766 and 0.364 for HA and 0.831 and 0.384 for KSB, and the corresponding true-positive rates (sensitivities) were 88.0% and 93.5% for BB, 78.0% and 88.0% for HA and 79.0% and 87.5% for KSB.

Table 5 The diagnostic accuracies of three new combinations and the standard combination in distinguishing between NPC cases and controls

Combination	New cutoff values[1]	Sensitivity (%)			Specificity (%)	AUC (95%CI)	p[5]
		Early stage (95%CI)	Advanced stage (95%CI)[2]	Average (95%CI)[3]	Control (95% CI)[4]		
BB+EBNA1-IgA	0.258	97.0 (94.6-99.3)	92.8 (89.2-96.4)	93.5 (90.1-96.9)	95.0 (92.0-98.0)	0.977 (0.963-0.991)	0.000

HA+EBNA1-IgA	0.379	97.0 (94.6-99.3)	86.2 (81.5-91.0)	88.0 (83.5-92.5)	94.0 (90.7-97.3)	0.961 (0.943-0.979)	0.000
KSB+EBNA1-IgA	0.191	93.9 (90.6-97.2)	94.6 (91.5-97.7)	94.5 (91.3-97.7)	87.0 (82.3-91.7)*	0.964 (0.947-0.981)	0.000
Standard combination	0.998	97.0 (94.6-99.3)	88.6 (84.2-93.0)	90.0 (85.8-94.2)	95.5 (92.6-98.4)	0.970 (0.956-0.985)	

[1]New cutoff value for NPC diagnosis was defined as the value with the largest Youden Index chosen from each ROC. [2]Differences in the sensitivity of early- and advanced-stage NPC were compared by chi-squared tests. *p<0.05 was considered as significant. [3]Differences in the sensitivities of the new combinations and the standard combination were compared by chi-squared tests. *p<0.05 was considered as significant. [4]Differences in the specificities of the new combinations and the standard combination were compared by chi-squared tests. *p<0.05 was considered as significant. [5] p-values were estimated by non-inferiority tests based on the bootstrap approach for AUC between new combinations and the standard combination.*p>0.05 was considered as significant.

DISCUSSION

In our study, seven recombinant VCA-IgA kits were evaluated, and of these, KSB, BB and HA had diagnostic effects as good as those of the standard kit in terms of sensitivity, specificity and AUC. Combining VCA-IgA with EBNA1-IgA by logistic regression models increased the diagnostic accuracy of these three kits, and all combinations performed as well as the standard combination in sensitivity, specificity and AUC. This is the first study to carry out a comprehensive evaluation of recombinant commercial diagnostic VCA-IgA (ELISA) kits in China, and new logistic models combining VCA-IgA with EBNA1-IgA were established. Furthermore, new cutoff values for VCA-IgA kits and their corresponding combinations were defined for the early detection of and screening for NPC.

In this study, we first calculated the diagnostic performance of each brand of VCA/IgA kit. The AUC of the standard VCA-IgA kit (EUROIMMUN) was 0.942 (95%CI 0.920-0.964), which was consistent with results from our previous studies and verified that the diagnostic performance of VCA-IgA was good and stable.[16 17] We also found that the sensitivities, specificities and AUCs of three kits – KSB, BB and HA – were as high as those of the standard kit, and no significant differences in sensitivity were found between early- and advanced-stage NPC. Moreover, all test–retest reproducibilities were excellent (>0.8). These results suggested that these three kits had equal diagnostic effects and can be substituted for the standard kit. The costs of these recombinant commercial diagnostic kits were only half that of the standard kit, making them more cost-effective.

We noticed that, in contrast to the standard kit with a combined native capsid protein, these testing kits contain primarily recombinant p18 capsid proteins (VCA-p18). VCA-p18 is a small capsid protein that contains several small peptide regions (epitopes) which can be combined to form a powerful diagnostic reagent for VCA-IgA antibody responses.[20] Some researchers have reported that VCA-p18 is the

major VCA antigen for IgA responses.[20 28] Our study showed that the AUCs of these VCA-p18 recombinant kits were more than 0.85, and three of them had the same diagnostic effects as the standard kit, suggesting that, although the manufacturing processes of some recombinant VCA-p18 kits still need to be improved, some of the recombinant kits can be substituted for the standard kit for NPC diagnosis.

As the serum antibody level (rOD) provides continuous data, the cutoff value for distinguishing between NPC cases and controls is critical for the early detection of and screening for NPC. A reasonable cutoff value can balance sensitivity and specificity. In the early detection of and screening for NPC, high sensitivity is required for the identification of high-risk individuals, and high specificity is required for reducing the rate of misdiagnosis and associated costs. According to the cutoff values proved by kits' instructions (always 1), the sensitivities of testing kits were always too low, whilst their specificities were always too high. For example, the sensitivity and the specificity of KSB are 0.780 and 0.925, respectively, suggesting that the old cutoff values should be adjusted. We established new cutoff values for distinguishing between NPC cases and controls by Youdon Indices, and then obtained reasonable sensitivities and specificities. After adjustment, the new cutoff value for KSB is 0.283, and the sensitivity and specificity are 0.890 and 0.875, respectively. Moreover, no differences were found between sensitivities and specificities of these three kits – KSB, BB and HA – and those of the standard kit. There were also no statistically significant differences in the sensitivities of these three kits for early- and advanced-stage NPC.

As for the standard VCA, we found that the combinations of VCA-IgA and EBNA1-IgA by logistic models increased the diagnostic accuracy for NPC from less than 0.946 to more than 0.961 in AUCs. Sensitivities and specificities also increased. For example, the sensitivity and specificity of BB increased from 0.865 and 0.920 to 0.935 and 0.955, respectively. VCA-IgA and EBNA1-IgA are two antibodies corresponding to EBV lytic-cycle proteins and latency gene products, respectively. Therefore, it is reasonable that host antibody responses for lytic-cycle and latency-associated EBV-related proteins can be complementary to each other in the diagnosis of NPC, and the combination of both could increase NPC diagnostic accuracy.[11 29] Furthermore, these three new combinations had diagnostic effects in sensitivities, specificities and AUCs equal to those of the standard combination, suggesting that the combinations of the three recombinant kits can be used for the early detection of and diagnostic screening for NPC. In this study, the control individuals came from NPC-endemic areas and belonged to a screening target population, so we attempted to define people at different risk levels by these new combinations for NPC screening. The NPC incidence rate in the screening target population was relatively low (about 50 per 100,000).[2 10 30] Thus, it is important that the false-positive rate be small enough to avoid unnecessary fiberoptic endoscopy/biopsies and psychological stress for the NPC screening participants. Conversely, the true-positive rate (equal to sensitivity) should be acceptable.[31] We

used two minimally acceptable false-positive rates of 3% and 7% as the high-risk and medium-risk cutoff values, respectively,[17] and the corresponding true-positive rates (sensitivities) for these three kits were 78.0% to 88.0% and 87.5% to 93.5%,[5] respectively. If the baseline serologic results fulfilled the definition of high risk, the participants were referred for diagnostic examinations, and different screening intervals were assigned to the high-, medium- and low-risk groups. The screening intervals for these groups are 1, 1 and 4 years, respectively.[17]

The study had some limitations. First, all cases and controls were from NPC-endemic areas of southern China; therefore, these results might not be applicable to other populations. Second, it was a diagnostic trial in case-control design, and new cutoff values of these new schemes for NPC screening from this study must be verified in prospective mass screenings.

CONCLUSIONS

Three recombinant VCA-IgA kits – KSB, BB and HA – had diagnostic effects equal to those of the standard kit. They can be substituted for the standard kit, and their combinations could be used in the early detection of and screening for NPC.

Contributors Su-Mei Cao and Qing Liu obtained the funding. Rui Gao, Lin Wang and Li-Fang Zhang contributed to study conception and design; Rui Gao conducted experiments; Lin Wang, Yan-Fang Ye, Jin-Lin Du, Shang-Hang Xie, Sui-Hong Chen, Jie Guo and Meng-Jie Yang, Chu-yang Lin acquired or cleaned the data; Rui Gao analysed and interpreted the data; Rui Gao, Su-Mei Cao and Qing Liu drafted or revised the manuscript. All authors critically revised the manuscript for important intellectual content. All authors read and approved the final version of the manuscript.

Funding This work was supported by the National Natural Science Foundation (No. 81373068) and the Special Fund for Public Health Trade (No. 201202014).

Competing interests The authors declare no conflicts of interest.

Ethics approval This study was approved by the Sun Yat-sen University Cancer Center IRB (YB2015-029-01).

Data sharing statement Technical appendix, statistical code, and dataset available from the corresponding author at caosm@sysucc.org.cn and liuqing@sysucc.org.cn.

REFERENCES

1. Ma J, Cao S. The epidemiology of nasopharyngeal carcinoma. In: Lu JJ, Cooper JS, Lee AWM, eds. *Nasopharyngeal cancer*. Berlin: Springer, 2010:1–8.
2. Jia W-H, Huang Q-H, Liao J, *et al*. Trends in incidence and mortality of nasopharyngeal carcinoma over a 20–25 year period (1978/1983–2002) in Sihui and Cangwu counties in southern China. *BMC Cancer* 2006;6:178–85.

3. Chang ET, Adami HO. The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev* 2006;15:1765–77.
4. Cao SM, Simons MJ, Qian CN. The prevalence and prevention of nasopharyngeal carcinoma in China. *Chinese J Cancer* 2011;30:114–9.
5. Sarmiento MP, Mejia MB. Preliminary assessment of nasopharyngeal carcinoma incidence in the Philippines: a second look at published data from four centers. *Chinese J Cancer* 2014;33:159–64.
6. Jia WH. Complex segregation analysis of nasopharyngeal carcinoma in Guangdong, China: evidence for a multifactorial mode of inheritance (complex segregation analysis of NPC in China). *Eur J Hum Genet* 2005;13:248–52.
7. Yu MC, Yuan JM. Epidemiology of nasopharyngeal carcinoma. *Semin Cancer Biol* 2002;12:421–29.
8. Xue WQ, Qin HD, Ruan HL, *et al*. Quantitative association of tobacco smoking with the risk of nasopharyngeal carcinoma: a comprehensive meta-analysis of studies conducted between 1979 and 2011. *Am J Epidemiol* 2013;178:325–38.
9. Chien YC, Chen JY, Liu MY, *et al*. Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men. *N Engl J Med* 2001;345:1877–82.
10. Cao SM, Liu Z, Jia WH, *et al*. Fluctuations of Epstein-Barr virus serological antibodies and risk for nasopharyngeal carcinoma: a prospective screening study with a 20-year follow-up. *PLoS One* 2011;6:e19100.
11. Zeng Y, Zhong JM, Li LY, *et al.* Follow-up studies on Epstein-Barr virus IgA/VCA antibody-positive persons in Zangwu County, China. *Intervirology* 1983;20:190–4.
12. Chen M, Liu M, Hsu S, *et al*. Use of bacterially expressed EBNA-1 protein cloned from a nasopharyngeal carcinoma (NPC) biopsy as a screening test for NPC patients. *J Med Virol* 2001;64:51–7.
13. Zong YS, Sham JS, Ng MH, *et al*. Immunoglobulin A against viral capsid antigen of Epstein-Barr virus and indirect mirror examination of the nasopharynx in the detection of asymptomatic nasopharyngeal carcinoma. *Cancer* 1992;69:3–7.
14. Yi Z, Yuxi L, Chunren L, *et al*. Application of an immunoenzymatic method and an immunoautoradiographic method for a mass survey of nasopharyngeal carcinoma. *Intervirology* 1980;13:162–8.
15. Tang B, Liu Q, Cui Y, *et al*. Protocols of screening, early detection and diagnosis for nasopharyngeal carcinoma. In: Dong Z, Qiao Y, eds. *Protocols of screening, early detection and diagnosis for cancer in P.R.China*. Beijing: People's Medical Publishing House of P.R. China, 2009:187–8.
16. Liu Y, Huang QH, Liu WL, *et al*. Establishment of VCA and EBNA1 IgA-based combination by enzyme-linked immunosorbent assay as preferred screening method for nasopharyngeal carcinoma: a two-stage design with a preliminary performance study and a mass screening in southern China. *Int J Cancer* 2012;131:406–16.
17. Liu ZW, Ji MF, Huang QH, *et al*. Two Epstein-Barr virus–related serologic antibody tests in nasopharyngeal carcinoma screening: results from the initial phase of a cluster randomized controlled trial in Southern China. *Am J Epidemiol*

2013;177:242–50.
18. Department of disease prevention and control, Ministry of health, China. Technical proposal for early diagnosis and treatment of cancer (2011) (Chinese). Peking: People's Medical Publishing House (PMPH) 2011:144-71.
19. Tang JW, Rohwader E, Chu IM, *et al*. Evaluation of EpsteinBarr virus antigen-based immunoassays for serological diagnosis of nasopharyngeal carcinoma. *J Clin Virol* 2007;40:284–8.
20. Middeldorp JM. Epstein-Barr virus-specific humoral immune responses in health and disease. *Curr Top Microbiol Immunol* 2015;391:289–323.
21. Mao YP, Li WF, Chen L, *et al*. A clinical verification of the Chinese 2008 staging system for nasopharyngeal carcinoma (in Chinese). *Ai Zheng* 2009;28:1022–8.
22. Obuchowski N. Testing for equivalence of diagnostic tests. *Am JRadiol* 1997;168:13–7.
23. Liu JP, Ma MCh, Wu ChY, *et al*. Tests of equivalence and noninferiority for diagnostic accuracy based on the paired areas under ROC curve. *Statist Med* 2006;25:1219–38.
24. Zhou XH, Obuchowski NA, McClish DK. *Statistical methods in diagnostic medicine.* New York: Wiley, 2002:188–92.
25. Chen WZ, Zhang JY. Application of non-inferiority test for diagnostic accuracy under the areas of ROC based on bootstrap approach and its macro programming development. *Modern Prev Med (Chinese)* 2010;37:3009–10.
26. Simel D, Samsa G, Matchar D. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991;44:763–70.
27. Fleiss JL. *Reliability of measurement*. New York: Wiley-Interscience, 1986:1–32.
28. Van Grunsven WM, Spaan WJ, Middeldorp JM. Localization and diagnostic application of immunodominant domains of the BFRF3-encoded Epstein-Barr virus capsid protein. *J Infect Dis* 1994;170:13–9.
29. Fachiroh J, Paramita DK, Hariwiyanto B, *et al*. Single assay combination of Epstein-Barr virus (EBV) EBNA1- and viral capsid antigen-p18-derived synthetic peptides for measuring anti-EBV Immunoglobulin G (IgG) and IgA antibody levels in sera from nasopharyngeal carcinoma patients: options for field screening. *J Clin Microbiol* 2006;44:1459–67.
30. Wei K, Xu Y, Liu J, *et al*. No incidence trends and no change in pathological proportions of nasopharyngeal carcinoma in Zhongshan in 1970–2007. *Asian Pacif J Cancer Prev* 2010;11:1595–9.
31. Pepe MS, Feng Z, Janes H, *et al.* Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst* 2008;100:1432–8.

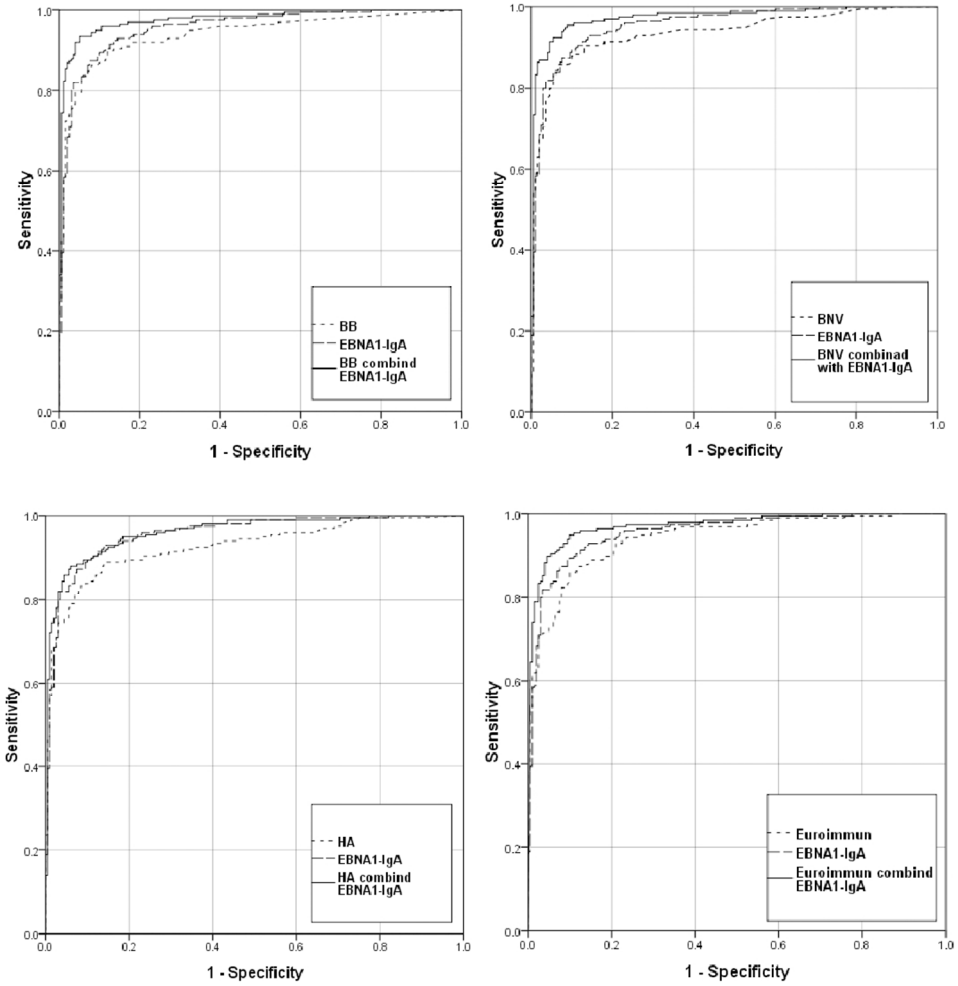

Figure1_ROCs for BB, BNV, HA, Euroimmun and their combination with EBNA1-IgA

94x96mm (300 x 300 DPI)

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)	1,2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	2,3
	4	Study objectives and hypotheses	3
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	4
Participants	**6**	Eligibility criteria	4
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	4
	8	Where and when potentially eligible participants were identified (setting, location and dates)	4
	9	Whether participants formed a consecutive, random or convenience series	4
Test methods	**10a**	Index test, in sufficient detail to allow replication	4,5
	10b	Reference standard, in sufficient detail to allow replication	4,5
	11	Rationale for choosing the reference standard (if alternatives exist)	4
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	5
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	5
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	4,5
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	4,5
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	5
	15	How indeterminate index test or reference standard results were handled	5
	16	How missing data on the index test and reference standard were handled	5
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	5
	18	Intended sample size and how it was determined	4,5
RESULTS			
Participants	**19**	Flow of participants, using a diagram	5,6
	20	Baseline demographic and clinical characteristics of participants	5,6
	21a	Distribution of severity of disease in those with the target condition	5,6
	21b	Distribution of alternative diagnoses in those without the target condition	5,6
	22	Time interval and any clinical interventions between index test and reference standard	5,6
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	6-9
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	6-9
	25	Any adverse events from performing the index test or the reference standard	6-9
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	11
	27	Implications for practice, including the intended use and clinical role of the index test	9-11
OTHER INFORMATION			
	28	Registration number and name of registry	11
	29	Where the full study protocol can be accessed	11
	30	Sources of funding and other support; role of funders	11

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.

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1,2
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	1,2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	2,3
	2b	Specific objectives or hypotheses	3
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	3,4
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	3,4
Participants	4a	Eligibility criteria for participants	3,4
	4b	Settings and locations where the data were collected	3,4
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	3,4
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	3,4
	6b	Any changes to trial outcomes after the trial commenced, with reasons	3,4
Sample size	7a	How sample size was determined	3,4
	7b	When applicable, explanation of any interim analyses and stopping guidelines	3,4
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	4
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	4
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	4
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	4
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	4

		assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	5
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	5
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	5
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	5-9
	13b	For each group, losses and exclusions after randomisation, together with reasons	5-9
Recruitment	14a	Dates defining the periods of recruitment and follow-up	8-9
	14b	Why the trial ended or was stopped	8-9
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	5,6
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	5,6
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	6-9
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	6-9
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	8-9
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	8
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	11
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	9-11
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	9-11
Other information			
Registration	23	Registration number and name of trial registry	11
Protocol	24	Where the full trial protocol can be accessed, if available	5,11
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	11

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

BMJ Open

Evaluation of seven recombinant VCA-IgA ELISA kits for the diagnosis of nasopharyngeal carcinoma in China: a case-control trial

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-013211.R1
Article Type:	Research
Date Submitted by the Author:	01-Feb-2017
Complete List of Authors:	Gao, Rui; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Wang, Lin; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Nasopharyngeal Carcinoma Zhang, Lifang; Guangzhou Women and Children\'s Medical Center, Division of Birth Cohort Study Ye, Yanfang; Sun Yat-Sen University, Sun Yat-sen Memorial Hospital Xie, Shanghang; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention Du, Jinlin; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Chen, Suihong; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Guo, Jie; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Yang, Mengjie ; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Lin, Chuyang; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Liu, Qing; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention Cao, Sumei; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Diagnostics, Epidemiology, Research methods
Keywords:	Nasopharyngeal carcinoma, Epstein-Barr virus, VCA-IgA, Diagnostic effect, Screening

Evaluation of seven recombinant VCA-IgA ELISA kits for the diagnosis of nasopharyngeal carcinoma in China: a case-control trial

Rui Gao[1,2*], Lin Wang[3*], Li-Fang Zhang[4] ,Yan-Fang Ye[5], Shang-Hang Xie[1], Jin-Lin Du[1,2], Sui-Hong Chen[1,2], Jie Guo[1,2], Meng-Jie Yang[1,2], Chu-yang Lin[1,2], Qing liu[1#], Su-Mei Cao[1#]

[1]Department of Cancer Prevention, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Guangzhou, 510060, P.R. China;
[2]School of Public Health, Sun Yat-Sen University, Guangzhou, 510060, P.R. China;
[3]Department of Nasopharyngeal Carcinoma, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Guangzhou, 510060, P.R. China;
[4]Division of Birth Cohort Study, Guangzhou Women and Children's Medical Center, Guangzhou, 510623, P.R. China;
[5]Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, 510120, P.R. China.
*These authors contributed equally to this work.

[#]Correspondence to
Prof. Su-Mei Cao, caosm@sysucc.org.cn;
Prof. Qing Liu, liuqing@sysucc.org.cn

Authors' E-mail:
Rui Gao: gaor5@mail2.sysu.edu.cn Lin Wang: wangl1@sysucc.org.cn
Li-Fang Zhang: lifanglydia@hotmail.com Yan-Fang Ye: yeyanfang1@gmail.com
Shang-Hang Xie: xieshh@sysucc.org.cn Jin-Lin Du: dujl3@mail2.sysu.edu.cn
Sui-Hong Chen: chensh43@mail2.sysu.edu.cn Jie Guo: guojiegina@126.com
Meng-Jie Yang: yangmj@sysucc.org.cn Chu-yang Lin: linchy23@mail2.sysu.edu.cn

Preferred reviewers:
1. Fang Fang, Fang.Fang@ki.se, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
2. Liming Liang, lliang@hsph.harvard.edu, Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA, US.
3. Qing Chen, qingchen@fimmu.com, Department of Epidemiology, School of Public Health and Tropical Medicine, Southern Medical University, Guangzhou, China

Key words: Nasopharyngeal carcinoma, Epstein-Barr virus, VCA-IgA, Diagnostic effect, Screening

Word count: 2695

ABSTRACT

Objective: Seven recombinant VCA-IgA ELISA kits are widely used in China, but their diagnostic effects have not been evaluated. In this study, we evaluated whether the diagnostic effects of these kits are similar to those of the standard kit (EUROIMMUN, Lübeck, Germany).

Methods: A diagnostic case-control trial was conducted, with 200 cases of nasopharyngeal carcinoma (NPC) and 200 controls from NPC-endemic areas in southern China. The areas under the curve (AUCs), the sensitivities and the specificities of testing kits were compared with those of the standard kit. The test–retest reliability of each kit was determined by intraclass correlation coefficient (ICC). Their diagnostic accuracy in combination with EBNA1-IgA was also evaluated in logistic models.

Results: Three testing kits – KSB, BB and HA – showed diagnostic accuracy equal to that of the standard kit, with good performance in the AUCs (0.926~0.945), and no significant differences in sensitivity were found between early- and advanced-stage NPCs. ICCs exceeded 0.8. Three logistic regression models were built, and the AUCs of these models (0.961~0.977) were better than those of the individual VCA-IgA kits. All new models had diagnostic accuracy equal to that of the standard kit. New cutoff values of these three kits and their corresponding combinations for researchers to replicate and use in NPC early detection and screening in the future were provided.

Conclusions: Three recombinant VCA-IgA kits – KSB, BB and HA – had diagnostic effects equal to those of the standard kit, and, in combination with EBNA1-IgA in logistic regression models, can be used in future screening for NPC.

Strengths and limitations of this study

• This is the first study to carry out a comprehensive evaluation of recombinant commercial diagnostic VCA-IgA (ELISA) kits in China, and logistic models combining VCA-IgA with EBNA1-IgA were established.

• New cutoff values for VCA-IgA kits and their corresponding combinations for researchers to replicate and use in NPC early detection and screening in the future were provided. • All cases and controls were from NPC-endemic areas of southern China, and thus these results might not be applicable to other populations.

• Cutoff values for NPC screening by means of these models described in this study must be verified in prospective mass screening.

INTRODUCTION

Nasopharyngeal carcinoma (NPC) is a common form of squamous-cell carcinoma in southern China and southeastern Asia. The annual incidence rate of NPC in southern China can reach 25 per 100,000 person-years, which is about 25-fold higher than in the rest of the world.[1–4] NPC is a complex disease caused by a combination of Epstein-Barr virus (EBV), chronic infection, the environment and host genes in a

multi-step process of carcinogenesis, but until now there have been no effective preventive measures.[5–9] Long-term survival rates differ substantially between patients with advanced- (stages III and IV) and early- (stages I and II) stage NPC. Four-year survival rates of early-stage NPC patients are 96.7% compared with 67.1% for those with advanced-stage NPC.[5] Mass screening has become the most practical method for improved early detection in and overall prognosis of NPC patients in the endemic areas.[10 11]

Serum antibodies against EBV-related antigens, especially immunoglobulin A (IgA) against viral capsid antigen (VCA-IgA), early antigen (EA-IgA), EBV nuclear antigen 1(EBNA1-IgA) and so on, remain elevated for an average of 38 months in the preclinical phase,[9–14] and serological tests for these markers are simple and inexpensive.[15–18] Therefore, since the 1970s, these tests have been used as screening markers for NPC in endemic areas. In our previous study, we evaluated the diagnostic performance of seven commercial EBV-related antibodies by enzyme-linked immunosorbent assay (ELISA) and found EBNA1-IgA (Zhongshan Biotech, China) and VCA-IgA (EUROIMMUN, Lübeck, Germany) ELISA to be the top two seromarkers, with AUCs of 0.95 (95%CI, 0.93–0.97) and 0.94 (95%CI, 0.92–0.97), respectively[16]. We further verified that the combination of VCA-IgA and EBNA1-IgA outperformed any individual EBV seromarkers, with AUC up to 0.97 (95%CI, 0.96, 0.99).[15–17] Thus, since 2011, the combination of VCA-IgA and EBNA1-IgA has been recommended as the standard tool for NPC screening in China.[18]

Nowadays, several kinds of commercial VCA-IgA kits based on recombinant peptides have been developed in China and are presently widely used for the early detection of and screening for NPC. However, their diagnostic performance for NPC alone and in combination with EBNA1-IgA has not been evaluated. In this study, we evaluated whether the effects of the NPC-diagnostic kits are comparable with those of the standard VCA-IgA kit and can be substituted for it. If so, we will further explore the combination diagnostic strategy with EBNA1-IgA for the early detection of and mass screening for NPC.

METHODS

Study population

Serum specimens were continuously collected from 200 patients hospitalised with NPC in the Sun Yat-sen University Cancer Center (SYSUCC) from January 2013 to June 2013. These cases were histologically confirmed by biopsy, and the clinical stages were classified according to the 2009 Union for International Cancer Control (UICC) criteria, including 33 patients with early-stage NPC (stages I, II) and 167 with advanced-stage NPC (stages III, IV). The inclusion criteria included being between 30 and 59 years of age and residing in one of the six high-endemic provinces of southern China (Guangdong, Guangxi, Jiangxi, Hunan, Fujian or Hainan Province). Other information, including demographic data, smoking, drinking histories and

family history of NPC, was collected by the physician in charge. All serum samples were collected before treatment.

The 200 healthy controls were randomly selected from among healthy people who participated in physical examinations at the Sihui Cancer Center (Sihui City, Guangdong Province, China) from July 2013 to September 2013 and were frequency-matched with cases by age (5-year age groups) and gender. All participants completed a short questionnaire to record demographic data, smoking, drinking histories and family history of NPC and donated 3 mL of blood.

This study was approved by the Clinical Research Ethics Committee of the SYSUCC (YB2015-029-01), and all participants provided written informed consent.

Detection of serological EBV antibodies

Serum and buffy coat were separated less than 4 hours after collection and stored at -80°C before being tested. None of the specimens was hemolytic or repeatedly frozen more than twice. Seven recombinant VCA-IgA kits, the standard VCA-IgA kit (EUROIMMUN) and the standard EBNA1-IgA kit (Zhongshan) were tested (table 1).

Table 1 Product information for eight brands of VCA-IgA kits and the EBNA1-IgA kit

Abbreviation for kits	Manufacturer
VCA-IgA	
KSB	Shenzhen Kang Sheng Bao Bio-Technology Co., Ltd.
BNV	Bioneovan Co., Ltd.
GBI	Beijing BGI-GBI Biotech Co., Ltd.
BB	Beijing Beier Bioengineering Co., Ltd.
HA	Shenzhen HuianBioscitech Co., Ltd.
HK	Shen Zhen HuaKang Co., Ltd.
ZS	ZhongShan Biotech Co., Ltd.
EUROIMMUN	EUROIMMUN Medizinische Labordiagnostika AG
EBNA1-IgA	ZhongShan Biotech Co., Ltd.

All samples were renumbered and tested blindly by one technician according to the manufacturers' instructions. Levels of antibodies were assessed by photometric measurement, which provided optical density (OD) values. Reference ODs (rOD) were obtained according to manufacturers' instructions by dividing OD values by a reference control. To investigate the test-retest reliability of each kit, 10% serum samples (40 samples) were randomly chosen and retested.

Statistical analysis

Demographic characteristics and NPC risk factors between cases and controls were compared by chi-squared tests. The cutoff value of each single kit was defined with

the largest Youden Indices (sensitivity+specificity-1) chosen from each receiver operating characteristic (ROC). The diagnostic efficacy of each kit was evaluated by AUC, and non-inferiority tests based on the bootstrap approach were performed to determine whether the AUCs of these recombinant testing kits were inferior to that of the standard kit (let Δ=0.05 be the pre-determined clinically meaningful equivalence limit).[19-22] The sensitivity and specificity of each kit were calculated, and their 95% confidence intervals (CIs) were estimated by the methods of Simel and colleagues.[23] Differences in sensitivities between early- and advanced-stage NPC with each kit were compared by Chi-squared tests (Fisher's exact test and McNemar's test will be specified while others Chi-squared tests means Person's Chi-square test). Intraclass correlation coefficients (ICC) were performed to determine test–retest reliability.

In order to prevent bias and study the virus factor only, we matched the baseline covariates (gender and age) and some of the important NPC risk factors (smoking, drinking and NPC history). Binary unconditional logistic regressions were used to establish formulae for VCA-IgA and EBNA1-IgA. The diagnostic efficacy of each formula was evaluated by sensitivity, specificity and AUC, compared with the standard formula, *Logit P* = -3.934 + 2.203VCA-IgA (EUROIMMUN) + 4.797EBNA1-IgA. The cutoff p-value in the corresponding logistic regression for distinguishing between NPC cases and controls was defined with the largest Youden Index chosen from each ROC. Two minimally acceptable false-positive rates (1-Specificity), 3% and 7%, were used empirically to establish the cutoff p-values for classifying different NPC risk subgroups.[16 17]

The non-inferiority tests were one-sided, and $p>0.05$ was considered to be non-inferior. Other tests were two-sided, and $p<0.05$ was considered to be statistically significant. Data were analyzed by SAS9.2 and SPSS16.0 software.

RESULTS

Baseline information

Baseline information on gender, age, smoking, drinking and NPC family history was comparable between cases and controls, and no statistically significant differences were found between them. Further, there were no statistically significant differences for these items between early- and advanced-stage cases (table 2).

Table 2 Demographic characteristics of NPC cases and controls

Categories	NPC Cases (N_1=200) No. (%)				Controls (N_2=200) No. (%)	p^5
	Early stage (n=33)	Advanced stage (n=167)	p^4	Total		
Gender			0.472			0.417
Male	27 (81.8)	127 (76.0)		154 (77.0)	147 (73.5)	

Female	6 (18.2)	40 (24.0)		46 (23.0)	53 (26.5)	
Age (years)			0.299			0.785
30~	6 (18.2)	47 (28.1)		53 (26.5)	47 (23.5)	
40~	13 (39.4)	70 (41.9)		83 (41.5)	87 (43.5)	
50~	14 (42.4)	50 (29.9)		64 (32.0)	66 (33.0)	
Smoking[1]			0.857			0.746
Yes	11 (33.3)	53 (31.7)		64 (32.0)	61 (30.5)	
No	22 (66.7)	114 (68.3)		136 (68.0)	139 (69.5)	
Drinking[2]			0.641			0.494
Yes	6 (18.2)	25 (15.0)		31 (15.5)	27 (13.5)	
No	27 (81.8)	142 (85.0)		169 (84.5)	173 (86.5)	
NPC family history[3]			0.732			0.224
Yes	3 (9.1)	13 (7.8)		16 (8.0)	10 (5.0)	
No	30 (90.9)	154 (92.2)		184 (92.0)	190 (95.0)	

[1]'Smoking' refers to people who smoked more than one cigarette every three days within half a year and included current and former smokers.[2]'Drinking' refers to people who consumed alcoholic beverages every week within half a year and included current and former drinkers. [3]'NPC family history' refers to people whose parents, children and siblings have or did have NPC. [4]Differences in early- and advanced-stage NPC were compared by chi-squared tests (Fisher's Exact Test for NPC family history). *$p<0.05$ was considered as statistically significant. [5]Differences in NPC Cases and Controls were compared by chi-squared tests. *$p<0.05$ was considered as statistically significant.

The diagnostic accuracies of eight brands of VCA-IgA kits and the EBNA1-IgA kit

Table 3 shows that the AUCs of four kits – KSB, BB, BNV and HA – were as high as that of the standard VCA-IgA kit (AUC, 0.942; 95%CI, 0.920-0.964). The AUCs, in order, were 0.945 for KSB (95%CI 0.925-0.966), 0.940 for BB (95%CI 0.916-0.964), 0.936 for BNV (95%CI 0.911-0.961) and 0.926 for HA (95%CI 0.900-0.953). In addition, the AUCs of GBI, HK and ZS were lower than that of the standard kit. Furthermore, no significant differences were found between early- and advanced-stage NPC in the sensitivities of six kits ($p>0.05$), except for BNV ($p=0.044$).

Table 3 The diagnostic accuracies of eight brands of VCA-IgA kits and the EBNA1-IgA kit in distinguishing between NPC cases and controls

Kits	Cutoff values[1]	Sensitivity (%)			Specificity (%)	AUC (95%CI)	P[3]
		Early stage (95%CI)	Advanced stage (95%CI)[2]	Total (95%CI)	Control (95% CI)		
VCA-IgA							

BB	0.58	75.8 (69.8-81.7)	88.6 (84.2-93.0)	86.5 (81.8-91.2)	92.0 (88.2-95.8)	0.940 (0.916-0.964)	0.002
BNV	0.923	72.7 (66.6-78.9)	88.0 (83.5-92.5)*	86.0 (81.2-90.8)	93.5 (90.1-96.9)	0.936 (0.911-0.961)	0.003
GBI	0.825	72.7 (66.6-78.9)	76.6 (70.8-82.5)	76.0 (70.1-81.9)	92.0 (88.2-95.8)	0.899 (0.868-0.930)	0.341*
HA	0.884	93.9 (90.6-97.2	88.0 (83.5-92.5)	89.0 (84.7-93.3)	86.0 (81.2-90.8)	0.926 (0.900-0.953)	0.012
HK	1.218	81.8 (76.5-87.2)	83.2 (78.1-88.4)	83.0 (77.8-88.2)	89.5 (85.3-93.7)	0.913 (0.884-0.942)	0.075*
KSB	0.283	100.0 (100.0-100.0)	86.8 (82.1-91.5)	89.0 (84.7-93.3)	87.5 (82.9-92.1)	0.945 (0.925-0.966)	0.000
ZS	0.418	75.8 (69.8-81.7)	74.3 (68.2-80.3)	74.5 (68.5-80.5)	87.5 (82.9-92.1)	0.868 (0.831-0.904)	0.878*
EUROIMMUN	1.561	87.9 (83.4-92.4)	85.6 (80.8-90.5)	86.0 (81.2-90.8)	90.0 (85.8-94.2)	0.942 (0.921-0.964)	
EBNA1-IgA	1.203	93.9 (90.6-97.2)	86.2 (81.5-91.0)	87.5 (82.9-92.1)	92.5 (88.8-96.2)	0.956 (0.937-0.975)	0.000

[1]Cutoff value for NPC diagnosis was defined as the value with the largest Youden Index chosen from each ROC. [2]Differences in the sensitivities of early- and advanced-stage NPC were compared by Person Chi-Squared tests. *p<0.05 was considered as statistically significant. [3]p values were estimated by non-inferiority tests based on the bootstrap approach for AUC between EUROIMMUN and other kits. *p<0.05 was considered as statistically significant while p>0.05 was consider to be inferior to the standard kit.

The test–retest reliabilities of eight brands of VCA-IgA kits and the EBNA1-IgA kit

Ten percent serum samples (40 samples) were randomly chosen and retested for calculation of the ICC of each brand of kit, VCA-IgA or EBNA1-IgA. The test-retest reliabilities of all kits were excellent (>0.75, excellent) according to Fleiss's classification[24] (table 4).

Table 4 The test–retest reliabilities of eight brands of VCA-IgA kits and the EBNA1-IgA kit

Kits	ICC*	95%CI
VCA-IgA		
BB	0.990	0.980-0.994
BNV	0.982	0.967-0.991
GBI	0.964	0.933-0.981
HA	0.975	0.952-0.987
HK	0.876	0.764-0.935
KSB	0.823	0.666-0.906

ZS	0.978	0.958-0.988
EUROIMMUN	0.913	0.830-0.955
EBNA1-IgA	0.981	0.964-0.990

*Less than 0.40- poor; Between 0.40 and 0.59- Fair;
Between 0.60 and 0.74- good; Between 0.75 and 1.00- Excellent

The diagnostic accuracies of the combinations of VCA-IgA and EBNA1-IgA with logistic models

We chose for testing three VCA-IgA kits with high AUCs, no differences in diagnoses for early- and advanced-stage NPC and excellent test-retest reliabilities, and then combined each with the EBNA1-IgA kit by logistic models. Three logistic regression models were established:

LogitP = -3.2323 + 0.8060VCA-IgA (BB) + 1.1044 EBNA1-IgA
LogitP = -2.7591 + 0.6380VCA-IgA (HA) + 1.0620EBNA1-IgA
LogitP = -2.6039 + 0.5312VCA-IgA (KSB) + 1.1673 EBNA1-IgA

In all these models, both VCA-IgA and EBNA1-IgA were statistically significant independent predictors of NPC risk ($p<0.05$), and the AUC of each combination was statistically significantly larger than that of each single VCA-IgA ($p<0.05$). The AUC of KSB increased from 0.945 (95%CI 0.925-0.966) to 0.964 (95%CI 0.947-0.981); BB increased from 0.940 (95%CI 0.916-0.964) to 0.977 (95%CI 0.963-0.991); and HA increased from 0.926 (95%CI 0.900-0.953) to 0.961 (95%CI 0.943-0.979) (figure 1).

Table 5 shows the diagnostic accuracies of the three new combinations and the standard combination [*Logit P*=-3.934+2.203VCA-IgA (EUROIMMUN) +4.797EBNA1-IgA] in distinguishing between NPC cases and controls. The AUCs of these three combinations were as high as that of the standard combination (AUC 0.970; 95%CI 0.956-0.985) ($p<0.05$). Furthermore, no statistically significant difference was found in the sensitivity of each combination between early- and advanced-stage NPC ($p>0.05$).

We used two minimally acceptable false-positive rates (1-Specificity) of 3% and 7% to define the high-risk and medium-risk cutoff values for the new combinations. The corresponding logistic regression p-values were 0.707 and 0.232 for BB, 0.766 and 0.364 for HA, 0.831 and 0.384 for KSB, and the corresponding true-positive rates (Sensitivities) were 88.0% and 93.5% for BB, 78.0% and 88.0% for HA and 79.0% and 87.5% for KSB.

Table 5 The diagnostic accuracies of three new combinations and the standard combination in distinguishing between NPC cases and controls

Combination	New cutoff	Sensitivity (%)			Specificity (%)	AUC	P[3]
		Early stage	Advanced stage	Total	Control	(95%CI)	

	values[1]	(95%CI)	(95%CI)[2]	(95%CI)[3]	(95% CI)[4]		
BB+EBNA1-IgA	0.258	97.0 (94.6-99.3)	92.8 (89.2-96.4)	93.5 (90.1-96.9)	95.0 (92.0-98.0)	0.977 (0.963-0.991)	<0.001
HA+EBNA1-IgA	0.379	97.0 (94.6-99.3)	86.2 (81.5-91.0)	88.0 (83.5-92.5)	94.0 (90.7-97.3)	0.961 (0.943-0.979)	<0.001
KSB+EBNA1-IgA	0.191	93.9 (90.6-97.2)	94.6 (91.5-97.7)	94.5 (91.3-97.7)	87.0 (82.3-91.7)*	0.964 (0.947-0.981)	<0.001
Standard combination	0.998	97.0 (94.6-99.3)	88.6 (84.2-93.0)	90.0 (85.8-94.2)	95.5 (92.6-98.4)	0.970 (0.956-0.985)	

[1]New cutoff value for NPC diagnosis was defined as the value with the largest Youden Index chosen from each ROC. [2]Differences in the sensitivity of early- and advanced-stage NPC were compared by Person Chi-Squared tests. *p<0.05 was considered as statistically significant. [3] p-values were estimated by non-inferiority tests based on the bootstrap approach for AUC between new combinations and the standard combination. *p<0.05 was considered as statistically significant while p>0.05 was consider to be inferior to the standard kit.

DISCUSSION

In our study, seven recombinant VCA-IgA kits were evaluated, and of these, KSB, BB and HA had diagnostic effects as good as those of the standard kit in terms of sensitivity, specificity and AUC. Combining VCA-IgA with EBNA1-IgA by logistic regression models increased the diagnostic accuracy of these three kits, and all combinations performed as well as the standard combination in sensitivity, specificity and AUC. This is the first study to carry out a comprehensive evaluation of recombinant commercial diagnostic VCA-IgA (ELISA) kits in China, and logistic models combining VCA-IgA with EBNA1-IgA were established. Furthermore, new cutoff values for these VCA-IgA kits and their corresponding combinations for researchers to replicate and use in NPC early detection and screening in the future were provided.

In this study, we first calculated the diagnostic performance of each brand of VCA/IgA kit. The AUC of the standard VCA-IgA kit (EUROIMMUN) was 0.942 (95%CI 0.920-0.964), which was consistent with results from our previous studies and verified that the diagnostic performance of VCA-IgA was good and stable.[16 17] We also found that the sensitivities, specificities and AUCs of three kits – KSB, BB and HA – were as high as those of the standard kit, and no significant differences in sensitivity were found between early- and advanced-stage NPC. Moreover, all test–retest reproducibilities were excellent (>0.75). These results suggested that these three kits had equal diagnostic effects and can be substituted for the standard kit. The costs of these recombinant commercial diagnostic kits were only half that of the standard kit, making them more cost-effective.

The EBV capsid antigen (VCA) is a late protein produced in the EBV lytic infection period. VCA contains a batch of capsid proteins, such as VCA-p18 (BFRF3),

VCA-p23 (BLRF2), gp125/110 (BALF4) and so on, which have unique immune dominants and virus-specific antigenic domains. These domains contain several small peptide regions (epitopes) which can be combined to form a powerful diagnostic reagent for VCA-IgA.[25-26] The capsid proteins in the EUROIMMUN kit[27] were extracted from the pyrolysis products of human B lymphocytes (P3HR1 cell line) infected by EBV and contained a combined native capsid protein of EBV. We noticed that, in contrast to the standard kit with a combined native capsid protein, these testing kits contain primarily recombinant p18 capsid proteins (VCA-p18). VCA-p18 is a small capsid protein that contains several small peptide regions (epitopes) which can be combined to form a powerful diagnostic reagent for VCA-IgA antibody responses.[26] Some researchers have reported that VCA-p18 is the major VCA antigen for IgA responses.[26 28] Our study showed that the AUCs of these VCA-p18 recombinant kits were more than 0.85, and three of them had the same diagnostic effects as the standard kit, suggesting that, although the manufacturing processes of some recombinant VCA-p18 kits still need to be improved, some of the recombinant kits can be substituted for the standard kit for NPC diagnosis.

As the serum antibody level (rOD) provides continuous data, the cutoff value for distinguishing between NPC cases and controls is critical for the early detection of and screening for NPC. A reasonable cutoff value can balance sensitivity and specificity. In the early detection of and screening for NPC, high sensitivity is required for the identification of high-risk individuals, and high specificity is required for reducing the rate of misdiagnosis and associated costs. According to the cutoff values provided by the kits' instructions, the sensitivities of testing kits were always too low, whilst their specificities were always too high. For example, the sensitivity and the specificity of KSB are 0.780 and 0.925, respectively, suggesting that the old cutoff values should be adjusted. We established new cutoff values for distinguishing between NPC cases and controls by Youdon Indices, and then obtained reasonable sensitivities and specificities. After adjustment, the new cutoff value for KSB is 0.283, and the sensitivity and specificity are 0.890 and 0.875, respectively. Moreover, no differences were found between sensitivities and specificities of these three kits – KSB, BB and HA – and those of the standard kit. There were also no statistically significant differences in the sensitivities of these three kits for early- and advanced-stage NPC.

As for the standard VCA, we found that the combinations of VCA-IgA and EBNA1-IgA by logistic models increased the diagnostic accuracy for NPC from less than 0.946 to more than 0.961 in AUCs. Sensitivities and specificities also increased. For example, the sensitivity and specificity of BB increased from 0.865 and 0.920 to 0.935 and 0.955, respectively. VCA-IgA and EBNA1-IgA are two antibodies corresponding to EBV lytic-cycle proteins and latency gene products, respectively. Therefore, it is reasonable that host antibody responses for lytic-cycle and latency-associated EBV-related proteins can be complementary to each other in the diagnosis of NPC, and the combination of both could increase NPC diagnostic

accuracy.[11 29] Furthermore, these three new combinations had diagnostic effects in sensitivities, specificities and AUCs equal to those of the standard combination, suggesting that the combinations of the three recombinant kits can be used for the early detection of and diagnostic screening for NPC. In this study, the control individuals came from NPC-endemic areas and belonged to a screening target population, so we attempted to define people at different risk levels by these new combinations for NPC screening. Compared with other common diseases, the NPC incidence rate in the screening target population was relatively low (about 50 per 100,000 person-years)[2 10 30] Thus, it is important that the false-positive rate be small enough to avoid unnecessary fiberoptic endoscopy/biopsies and psychological stress for the NPC screening participants. Conversely, the true-positive rate (equal to sensitivity) should be acceptable.[31] We used two minimally acceptable false-positive rates of 3% and 7% as the high-risk and medium-risk cutoff values, respectively,[17] and the corresponding true-positive rates (sensitivities) for these three kits were 78.0% to 88.0% and 87.5% to 93.5%,[5] respectively. If the baseline serologic results fulfilled the definition of high risk, the participants were referred for diagnostic examinations, and different screening intervals were assigned to the high-, medium- and low-risk groups. The screening intervals for these groups are 1, 1 and 4 years, respectively.[17]

The study had some limitations. First, this study was a Single-center study and all cases and controls were from NPC-endemic areas of southern China (controls were from hospital); therefore, these results might not be applicable to other populations. Second, though we have confidence VCA-IgA in this study have the same sensitivities for detection of early and advanced stage NPC patients[10 11 32], due to the low percentage (less than 20%) of early stage in clinic, we can only collect 33 early stage NPC participants in our study period. Third, it was a diagnostic trial in case-control design, and new cutoff values of these new schemes for NPC screening from this study must be verified in prospective mass screenings.

CONCLUSIONS

Three recombinant VCA-IgA kits – KSB, BB and HA – had diagnostic effects equal to those of the standard kit. They can be substituted for the standard kit, and their combinations could be used in the early detection of and screening for NPC.

Contributors Su-Mei Cao and Qing Liu obtained the funding. Rui Gao, Lin Wang and Li-Fang Zhang contributed to study conception and design; Rui Gao conducted experiments; Lin Wang, Yan-Fang Ye, Jin-Lin Du, Shang-Hang Xie, Sui-Hong Chen, Jie Guo and Meng-Jie Yang, Chu-yang Lin acquired or cleaned the data; Rui Gao analysed and interpreted the data; Rui Gao, Su-Mei Cao and Qing Liu drafted or revised the manuscript. All authors critically revised the manuscript for important intellectual content. All authors read and approved the final version of the manuscript.

Funding This work was supported by the National Natural Science Foundation (No. 81373068) and the Special Fund for Public Health Trade (No. 201202014).

Competing interests The authors declare no conflicts of interest.

Ethics approval This study was approved by the Sun Yat-sen University Cancer Center IRB (YB2015-029-01).

Data sharing statement Technical appendix, statistical code, and dataset available from the corresponding author at caosm@sysucc.org.cn and liuqing@sysucc.org.cn.

REFERENCES

1. Ma J, Cao S. The epidemiology of nasopharyngeal carcinoma. In: Lu JJ, Cooper JS, Lee AWM, eds. *Nasopharyngeal cancer*. Berlin: Springer, 2010:1–8.
2. Jia W-H, Huang Q-H, Liao J, *et al*. Trends in incidence and mortality of nasopharyngeal carcinoma over a 20–25 year period (1978/1983–2002) in Sihui and Cangwu counties in southern China. *BMC Cancer* 2006;6:178–85.
3. Chang ET, Adami HO. The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev* 2006;15:1765–77.
4. Cao SM, Simons MJ, Qian CN. The prevalence and prevention of nasopharyngeal carcinoma in China. *Chinese J Cancer* 2011;30:114–9.
5. Sarmiento MP, Mejia MB. Preliminary assessment of nasopharyngeal carcinoma incidence in the Philippines: a second look at published data from four centers. *Chinese J Cancer* 2014;33:159–64.
6. Jia WH. Complex segregation analysis of nasopharyngeal carcinoma in Guangdong, China: evidence for a multifactorial mode of inheritance (complex segregation analysis of NPC in China). *Eur J Hum Genet* 2005;13:248–52.
7. Yu MC, Yuan JM. Epidemiology of nasopharyngeal carcinoma. *Semin Cancer Biol* 2002;12:421–29.
8. Xue WQ, Qin HD, Ruan HL, *et al*. Quantitative association of tobacco smoking with the risk of nasopharyngeal carcinoma: a comprehensive meta-analysis of studies conducted between 1979 and 2011. *Am J Epidemiol* 2013;178:325–38.
9. Chien YC, Chen JY, Liu MY, *et al*. Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men. *N Engl J Med* 2001;345:1877–82.
10. Cao SM, Liu Z, Jia WH, *et al*. Fluctuations of Epstein-Barr virus serological antibodies and risk for nasopharyngeal carcinoma: a prospective screening study with a 20-year follow-up. *PLoS One* 2011;6:e19100.
11. Zeng Y, Zhong JM, Li LY, *et al.* Follow-up studies on Epstein-Barr virus IgA/VCA antibody-positive persons in Zangwu County, China. *Intervirology* 1983;20:190–4.
12. Chen M, Liu M, Hsu S, *et al*. Use of bacterially expressed EBNA-1 protein cloned from a nasopharyngeal carcinoma (NPC) biopsy as a screening test for NPC patients. *J Med Virol* 2001;64:51–7.
13. Zong YS, Sham JS, Ng MH, *et al*. Immunoglobulin A against viral capsid antigen of Epstein-Barr virus and indirect mirror examination of the nasopharynx in the

detection of asymptomatic nasopharyngeal carcinoma. *Cancer* 1992;69:3–7.
14. Yi Z, Yuxi L, Chunren L, *et al*. Application of an immunoenzymatic method and an immunoautoradiographic method for a mass survey of nasopharyngeal carcinoma. *Intervirology* 1980;13:162–8.
15. Tang B, Liu Q, Cui Y, *et al*. Protocols of screening, early detection and diagnosis for nasopharyngeal carcinoma. In: Dong Z, Qiao Y, eds. *Protocols of screening, early detection and diagnosis for cancer in P.R.China*. Beijing: People's Medical Publishing House of P.R. China, 2009:187–8.
16. Liu Y, Huang QH, Liu WL, *et al*. Establishment of VCA and EBNA1 IgA-based combination by enzyme-linked immunosorbent assay as preferred screening method for nasopharyngeal carcinoma: a two-stage design with a preliminary performance study and a mass screening in southern China. *Int J Cancer* 2012;131:406–16.
17. Liu ZW, Ji MF, Huang QH, *et al*. Two Epstein-Barr virus–related serologic antibody tests in nasopharyngeal carcinoma screening: results from the initial phase of a cluster randomized controlled trial in Southern China. *Am J Epidemiol* 2013;177:242–50.
18. Department of disease prevention and control, Ministry of health, China. Technical proposal for early diagnosis and treatment of cancer (2011) (Chinese). Peking: People's Medical Publishing House (PMPH) 2011:144-71.
19. Obuchowski N. Testing for equivalence of diagnostic tests. *Am JRadiol* 1997;168:13–7.
20. Liu JP, Ma MCh, Wu ChY, *et al*. Tests of equivalence and noninferiority for diagnostic accuracy based on the paired areas under ROC curve. *Statist Med* 2006;25:1219–38.
21. Zhou XH, Obuchowski NA, McClish DK. *Statistical methods in diagnostic medicine.* New York: Wiley, 2002:188–92.
22. Chen WZ, Zhang JY. Application of non-inferiority test for diagnostic accuracy under the areas of ROC based on bootstrap approach and its macro programming development. *Modern Prev Med (Chinese)* 2010;37:3009–10.
23. Simel D, Samsa G, Matchar D. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991;44:763–70.
24. Fleiss JL. *Reliability of measurement*. The designand analysis of clinical experiments, 1st edn. New York: Wiley-Interscience, 1986:1–32.
25. Tang JW, Rohwader E, Chu IM, *et al*. Evaluation of EpsteinBarr virus antigen-based immunoassays for serological diagnosis of nasopharyngeal carcinoma. *J Clin Virol* 2007;40:284–8.
26. Middeldorp JM. Epstein-Barr virus-specific humoral immune responses in health and disease. *Curr Top Microbiol Immunol* 2015;391:289–323.
27. Mao YP, Li WF, Chen L, *et al*. A clinical verification of the Chinese 2008 staging system for nasopharyngeal carcinoma (in Chinese). *Ai Zheng* 2009;28:1022–8.
28. Van Grunsven WM, Spaan WJ, Middeldorp JM. Localization and diagnostic application of immunodominant domains of the BFRF3-encoded Epstein-Barr virus capsid protein. *J Infect Dis* 1994;170:13–9.
29. Fachiroh J, Paramita DK, Hariwiyanto B, *et al*. Single assay combination of

Epstein-Barr virus (EBV) EBNA1- and viral capsid antigen-p18-derived synthetic peptides for measuring anti-EBV Immunoglobulin G (IgG) and IgA antibody levels in sera from nasopharyngeal carcinoma patients: options for field screening. *J Clin Microbiol* 2006;44:1459–67.

30. Wei K, Xu Y, Liu J, *et al*. No incidence trends and no change in pathological proportions of nasopharyngeal carcinoma in Zhongshan in 1970–2007. *Asian Pacif J Cancer Prev* 2010;11:1595–9.

31. Pepe MS, Feng Z, Janes H, *et al.* Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst* 2008;100:1432–8.

32. Stolzenberg MC, et al. Purified recombinant EBV desoxyribonuclease in serological diagnosis of nasopharyngeal carcinoma. Int J Cancer. 1996 May 3;66(3):337-41

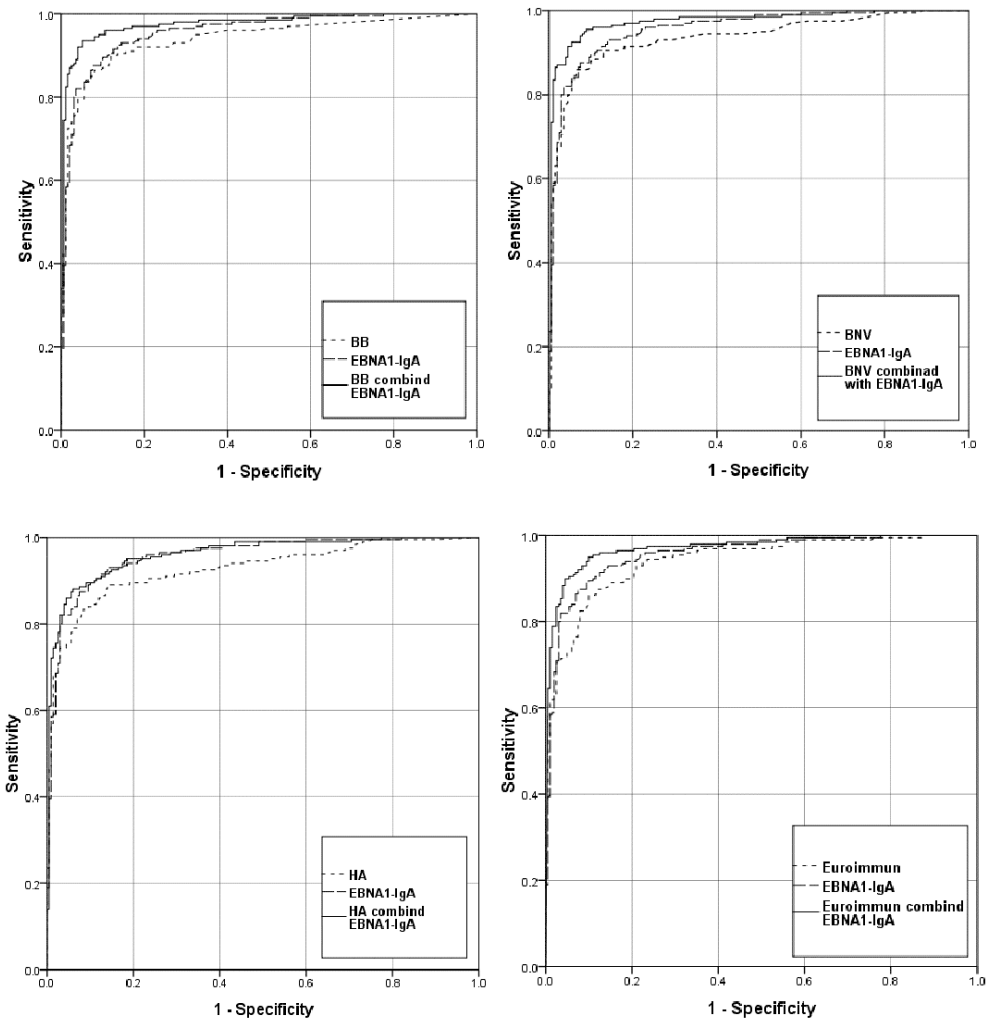

Figure1 ROCs for BB, BNV, HA, Euroimmun and their combination with EBNA1-IgA

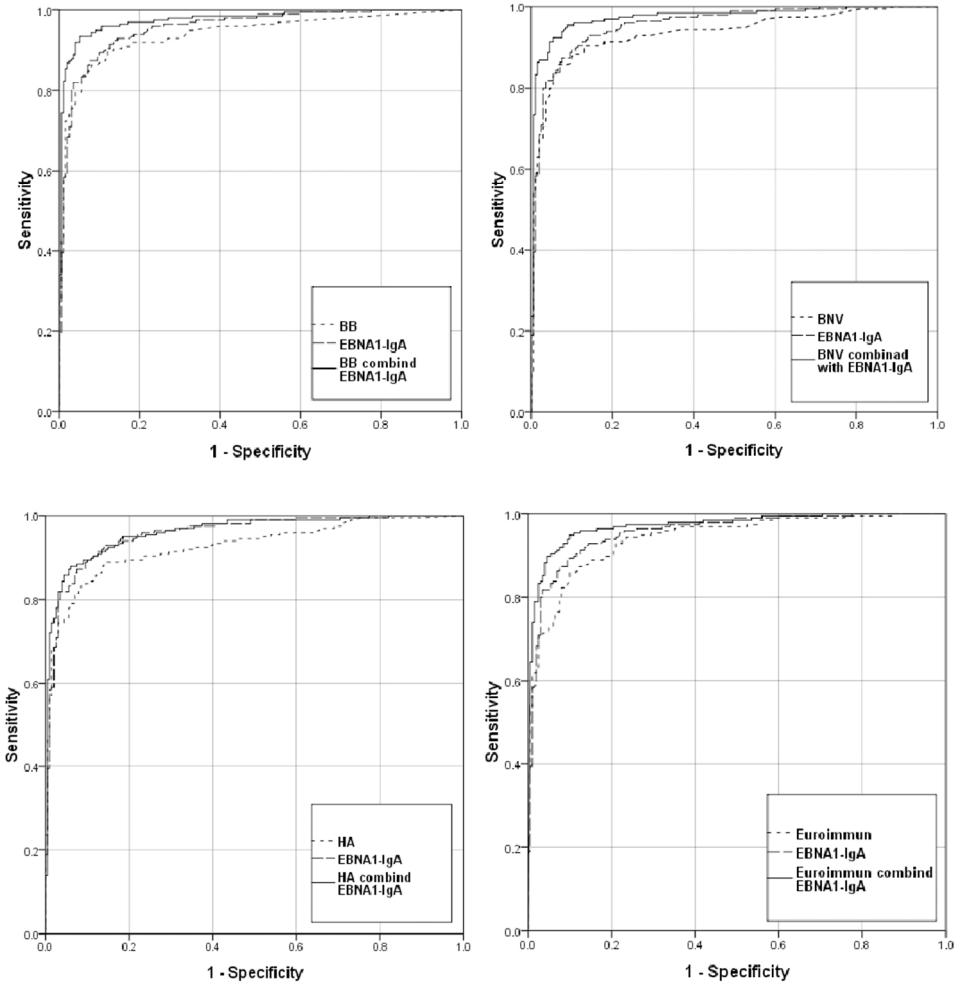

Figure1_ROCs for BB, BNV, HA, Euroimmun and their combination with EBNA1-IgA

94x96mm (300 x 300 DPI)

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)	1,2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	2,3
	4	Study objectives and hypotheses	3
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	4
Participants	**6**	Eligibility criteria	4
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	4
	8	Where and when potentially eligible participants were identified (setting, location and dates)	4
	9	Whether participants formed a consecutive, random or convenience series	4
Test methods	**10a**	Index test, in sufficient detail to allow replication	4,5
	10b	Reference standard, in sufficient detail to allow replication	4,5
	11	Rationale for choosing the reference standard (if alternatives exist)	4
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	5
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	5
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	4,5
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	4,5
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	5
	15	How indeterminate index test or reference standard results were handled	5
	16	How missing data on the index test and reference standard were handled	5
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	5
	18	Intended sample size and how it was determined	4,5
RESULTS			
Participants	**19**	Flow of participants, using a diagram	5,6
	20	Baseline demographic and clinical characteristics of participants	5,6
	21a	Distribution of severity of disease in those with the target condition	5,6
	21b	Distribution of alternative diagnoses in those without the target condition	5,6
	22	Time interval and any clinical interventions between index test and reference standard	5,6
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	6-9
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	6-9
	25	Any adverse events from performing the index test or the reference standard	6-9
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	11
	27	Implications for practice, including the intended use and clinical role of the index test	9-11
OTHER INFORMATION			
	28	Registration number and name of registry	11
	29	Where the full study protocol can be accessed	11
	30	Sources of funding and other support; role of funders	11

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.

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1,2
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	1,2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	2,3
	2b	Specific objectives or hypotheses	3
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	3,4
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	3,4
Participants	4a	Eligibility criteria for participants	3,4
	4b	Settings and locations where the data were collected	3,4
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	3,4
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	3,4
	6b	Any changes to trial outcomes after the trial commenced, with reasons	3,4
Sample size	7a	How sample size was determined	3,4
	7b	When applicable, explanation of any interim analyses and stopping guidelines	3,4
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	4
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	4
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	4
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	4
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	4

		assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	5
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	5
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	5
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	5-9
	13b	For each group, losses and exclusions after randomisation, together with reasons	5-9
Recruitment	14a	Dates defining the periods of recruitment and follow-up	8-9
	14b	Why the trial ended or was stopped	8-9
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	5,6
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	5,6
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	6-9
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	6-9
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	8-9
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	8
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	11
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	9-11
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	9-11
Other information			
Registration	23	Registration number and name of trial registry	11
Protocol	24	Where the full trial protocol can be accessed, if available	5,11
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	11

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

BMJ Open

Evaluation of seven recombinant VCA-IgA ELISA kits for the diagnosis of nasopharyngeal carcinoma in China: a case-control trial

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-013211.R2
Article Type:	Research
Date Submitted by the Author:	29-Mar-2017
Complete List of Authors:	Gao, Rui; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Wang, Lin; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Nasopharyngeal Carcinoma Zhang, Lifang; Guangzhou Women and Children\'s Medical Center, Division of Birth Cohort Study Ye, Yanfang; Sun Yat-Sen University, Sun Yat-sen Memorial Hospital Xie, Shanghang; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention Du, Jinlin; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Chen, Suihong; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Guo, Jie; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Yang, Mengjie ; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Lin, Chuyang; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention; Sun Yat-Sen University, School of Public Health Liu, Qing; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention Cao, Sumei; Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Department of Cancer Prevention
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Diagnostics, Epidemiology, Research methods
Keywords:	Nasopharyngeal carcinoma, Epstein-Barr virus, VCA-IgA, Diagnostic effect, Screening

Evaluation of seven recombinant VCA-IgA ELISA kits for the diagnosis of nasopharyngeal carcinoma in China: a case-control trial

Rui Gao[1,2*], Lin Wang[3*], Li-Fang Zhang[4] ,Yan-Fang Ye[5], Shang-Hang Xie[1], Jin-Lin Du[1,2], Sui-Hong Chen[1,2], Jie Guo[1,2], Meng-Jie Yang[1,2], Chu-yang Lin[1,2], Qing liu[1#], Su-Mei Cao[1#]

[1]Department of Cancer Prevention, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Guangzhou, 510060, P.R. China;
[2]School of Public Health, Sun Yat-Sen University, Guangzhou, 510060, P.R. China;
[3]Department of Nasopharyngeal Carcinoma, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Guangzhou, 510060, P.R. China;
[4]Division of Birth Cohort Study, Guangzhou Women and Children's Medical Center, Guangzhou, 510623, P.R. China;
[5]Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, 510120, P.R. China.
*These authors contributed equally to this work.

[#]Correspondence to
Prof. Su-Mei Cao, caosm@sysucc.org.cn;
Prof. Qing Liu, liuqing@sysucc.org.cn

Authors' E-mail:
Rui Gao: gaor5@mail2.sysu.edu.cn Lin Wang: wangl1@sysucc.org.cn
Li-Fang Zhang: lifanglydia@hotmail.com Yan-Fang Ye: yeyanfang1@gmail.com
Shang-Hang Xie: xieshh@sysucc.org.cn Jin-Lin Du: dujl3@mail2.sysu.edu.cn
Sui-Hong Chen: chensh43@mail2.sysu.edu.cn Jie Guo: guojiegina@126.com
Meng-Jie Yang: yangmj@sysucc.org.cn Chu-yang Lin: linchy23@mail2.sysu.edu.cn

Preferred reviewers:
1. Fang Fang, Fang.Fang@ki.se, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
2. Liming Liang, lliang@hsph.harvard.edu, Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA, US.
3. Qing Chen, qingchen@fimmu.com, Department of Epidemiology, School of Public Health and Tropical Medicine, Southern Medical University, Guangzhou, China

Key words: Nasopharyngeal carcinoma, Epstein-Barr virus, VCA-IgA, Diagnostic effect, Screening

Word count(Excluding References): 4168

ABSTRACT

Objective: Seven recombinant VCA-IgA ELISA kits are widely used in China, but their diagnostic effects have not been evaluated. In this study, we evaluated whether the diagnostic effects of these kits are similar to those of the standard kit (EUROIMMUN, Lübeck, Germany).

Methods: A diagnostic case-control trial was conducted, with 200 cases of nasopharyngeal carcinoma (NPC) and 200 controls from NPC-endemic areas in southern China. The areas under the curve (AUCs), the sensitivities and the specificities of testing kits were compared with those of the standard kit. The test–retest reliability of each kit was determined by intraclass correlation coefficient (ICC). Their diagnostic accuracy in combination with EBNA1-IgA was also evaluated in logistic models.

Results: Three testing kits – KSB, BB and HA – showed diagnostic accuracy equal to that of the standard kit, with good performance in the AUCs (0.926~0.945), and no significant differences in sensitivity were found between early- and advanced-stage NPCs. ICCs exceeded 0.8. Three logistic regression models were built, and the AUCs of these models (0.961~0.977) were better than those of the individual VCA-IgA kits. All new models had diagnostic accuracy equal to that of the standard kit. New cutoff values of these three kits and their corresponding combinations for researchers to replicate and use in NPC early detection and screening in the future were provided.

Conclusions: Three recombinant VCA-IgA kits – KSB, BB and HA – had diagnostic effects equal to those of the standard kit, and, in combination with EBNA1-IgA in logistic regression models, can be used in future screening for NPC.

Strengths and limitations of this study

- This is the first study to carry out a comprehensive evaluation of recombinant commercial diagnostic VCA-IgA (ELISA) kits in China, and logistic models combining VCA-IgA with EBNA1-IgA were established.
- New cutoff values for VCA-IgA kits and their corresponding combinations for researchers to replicate and use in NPC early detection and screening in the future were provided.
- All cases and controls were from NPC-endemic areas of southern China, and thus these results might not be applicable to other populations.
- Only 33 early stage NPC cases were collected. Controls were recruited from rural area, but half of the NPC cases were from urban areas.
- Cutoff values for NPC screening by means of these models described in this study must be verified in prospective mass screening.

INTRODUCTION

Nasopharyngeal carcinoma (NPC) is a common form of squamous-cell carcinoma in southern China and southeastern Asia. The annual incidence rate of NPC in southern China can reach 25 per 100,000 person-years, which is about 25-fold higher than in the

rest of the world.[1–4] NPC is a complex disease caused by a combination of Epstein-Barr virus (EBV), chronic infection, the environment and host genes in a multi-step process of carcinogenesis, but until now there have been no effective preventive measures.[5–9] Long-term survival rates differ substantially between patients with advanced- (stages III and IV) and early- (stages I and II) stage NPC. Four-year survival rates of early-stage NPC patients are 96.7% compared with 67.1% for those with advanced-stage NPC.[5] Mass screening has become the most practical method for improved early detection in and overall prognosis of NPC patients in the endemic areas.[10 11]

Serum antibodies against EBV-related antigens, especially immunoglobulin A (IgA) against viral capsid antigen (VCA-IgA), early antigen (EA-IgA), EBV nuclear antigen 1(EBNA1-IgA) and so on, remain elevated for an average of 38 months in the preclinical phase,[9–14] and serological tests for these markers are simple and inexpensive.[15–18] Therefore, since the 1970s, these tests have been used as screening markers for NPC in endemic areas. In our previous study, we evaluated the diagnostic performance of seven commercial EBV-related antibodies by enzyme-linked immunosorbent assay (ELISA) and found EBNA1-IgA (Zhongshan Biotech, China) and VCA-IgA (EUROIMMUN, Lübeck, Germany) ELISA to be the top two seromarkers, with AUCs of 0.95 (95%CI, 0.93–0.97) and 0.94 (95%CI, 0.92–0.97), respectively[16]. We further verified that the combination of VCA-IgA and EBNA1-IgA outperformed any individual EBV seromarkers, with AUC up to 0.97 (95%CI, 0.96, 0.99).[15–17] Thus, since 2011, the combination of VCA-IgA and EBNA1-IgA has been recommended as the standard tool for NPC screening in China.[18]

Nowadays, several kinds of commercial VCA-IgA kits based on recombinant peptides have been developed in China and are presently widely used for the early detection of and screening for NPC. However, their diagnostic performance for NPC alone and in combination with EBNA1-IgA has not been evaluated. In this study, we evaluated whether the effects of the NPC-diagnostic kits are comparable with those of the standard VCA-IgA kit and can be substituted for it. If so, we will further explore the combination diagnostic strategy with EBNA1-IgA for the early detection of and mass screening for NPC.

METHODS

Study population

Serum specimens were continuously collected from 200 patients hospitalised with NPC in the Sun Yat-sen University Cancer Center (SYSUCC) from January 2013 to June 2013. These cases were histologically confirmed by biopsy, and the clinical stages were classified according to the 2009 Union for International Cancer Control (UICC) criteria, including 33 patients with early-stage NPC (stages I, II) and 167 with advanced-stage NPC (stages III, IV). The inclusion criteria included being between 30 and 59 years of age and residing in one of the six high-endemic provinces

of southern China (Guangdong, Guangxi, Jiangxi, Hunan, Fujian or Hainan Province). Other information, including demographic data, smoking, drinking histories and family history of NPC, was collected by the physician in charge. All serum samples were collected before treatment.

The 200 healthy controls were randomly selected from among healthy people who participated in physical examinations at the Sihui Cancer Center (Sihui City, Guangdong Province, China) from July 2013 to September 2013 and were frequency-matched with cases by age (5-year age groups) and gender. All participants completed a short questionnaire to record demographic data, smoking, drinking histories and family history of NPC and donated 3 mL of blood.

This study was approved by the Clinical Research Ethics Committee of the SYSUCC (YB2015-029-01), and all participants provided written informed consent.

Detection of serological EBV antibodies

Serum and buffy coat were separated less than 4 hours after collection and stored at -80°C before being tested. None of the specimens was hemolytic or repeatedly frozen more than twice. Seven recombinant VCA-IgA kits, the standard VCA-IgA kit (EUROIMMUN) and the standard EBNA1-IgA kit (Zhongshan) were tested (table 1).

Table 1 Product information for eight brands of VCA-IgA kits and the EBNA1-IgA kit

Abbreviation for kits	Manufacturer
VCA-IgA	
KSB	Shenzhen Kang Sheng Bao Bio-Technology Co., Ltd.
BNV	Bioneovan Co., Ltd.
GBI	Beijing BGI-GBI Biotech Co., Ltd.
BB	Beijing Beier Bioengineering Co., Ltd.
HA	Shenzhen HuianBioscitech Co., Ltd.
HK	Shen Zhen HuaKang Co., Ltd.
ZS	ZhongShan Biotech Co., Ltd.
EUROIMMUN	EUROIMMUN Medizinische Labordiagnostika AG
EBNA1-IgA	ZhongShan Biotech Co., Ltd.

All samples were renumbered and tested blindly by one technician according to the manufacturers' instructions. Levels of antibodies were assessed by photometric measurement, which provided optical density (OD) values. Reference ODs (rOD) were obtained according to manufacturers' instructions by dividing OD values by a reference control. To investigate the test-retest reliability of each kit, 10% serum samples (40 samples) were randomly chosen and retested.

Statistical analysis

Demographic characteristics and NPC risk factors between cases and controls were compared by chi-squared tests. The cutoff value of each single kit was defined with the largest Youden Indices (sensitivity+specificity-1) chosen from each receiver operating characteristic (ROC). The diagnostic efficacy of each kit was evaluated by AUC, and non-inferiority tests based on the bootstrap approach were performed to determine whether the AUCs of these recombinant testing kits were inferior to that of the standard kit (let Δ=0.05 be the pre-determined clinically meaningful equivalence limit).[19-22] The sensitivity and specificity of each kit were calculated, and their 95% confidence intervals (CIs) were estimated by the methods of Simel and colleagues (The 95%CIs of sensitivities for early stage groups was estimated by look-up table method for binomial distribution because the sample size was less than 50).[23 24] Differences in sensitivities between early- and advanced-stage NPC with each kit were compared by Chi-squared tests (Fisher's exact test and McNemar's test will be specified while others Chi-squared tests means Person's Chi-square test). Intraclass correlation coefficients (ICC) were performed to determine test–retest reliability.

In order to prevent bias and study the virus factor only, we matched the baseline covariates (gender and age) and some of the important NPC risk factors (smoking, drinking and NPC history). Binary unconditional logistic regressions were used to establish formulae for VCA-IgA and EBNA1-IgA. The diagnostic efficacy of each formula was evaluated by sensitivity, specificity and AUC, compared with the standard formula, *Logit P* = -3.934 + 2.203VCA-IgA (EUROIMMUN) + 4.797EBNA1-IgA. The cutoff p-value in the corresponding logistic regression for distinguishing between NPC cases and controls was defined with the largest Youden Index chosen from each ROC. Two minimally acceptable false-positive rates (1-Specificity), 3% and 7%, were used empirically to establish the cutoff p-values for classifying different NPC risk subgroups.[16 17]

The non-inferiority tests were one-sided, and $p>0.05$ was considered to be non-inferior. Other tests were two-sided, and $p<0.05$ was considered to be statistically significant. Data were analyzed by SAS9.2 and SPSS16.0 software.

RESULTS

Baseline information

Baseline information on gender, age, smoking, drinking and NPC family history was comparable between cases and controls, and no statistically significant differences were found between them. Further, there were no statistically significant differences for these items between early- and advanced-stage cases (table 2).

Table 2 Demographic characteristics of NPC cases and controls

Categories	NPC Cases (N_1=200) No. (%)				Controls (N_2=200)	p[5]
	Early stage	Advanced stage	p[4]	Total		

	(n=33)	(n=167)		No. (%)		
Gender			0.472			0.417
Male	27 (81.8)	127 (76.0)		154 (77.0)	147 (73.5)	
Female	6 (18.2)	40 (24.0)		46 (23.0)	53 (26.5)	
Age (years)			0.299			0.785
30~	6 (18.2)	47 (28.1)		53 (26.5)	47 (23.5)	
40~	13 (39.4)	70 (41.9)		83 (41.5)	87 (43.5)	
50~	14 (42.4)	50 (29.9)		64 (32.0)	66 (33.0)	
Smoking[1]			0.857			0.746
Yes	11 (33.3)	53 (31.7)		64 (32.0)	61 (30.5)	
No	22 (66.7)	114 (68.3)		136 (68.0)	139 (69.5)	
Drinking[2]			0.641			0.494
Yes	6 (18.2)	25 (15.0)		31 (15.5)	27 (13.5)	
No	27 (81.8)	142 (85.0)		169 (84.5)	173 (86.5)	
NPC family history[3]			0.732			0.224
Yes	3 (9.1)	13 (7.8)		16 (8.0)	10 (5.0)	
No	30 (90.9)	154 (92.2)		184 (92.0)	190 (95.0)	

[1]'Smoking' refers to people who smoked more than one cigarette every three days within half a year and included current and former smokers.[2]'Drinking' refers to people who consumed alcoholic beverages every week within half a year and included current and former drinkers. [3]'NPC family history' refers to people whose parents, children and siblings have or did have NPC. [4]Differences in early- and advanced-stage NPC were compared by chi-squared tests (Fisher's Exact Test for NPC family history). *p<0.05 was considered as statistically significant. [5]Differences in NPC Cases and Controls were compared by chi-squared tests. *p<0.05 was considered as statistically significant.

The diagnostic accuracies of eight brands of VCA-IgA kits and the EBNA1-IgA kit

Table 3 shows that the AUCs of four kits – KSB, BB, BNV and HA – were as high as that of the standard VCA-IgA kit (AUC, 0.942; 95%CI, 0.920-0.964). The AUCs, in order, were 0.945 for KSB (95%CI 0.925-0.966), 0.940 for BB (95%CI 0.916-0.964), 0.936 for BNV (95%CI 0.911-0.961) and 0.926 for HA (95%CI 0.900-0.953). In addition, the AUCs of GBI, HK and ZS were lower than that of the standard kit. Furthermore, no significant differences were found between early- and advanced-stage NPC in the sensitivities of six kits (p>0.05), except for BNV (p=0.044).

Table 3 The diagnostic accuracies of eight brands of VCA-IgA kits and the EBNA1-IgA kit in distinguishing between NPC cases and controls

Kits	Cutoff values	Sensitivity (%)			Specificity (%)	AUC (95%CI)	P[3]
		Early stage	Advanced stage	Total	Control		

	[1]	(95%CI)	(95%CI)[2]	(95%CI)	(95% CI)		
VCA-IgA							
BB	0.58	75.8 (58.0-89.0)	88.6 (84.2-93.0)	86.5 (81.8-91.2)	92.0 (88.2-95.8)	0.940 (0.916-0.964)	0.002
BNV	0.923	72.7 (54.0-87.0)	88.0 (83.5-92.5)*	86.0 (81.2-90.8)	93.5 (90.1-96.9)	0.936 (0.911-0.961)	0.003
GBI	0.825	72.7 (54.0-87.0)	76.6 (70.8-82.5)	76.0 (70.1-81.9)	92.0 (88.2-95.8)	0.899 (0.868-0.930)	0.341*
HA	0.884	93.9 (80.0-99.0)	88.0 (83.5-92.5)	89.0 (84.7-93.3)	86.0 (81.2-90.8)	0.926 (0.900-0.953)	0.012
HK	1.218	81.8 (64.0-93.0)	83.2 (78.1-88.4)	83.0 (77.8-88.2)	89.5 (85.3-93.7)	0.913 (0.884-0.942)	0.075*
KSB	0.283	100.0 (89.0-100.0)	86.8 (82.1-91.5)	89.0 (84.7-93.3)	87.5 (82.9-92.1)	0.945 (0.925-0.966)	0.000
ZS	0.418	75.8 (58.0-89.0)	74.3 (68.2-80.3)	74.5 (68.5-80.5)	87.5 (82.9-92.1)	0.868 (0.831-0.904)	0.878*
EUROIMMUN	1.561	87.9 (72.0-97.0)	85.6 (80.8-90.5)	86.0 (81.2-90.8)	90.0 (85.8-94.2)	0.942 (0.921-0.964)	
EBNA1-IgA	1.203	93.9 (80.0-99.0)	86.2 (81.5-91.0)	87.5 (82.9-92.1)	92.5 (88.8-96.2)	0.956 (0.937-0.975)	0.000

[1]Cutoff value for NPC diagnosis was defined as the value with the largest Youden Index chosen from each ROC. [2]Differences in the sensitivities of early- and advanced-stage NPC were compared by Person Chi-Squared tests. *p<0.05 was considered as statistically significant. [3]p values were estimated by non-inferiority tests based on the bootstrap approach for AUC between EUROIMMUN and other kits. *p<0.05 was considered as statistically significant while p>0.05 was consider to be inferior to the standard kit.

The test–retest reliabilities of eight brands of VCA-IgA kits and the EBNA1-IgA kit

Ten percent serum samples (40 samples) were randomly chosen and retested for calculation of the ICC of each brand of kit, VCA-IgA or EBNA1-IgA. The test-retest reliabilities of all kits were excellent (>0.75, excellent) according to Fleiss's classification[25] (table 4).

Table 4 The test–retest reliabilities of eight brands of VCA-IgA kits and the EBNA1-IgA kit

Kits	ICC*	95%CI
VCA-IgA		
BB	0.990	0.980-0.994
BNV	0.982	0.967-0.991
GBI	0.964	0.933-0.981
HA	0.975	0.952-0.987

HK	0.876	0.764-0.935
KSB	0.823	0.666-0.906
ZS	0.978	0.958-0.988
EUROIMMUN	0.913	0.830-0.955
EBNA1-IgA	0.981	0.964-0.990

*Less than 0.40- poor; Between 0.40 and 0.59- Fair;
Between 0.60 and 0.74- good; Between 0.75 and 1.00- Excellent

The diagnostic accuracies of the combinations of VCA-IgA and EBNA1-IgA with logistic models

We chose for testing three VCA-IgA kits with high AUCs, no differences in diagnoses for early- and advanced-stage NPC and excellent test-retest reliabilities, and then combined each with the EBNA1-IgA kit by logistic models. Three logistic regression models were established:

LogitP = -3.2323 + 0.8060VCA-IgA (BB) + 1.1044 EBNA1-IgA
LogitP = -2.7591 + 0.6380VCA-IgA (HA) + 1.0620EBNA1-IgA
LogitP = -2.6039 + 0.5312VCA-IgA (KSB) + 1.1673 EBNA1-IgA

In all these models, both VCA-IgA and EBNA1-IgA were statistically significant independent predictors of NPC risk ($p<0.05$), and the AUC of each combination was statistically significantly larger than that of each single VCA-IgA ($p<0.05$). The AUC of KSB increased from 0.945 (95%CI 0.925-0.966) to 0.964 (95%CI 0.947-0.981); BB increased from 0.940 (95%CI 0.916-0.964) to 0.977 (95%CI 0.963-0.991); and HA increased from 0.926 (95%CI 0.900-0.953) to 0.961 (95%CI 0.943-0.979) (figure 1).

Table 5 shows the diagnostic accuracies of the three new combinations and the standard combination [*Logit P*=-3.934+2.203VCA-IgA (EUROIMMUN) +4.797EBNA1-IgA] in distinguishing between NPC cases and controls. The AUCs of these three combinations were as high as that of the standard combination (AUC 0.970; 95%CI 0.956-0.985) ($p<0.05$). Furthermore, no statistically significant difference was found in the sensitivity of each combination between early- and advanced-stage NPC ($p>0.05$).

We used two minimally acceptable false-positive rates (1-Specificity) of 3% and 7% to define the high-risk and medium-risk cutoff values for the new combinations. The corresponding logistic regression p-values were 0.707 and 0.232 for BB, 0.766 and 0.364 for HA, 0.831 and 0.384 for KSB, and the corresponding true-positive rates (Sensitivities) were 88.0% and 93.5% for BB, 78.0% and 88.0% for HA and 79.0% and 87.5% for KSB.

Table 5 The diagnostic accuracies of three new combinations and the standard combination in distinguishing between NPC cases and controls

Combination	New cutoff values[1]	Sensitivity (%)			Specificity (%)	AUC (95%CI)	P[3]
		Early stage (95%CI)	Advanced stage (95%CI)[2]	Total (95%CI)[3]	Control (95% CI)[4]		
BB+EBNA1-IgA	0.258	97.0 (85.0-100)	92.8 (89.2-96.4)	93.5 (90.1-96.9)	95.0 (92.0-98.0)	0.977 (0.963-0.991)	<0.001
HA+EBNA1-IgA	0.379	97.0 (85.0-100)	86.2 (81.5-91.0)	88.0 (83.5-92.5)	94.0 (90.7-97.3)	0.961 (0.943-0.979)	<0.001
KSB+EBNA1-IgA	0.191	93.9 (80-100)	94.6 (91.5-97.7)	94.5 (91.3-97.7)	87.0 (82.3-91.7)*	0.964 (0.947-0.981)	<0.001
Standard combination	0.998	97.0 (85-100)	88.6 (84.2-93.0)	90.0 (85.8-94.2)	95.5 (92.6-98.4)	0.970 (0.956-0.985)	

[1]New cutoff value for NPC diagnosis was defined as the value with the largest Youden Index chosen from each ROC. [2]Differences in the sensitivity of early- and advanced-stage NPC were compared by Person Chi-Squared tests. *p<0.05 was considered as statistically significant. [3] p-values were estimated by non-inferiority tests based on the bootstrap approach for AUC between new combinations and the standard combination. *p<0.05 was considered as statistically significant while p>0.05 was consider to be inferior to the standard kit.

DISCUSSION

In our study, seven recombinant VCA-IgA kits were evaluated, and of these, KSB, BB and HA had diagnostic effects as good as those of the standard kit in terms of sensitivity, specificity and AUC. Combining VCA-IgA with EBNA1-IgA by logistic regression models increased the diagnostic accuracy of these three kits, and all combinations performed as well as the standard combination in sensitivity, specificity and AUC. This is the first study to carry out a comprehensive evaluation of recombinant commercial diagnostic VCA-IgA (ELISA) kits in China, and logistic models combining VCA-IgA with EBNA1-IgA were established. Furthermore, new cutoff values for these VCA-IgA kits and their corresponding combinations for researchers to replicate and use in NPC early detection and screening in the future were provided.

In this study, we first calculated the diagnostic performance of each brand of VCA/IgA kit. The AUC of the standard VCA-IgA kit (EUROIMMUN) was 0.942 (95%CI 0.920-0.964), which was consistent with results from our previous studies and verified that the diagnostic performance of VCA-IgA was good and stable.[16 17] We also found that the sensitivities, specificities and AUCs of three kits – KSB, BB and HA – were as high as those of the standard kit, and no significant differences in sensitivity were found between early- and advanced-stage NPC. Moreover, all test–retest reproducibilities were excellent (>0.75) and the CVs of Differences Values of test and retest result of all assays were shown in Supplementary Table2. These results suggested that these three kits had equal diagnostic effects and can be substituted for the standard kit. The costs of these recombinant commercial diagnostic

kits were only half that of the standard kit, making them more cost-effective.

The EBV capsid antigen (VCA) is a late protein produced in the EBV lytic infection period. VCA contains a batch of capsid proteins, such as VCA-p18 (BFRF3), VCA-p23 (BLRF2), gp125/110 (BALF4) and so on, which have unique immune dominants and virus-specific antigenic domains. These domains contain several small peptide regions (epitopes) which can be combined to form a powerful diagnostic reagent for VCA-IgA.[26-27] The capsid proteins in the EUROIMMUN kit[28] were extracted from the pyrolysis products of human B lymphocytes (P3HR1 cell line) infected by EBV and contained a combined native capsid protein of EBV. We noticed that, in contrast to the standard kit with a combined native capsid protein, these testing kits contain primarily recombinant p18 capsid proteins (VCA-p18). VCA-p18 is a small capsid protein that contains several small peptide regions (epitopes) which can be combined to form a powerful diagnostic reagent for VCA-IgA antibody responses.[27] Some researchers have reported that VCA-p18 is the major VCA antigen for IgA responses.[27 29] Our study showed that the AUCs of these VCA-p18 recombinant kits were more than 0.85, and three of them had the same diagnostic effects as the standard kit, suggesting that, although the manufacturing processes of some recombinant VCA-p18 kits still need to be improved, some of the recombinant kits can be substituted for the standard kit for NPC diagnosis.

As the serum antibody level (rOD) provides continuous data, the cutoff value for distinguishing between NPC cases and controls is critical for the early detection of and screening for NPC. A reasonable cutoff value can balance sensitivity and specificity. In the early detection of and screening for NPC, high sensitivity is required for the identification of high-risk individuals, and high specificity is required for reducing the rate of misdiagnosis and associated costs. According to the cutoff values provided by the kits' instructions, the sensitivities of testing kits were always too low, whilst their specificities were always too high. For example, the sensitivity and the specificity of KSB are 0.780 and 0.925, respectively, suggesting that the old cutoff values should be adjusted (Supplementary Table1). We established new cutoff values for distinguishing between NPC cases and controls by Youdon Indices, and then obtained reasonable sensitivities and specificities. After adjustment, the new cutoff value for KSB is 0.283, and the sensitivity and specificity are 0.890 and 0.875, respectively. Moreover, no differences were found between sensitivities and specificities of these three kits – KSB, BB and HA – and those of the standard kit. Due to the low percentage (less than 20%) of early stage in clinic, we can only collect 33 early stage NPC participants in our study. Analyzing the sensitivities by pooling early and late stage together only was not appropriate. So we did subgroup analysis and found there were also no statistically significant differences in the sensitivities of these three kits for early- and advanced-stage NPC. Furthermore, there were no differences between the early stage sensitivities of these three kits and that of the standard kit too (0.202 for BB. 0672 for HA, 0.112 for KSB).

As for the standard VCA, we found that the combinations of VCA-IgA and EBNA1-IgA by logistic models increased the diagnostic accuracy for NPC from less than 0.946 to more than 0.961 in AUCs. Sensitivities and specificities also increased. For example, the sensitivity and specificity of BB increased from 0.865 and 0.920 to 0.935 and 0.955, respectively. VCA-IgA and EBNA1-IgA are two antibodies corresponding to EBV lytic-cycle proteins and latency gene products, respectively. Therefore, it is reasonable that host antibody responses for lytic-cycle and latency-associated EBV-related proteins can be complementary to each other in the diagnosis of NPC, and the combination of both could increase NPC diagnostic accuracy.[11 30] Furthermore, these three new combinations had diagnostic effects in sensitivities (including subgroup analysis), specificities and AUCs equal to those of the standard combination, suggesting that the combinations of the three recombinant kits can be used for the early detection of and diagnostic screening for NPC. In this study, the control individuals came from NPC-endemic areas and belonged to a screening target population, so we attempted to define people at different risk levels by these new combinations for NPC screening. Compared with other common diseases, the NPC incidence rate in the screening target population was relatively low (about 50 per 100,000 person-years)[2 10 31] Thus, it is important that the false-positive rate be small enough to avoid unnecessary fiberoptic endoscopy/biopsies and psychological stress for the NPC screening participants. Conversely, the true-positive rate (equal to sensitivity) should be acceptable.[32] We used two minimally acceptable false-positive rates of 3% and 7% as the high-risk and medium-risk cutoff values, respectively,[17] and the corresponding true-positive rates (sensitivities) for these three kits were 78.0% to 88.0% and 87.5% to 93.5%,[5] respectively. If the baseline serologic results fulfilled the definition of high risk, the participants were referred for diagnostic examinations, and different screening intervals were assigned to the high-, medium- and low-risk groups. The screening intervals for these groups are 1, 1 and 4 years, respectively.[17]

The study had some limitations. First, this study was a Single-center study and all cases and controls were from NPC-endemic areas of southern China (controls were from hospital); therefore, these results might not be applicable to other populations. Second, Due to the low percentage (less than 20%) of early stage in clinic, we can only collect 33 early stage NPC participants in our study period. But the phenomenon also indicated that, most patients are typically not detected until NPC is in an advanced stage. Founding out such people was also very meaningful in real life. Third, controls were recruited from rural area, but half of the NPC cases were from urban areas (rural:urban=95:105). Though no evidence showed that there were different infection rate between rural and urban people, it might cause some other unknown bias. Fourth, it was a diagnostic trial in case-control design, and new cutoff values of these new schemes for NPC screening from this study must be verified in prospective mass screenings.

CONCLUSIONS

Three recombinant VCA-IgA kits – KSB, BB and HA – had diagnostic effects equal to those of the standard kit. They can be substituted for the standard kit, and their combinations could be used in the early detection of and screening for NPC.

Contributors Su-Mei Cao and Qing Liu obtained the funding. Rui Gao, Lin Wang and Li-Fang Zhang contributed to study conception and design; Rui Gao conducted experiments; Lin Wang, Yan-Fang Ye, Jin-Lin Du, Shang-Hang Xie, Sui-Hong Chen, Jie Guo and Meng-Jie Yang, Chu-yang Lin acquired or cleaned the data; Rui Gao analysed and interpreted the data; Rui Gao, Su-Mei Cao and Qing Liu drafted or revised the manuscript. All authors critically revised the manuscript for important intellectual content. All authors read and approved the final version of the manuscript.

Funding This work was supported by the National Natural Science Foundation (No. 81373068) and the Special Fund for Public Health Trade (No. 201202014).

Competing interests The authors declare no conflicts of interest.

Ethics approval This study was approved by the Sun Yat-sen University Cancer Center IRB (YB2015-029-01).

Data sharing statement Technical appendix, statistical code, and dataset available from the corresponding author at caosm@sysucc.org.cn and liuqing@sysucc.org.cn.

REFERENCES

1. Ma J, Cao S. The epidemiology of nasopharyngeal carcinoma. In: Lu JJ, Cooper JS, Lee AWM, eds. *Nasopharyngeal cancer*. Berlin: Springer, 2010:1–8.
2. Jia W-H, Huang Q-H, Liao J, *et al*. Trends in incidence and mortality of nasopharyngeal carcinoma over a 20–25 year period (1978/1983–2002) in Sihui and Cangwu counties in southern China. *BMC Cancer* 2006;6:178–85.
3. Chang ET, Adami HO. The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev* 2006;15:1765–77.
4. Cao SM, Simons MJ, Qian CN. The prevalence and prevention of nasopharyngeal carcinoma in China. *Chinese J Cancer* 2011;30:114–9.
5. Sarmiento MP, Mejia MB. Preliminary assessment of nasopharyngeal carcinoma incidence in the Philippines: a second look at published data from four centers. *Chinese J Cancer* 2014;33:159–64.
6. Jia WH. Complex segregation analysis of nasopharyngeal carcinoma in Guangdong, China: evidence for a multifactorial mode of inheritance (complex segregation analysis of NPC in China). *Eur J Hum Genet* 2005;13:248–52.
7. Yu MC, Yuan JM. Epidemiology of nasopharyngeal carcinoma. *Semin Cancer Biol* 2002;12:421–29.
8. Xue WQ, Qin HD, Ruan HL, *et al*. Quantitative association of tobacco smoking with the risk of nasopharyngeal carcinoma: a comprehensive meta-analysis of studies conducted between 1979 and 2011. *Am J Epidemiol* 2013;178:325–38.

9. Chien YC, Chen JY, Liu MY, *et al*. Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men. *N Engl J Med* 2001;345:1877–82.
10. Cao SM, Liu Z, Jia WH, *et al*. Fluctuations of Epstein-Barr virus serological antibodies and risk for nasopharyngeal carcinoma: a prospective screening study with a 20-year follow-up. *PLoS One* 2011;6:e19100.
11. Zeng Y, Zhong JM, Li LY, *et al.* Follow-up studies on Epstein-Barr virus IgA/VCA antibody-positive persons in Zangwu County, China. *Intervirology* 1983;20:190–4.
12. Chen M, Liu M, Hsu S, *et al*. Use of bacterially expressed EBNA-1 protein cloned from a nasopharyngeal carcinoma (NPC) biopsy as a screening test for NPC patients. *J Med Virol* 2001;64:51–7.
13. Zong YS, Sham JS, Ng MH, *et al*. Immunoglobulin A against viral capsid antigen of Epstein-Barr virus and indirect mirror examination of the nasopharynx in the detection of asymptomatic nasopharyngeal carcinoma. *Cancer* 1992;69:3–7.
14. Yi Z, Yuxi L, Chunren L, *et al*. Application of an immunoenzymatic method and an immunoautoradiographic method for a mass survey of nasopharyngeal carcinoma. *Intervirology* 1980;13:162–8.
15. Tang B, Liu Q, Cui Y, *et al*. Protocols of screening, early detection and diagnosis for nasopharyngeal carcinoma. In: Dong Z, Qiao Y, eds. *Protocols of screening, early detection and diagnosis for cancer in P.R.China*. Beijing: People's Medical Publishing House of P.R. China, 2009:187–8.
16. Liu Y, Huang QH, Liu WL, *et al*. Establishment of VCA and EBNA1 IgA-based combination by enzyme-linked immunosorbent assay as preferred screening method for nasopharyngeal carcinoma: a two-stage design with a preliminary performance study and a mass screening in southern China. *Int J Cancer* 2012;131:406–16.
17. Liu ZW, Ji MF, Huang QH, *et al*. Two Epstein-Barr virus–related serologic antibody tests in nasopharyngeal carcinoma screening: results from the initial phase of a cluster randomized controlled trial in Southern China. *Am J Epidemiol* 2013;177:242–50.
18. Department of disease prevention and control, Ministry of health, China. Technical proposal for early diagnosis and treatment of cancer (2011) (Chinese). Peking: People's Medical Publishing House (PMPH) 2011:144-71.
19. Obuchowski N. Testing for equivalence of diagnostic tests. *Am JRadiol* 1997;168:13–7.
20. Liu JP, Ma MCh, Wu ChY, *et al*. Tests of equivalence and noninferiority for diagnostic accuracy based on the paired areas under ROC curve. *Statist Med* 2006;25:1219–38.
21. Zhou XH, Obuchowski NA, McClish DK. *Statistical methods in diagnostic medicine.* New York: Wiley, 2002:188–92.
22. Chen WZ, Zhang JY. Application of non-inferiority test for diagnostic accuracy under the areas of ROC based on bootstrap approach and its macro programming development. *Modern Prev Med (Chinese)* 2010;37:3009–10.
23. Simel D, Samsa G, Matchar D. Likelihood ratios with confidence: sample size

estimation for diagnostic test studies. *J Clin Epidemiol* 1991;44:763–70.
24. Medical Statistics (Third Edition) (Chinese). Peking: People's Medical Publishing House (PMPH) 2010:93,710.
25. Fleiss JL. *Reliability of measurement*. The designand analysis of clinical experiments, 1st edn. New York: Wiley-Interscience, 1986:1–32.
26. Tang JW, Rohwader E, Chu IM, *et al*. Evaluation of EpsteinBarr virus antigen-based immunoassays for serological diagnosis of nasopharyngeal carcinoma. *J Clin Virol* 2007;40:284–8.
27. Middeldorp JM. Epstein-Barr virus-specific humoral immune responses in health and disease. *Curr Top Microbiol Immunol* 2015;391:289–323.
28. Mao YP, Li WF, Chen L, *et al*. A clinical verification of the Chinese 2008 staging system for nasopharyngeal carcinoma. *Ai Zheng*(*Chinese*). 2009;28:1022–8.
29. Van Grunsven WM, Spaan WJ, Middeldorp JM. Localization and diagnostic application of immunodominant domains of the BFRF3-encoded Epstein-Barr virus capsid protein. *J Infect Dis* 1994;170:13–9.
30. Fachiroh J, Paramita DK, Hariwiyanto B, *et al*. Single assay combination of Epstein-Barr virus (EBV) EBNA1- and viral capsid antigen-p18-derived synthetic peptides for measuring anti-EBV Immunoglobulin G (IgG) and IgA antibody levels in sera from nasopharyngeal carcinoma patients: options for field screening. *J Clin Microbiol* 2006;44:1459–67.
31. Wei K, Xu Y, Liu J, *et al*. No incidence trends and no change in pathological proportions of nasopharyngeal carcinoma in Zhongshan in 1970–2007. *Asian Pacif J Cancer Prev* 2010;11:1595–9.
32. Pepe MS, Feng Z, Janes H, *et al.* Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst* 2008;100:1432–8.

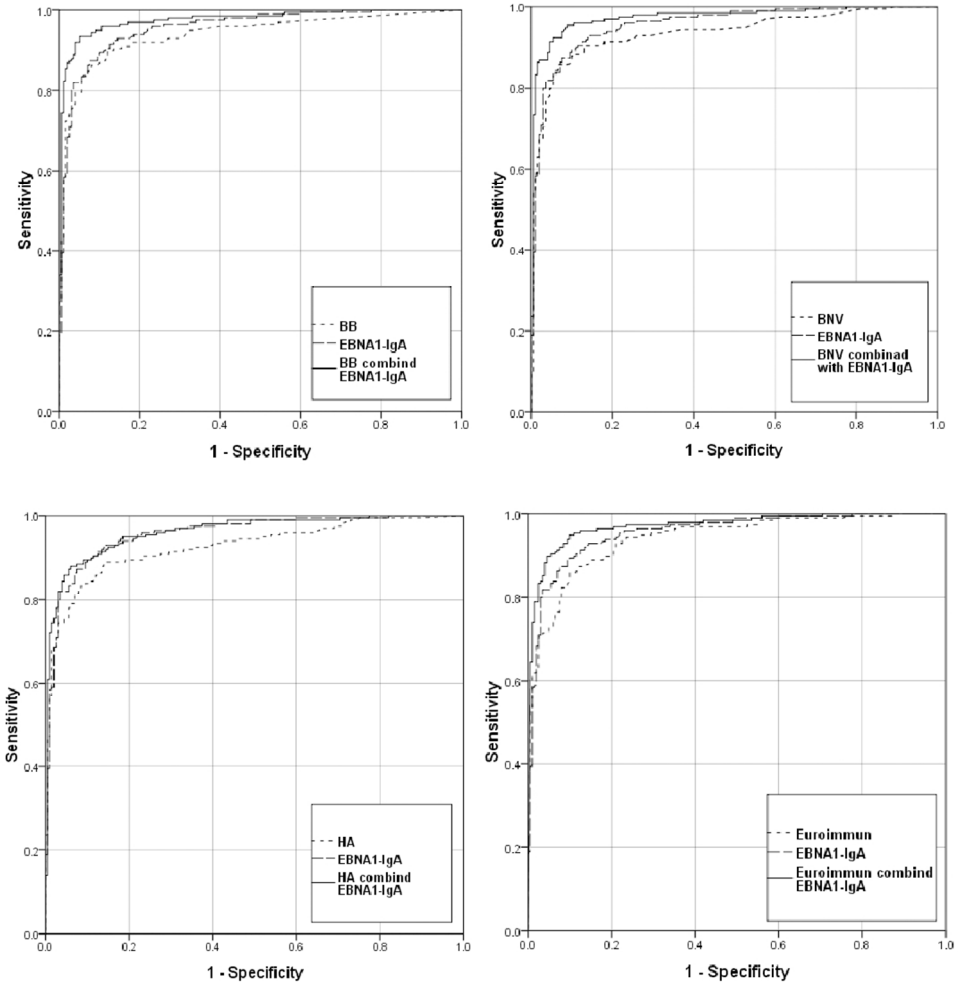

Figure1_ROCs for BB, BNV, HA, Euroimmun and their combination with EBNA1-IgA

94x96mm (300 x 300 DPI)

Supplementary Tables

Table 1. Sensitivities and Specificities based on manufacturers' cutoffs of VCA-IgA kits and the EBNA1-IgA kit

Kits	**Sensitivity (%)**(95%CI)			**Specificity (%)** (95% CI)
	Early stage (95%CI)	Advanced stage (95%CI)	Total (95%CI)	
VCA-IgA				
BB	69.7(51.0-84.0)	85.6(80.8-90.5)	83.0 (77.8-88.2)	94.5 (91.3-97.7)
BNV	72.7(54.0-87.0)	86.8(82.1-91.5)	84.5 (79.5-89.5)	94.0 (90.7-97.3)
GBI	69.7(51.0-84.0)	71.9(65.6-78.1)	71.5 (65.2-77.8)	93.0 (89.5-96.5)
HA	90.9(76.0-98.0)	85.6(80.8-90.5)	86.5 (81.8-91.2)	87.0 (82.3-91.7)
HK	84.8(68.0-95.0)	85.0(80.1-90.0)	85.0 (80.1-89.9)	85.5 (80.6-90.4)
KSB	81.8(64.0-93.0)	77.2(71.4-83.1)	78.0 (72.3-83.7)	92.5 (88.8-96.2)
ZS	57.6(39.0-74.0)	59.3(52.5-66.1)	59.0 (52.2-65.8)	95.5 (92.6-98.4)
Euroimmun	97.0(85.0-100.0)	90.4(86.3-94.5)	91.5 (87.6-95.4)	79.5 (73.9-85.1)
EBNA1-IgA	93.9(80.0-99.0)	88.6(84.2-93.0)	89.5 (84.7-93.3)	90.0 (85.8-94.2)

Table 2. CVs of Differences Values of test and retest result of VCA-IgA kits and the EBNA1-IgA kit

Kits	**CV**
VCA-IgA	
BB	0.69
BNV	0.65
GBI	1.00
HA	0.34
HK	1.55
KSB	0.86
ZS	1.24
Euroimmun	0.63
EBNA1-IgA	0.55

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)	1,2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	2,3
	4	Study objectives and hypotheses	3
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	4
Participants	**6**	Eligibility criteria	4
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	4
	8	Where and when potentially eligible participants were identified (setting, location and dates)	4
	9	Whether participants formed a consecutive, random or convenience series	4
Test methods	**10a**	Index test, in sufficient detail to allow replication	4,5
	10b	Reference standard, in sufficient detail to allow replication	4,5
	11	Rationale for choosing the reference standard (if alternatives exist)	4
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	5
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	5
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	4,5
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	4,5
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	5
	15	How indeterminate index test or reference standard results were handled	5
	16	How missing data on the index test and reference standard were handled	5
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	5
	18	Intended sample size and how it was determined	4,5
RESULTS			
Participants	**19**	Flow of participants, using a diagram	5,6
	20	Baseline demographic and clinical characteristics of participants	5,6
	21a	Distribution of severity of disease in those with the target condition	5,6
	21b	Distribution of alternative diagnoses in those without the target condition	5,6
	22	Time interval and any clinical interventions between index test and reference standard	5,6
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	6-9
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	6-9
	25	Any adverse events from performing the index test or the reference standard	6-9
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	11
	27	Implications for practice, including the intended use and clinical role of the index test	9-11
OTHER INFORMATION			
	28	Registration number and name of registry	11
	29	Where the full study protocol can be accessed	11
	30	Sources of funding and other support; role of funders	11

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.

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1,2
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	1,2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	2,3
	2b	Specific objectives or hypotheses	3
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	3,4
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	3,4
Participants	4a	Eligibility criteria for participants	3,4
	4b	Settings and locations where the data were collected	3,4
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	3,4
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	3,4
	6b	Any changes to trial outcomes after the trial commenced, with reasons	3,4
Sample size	7a	How sample size was determined	3,4
	7b	When applicable, explanation of any interim analyses and stopping guidelines	3,4
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	4
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	4
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	4
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	4
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	4

		assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	5
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	5
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	5
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	5-9
	13b	For each group, losses and exclusions after randomisation, together with reasons	5-9
Recruitment	14a	Dates defining the periods of recruitment and follow-up	8-9
	14b	Why the trial ended or was stopped	8-9
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	5,6
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	5,6
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	6-9
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	6-9
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	8-9
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	8
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	11
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	9-11
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	9-11
Other information			
Registration	23	Registration number and name of trial registry	11
Protocol	24	Where the full trial protocol can be accessed, if available	5,11
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	11

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.