SCREENING TEST ACCURACY STUDIES: HOW VALID ARE OUR CONCLUSIONS?

Application to visual inspection methods for cervical screening

Cédric Mahé[1], Lynne Gaffikin[2,3]

[1]International Agency for Research on Cancer, Lyon, France

[2]JHPIEGO corporation, Baltimore, MD, USA; [3]Evaluation and Research Technologies for Health Inc, Baltimore, MD, USA

Visual inspection with acetic acid (VIA)

Rational

Usefulness of the Pap test in reducing cervical cancer mortality is generally acknowledged in countries with well-organized prevention programs, but successful implementation is challenging in low-resource settings.

Definition

Naked- eye inspection of the cervix one minute after the application of a 3–5% dilute solution of acetic acid. The cervix is examined using a bright light source such as a torch or halogen focus lamp. Test positivity is based on the appearance of acetowhite areas in the transformation zone (TZ), close to the squamocolumnar junction (SCJ) or the os.

Advantage

- Less equipment and no laboratory facilities required
- No highly specialized training required
- Results are immmediately available leaving room for immediate treatment

Drawback

- Subjectivity: need for training and quality control

To date, has VIA been properly evaluated?

Essential pre-requisite

- The test should be convenient to, safe for and acceptable by target community members.
- The test should be relatively inexpensive both to obtain and provide

Reliability

Test reliability assesses the degree to which repeated measurements of the test (by the same provider or not) yields the same result.

A visual test will likely be less reliable

Key factor affecting reliability:

- Initial training (needs for competency-based training)
- Maintainance of the skills post-training
- Number of providers involved
- Level of comfort making definitive statement (depend on the case management protocol related to positive test results)
- Number of steps involved in testing

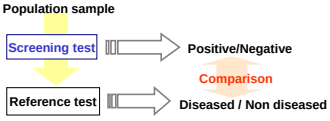
To ensure reproducibility and accuracy of the findings, test reliability should be assessed before any evaluation of test characteristics.

Test characteristics

Accuracy is measured by specificity and sensitivity

Sensitivity and specificity measure the ability of a test to correctly *identify* people as diseased or non-diseased, respectively. In principle, they do not vary according to disease prevalence (can be compared between studies)

Estimation using cross-sectional study



Pitfalls in the evaluation of screening tests

Test	Early disease: Present (+)	Absent (-)	
+	a	b	a+b
-	c	d	c+d
	a+c	b+d	

Sensitivity Se = a/a+c

Specificity Sp = d/b+d

Study sample size requirements are often neglected

Example: If we assume a disease prevalence of 1% and we expect an estimated sensitivity and specificity of 50% with a precision of +/- 5% (10% width), we will need to enroll about 9,600 subjects!!!!!!

Validity of test accuracy studies

Internal validity: credibility of results

Reference standard application

- Final disease status should be obtained for all subjects (work-up or verification bias):
- Both the new test and the reference test must be performed as close as possible in time
- Test results must be determined independently of previous results or other clinical information known to be associated with disease risk (information bias)

Reference standard definition

- The reference test should be accurate, to prevent misclassification bias (meaning the test may "misclassify" some people into the wrong disease category).

The sensitivity will probably be inflated when using two tests that rely on the visualizing the cervix (e.g. visual inspection and colposcopy) because these two tests will probably miss the same lesions.

Using a combination of colposcopy and biopsy reduces overall study verification bias, but introduces some misclassification bias (assuming colposcopy is not as accurate as biopsy) and there is some residual verification bias (resulting from the fact that colposcopy negatives do not get biopsied).

Spectrum of the disease

When the full "spectrum" of the disease (e.g. normal to cervical cancer) is not represented, estimates may be distorted because individuals with levels close to the test cut-point are more likely to be misclassified.

External validity: applicability to other population

Participant characteristics

- Visual-based tests are more difficult to perform well in women with a high rate of sexually transmitted diseases or in older, menopausal women
- People participating can be assumed to be more "health conscientious" and therefore not representative of the whole population at risk (selection bias)

Site characteristic

- High caseload (referral or STD clinics) or more representative low case-load (primary care, family planning clinics or mobile clinics)?

Drop outs or indeterminate results

- No systematic bias affecting who drops out and does not show up for the reference test or when the test is indeterminate

Failed or uninterpretable results should be included in the analysis or the report, because such results may affect interpretation of the test's usefulness.

Standardization of test positivity

Denomination	Possible thresholds		Characteristics
Normal	negative	negative	Normal looking cervix: no white lesion, smooth, uniform, featureless Atypical cervix: ectopion, polyp, cervicitis, inflammation, Nabothian cysts
Indeterminate	positive	negative	Severe inflammation or cervicitis so that cervix cannot be adequately assessed for acetowhite lesion
Ill-definite lesion	positive	positive	Pale white lesion (acetowhite lesion), poorly circumscribed and faintly acetowhite Focal, small punctuated areas of acetowhitening usually involving the TZ
Definite lesion	positive	positive	Dense white lesion with sharp border; one border abutting the SCJ
Suspicious cancer	positive	positive	Cervical ulcer or growth cauliflower-like growth or ulcer Fungation mass

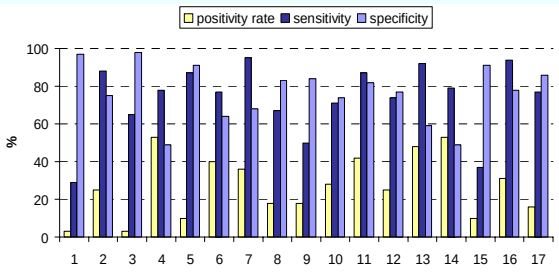
Case study: published test characteristics on VIA

To date, 17 cross-sectional studies published BUT

- **Sensitivity estimates range from 29% to 95% and specificity from 49% to 91%.**
- **Even within a study involving 55,000 women from 11 sites with a common training, protocol and test definition, substantial variability in positivity rates (7–27%), sensitivity (56–94%) and specificity (74–94%) was observed.**

The heterogeneity of the estimates from the 17 studies can be explained, in part, by:

- **The lack of standardized test definition:** this may explain the wide range of test positivity reported (3 to 53%)
- **The number, background skills and experience of the providers**: providers involved ranged from midwives to gynecologists with varying levels of training; and the number of providers involved was rarely specified.
- **The disease spectrum**: some studies involved symptomatic/ referred women, i.e. with a higher rate of cervical abnormalities.
- **Underlying prevalence of STD/cervical disease**: this level can be very different between developed and developing countries.
- **Target age group**: ages ranged from 15 to 83 years while the test is mainly suitable for women 25-55.
- **The choice of the gold standard**: it varies from 4-quadrant biopsies to colposcopy only.
- **The application of the gold standard**: only 7 of these studies were associated with minimal or no verification bias by design (gold standard applied to all screened women irrespective of the test result).
- **The quality/accuracy of the disease definition**: varying abilities of colposcopists to detect lesions (especially less apparent ones) and of pathologists to interpret histology accurately affect greatly the assessment of test performance
- **The absence of blinding**: even though it may be specified by design, blinding between VIA and colposcopy, performed during the same visit, is often difficult operationally and this may introduce information bias.



Comparison of the positivity rate, sensitivity and specificity for the 17 published studies

(see table below for details)

	Study	Sample size	Population (age, recruitment, location)	Provider	Reference diagnosis	Positivity rate	Sensitivity (95%CI)	Specificity (95%CI)	Comments
1	Slawson *et al.* (1992)	2,690	15-45, Family practice,USA	Clinician	Colposcopy/histology for Pap+ or VIA+	3	29 (14-48)	97 (96-98)	Not designed for test accuracy estimation
2	Cecchini *et al.* (1993)	2,036	17-83, Opportunistic, Italy	Smear taker	Colposcopy/histology for Pap+, VIA+ or cervicography+	25	88 (47-99)	75 (73-77)	Not designed for test accuracy estimation
3	Megevand *et al.* (1996)	2,426	20-83, Opportunistic, South-Africa	Nurse	Colposcopy/histology for Pap+ or VIA+	3	65 (45-81)	98 (97-98)	Not designed for test accuracy estimation
4	Londhe *et al.* (1997)	372	? , Opportunistic, India	Clinician	Colposcopy	53	78 (56-93)	49 (43-54)	Only 74% of the patient underwent colposcopy
5	Sankaranarayanan *et al.* (1998)	2,935	20+, Opportunistic, India	Cytotech	Colposcopy/histology for Pap+ or VIA+	10	87 (75-95)	91 (90-92)	Not designed for test accuracy estimation
6	Univ. of Zimbabwe / JHPIEGO (1999)	2,148	25-55, Opportunistic, Zimbabwe	Midwife	Colposcopy/histology	40	77 (70-82)	64 (62-66)	
7	Sankaranarayanan *et al.* (1999)	1,268	22-70, Opportunistic,India	Nurse	Colposcopy/histology for Pap+ or VIA+	36	95 (88-99)	68 (65-70)	Not designed for test accuracy estimation
8	Denny *et al.* (2000)	2,944	35-65, Opportunistic, South Africa	Nurse	Colposcopy/histology for Pap+, VIA+,HPV+ or cervicography+	18	67 (56-77)	83 (82-85)	Minimal verification bias
9	Cronjé *et al.* (2001)	6,298	Mean 34, Opportunistic, South Africa	Nurse	Histology for VIA+ and 20% of VIA-	18	50 (44-55)	84 (83-85)	Unbiased estimation not possible from the paper data
10	Belinson *et al.* (2001)	1,997	35-45, Opportunistic, China	Gynecol.	4-quadrant histology + endo-cervical curetage	28	71 (60-80)	74 (72-76)	
11	Singh *et al.* (2001)	402	Mean 37, Referred, India	Gynecol.	Colposcopy/histology	42	87 (81-92)	82 (76-86)	
12	Denny *et al.* (2002)	2,754	35-65,Opportunistic, South Africa	Nurse	Colposcopy/histology for Pap+, HPV+, VIA+ or cervicography+	25	74 (64-82)	77 (75-78)	Minimal verification bias
13	Rodriguez-Reyes *et al.* (2002)	376	19-45, Opportunistic, Mexico	?	Histology	48	92 (81-98)	59 (53-64)	
14	Cronjé *et al.* (2003)	1,093	21-65, Opportunistic, South Africa	Nurse	Histology	53	79 (?-?)	49 (?-?)	
15	Ngelangel *et al.* (2003)	3,316	25-65, Opportunistic,Philippine	Nurse	Colposcopy/histology	10	37 (27-49)	91 (90-92)	
16	Tayyeb *et al.* (2003)	501	30-60, Opportunistic, Pakistan	?	Colposcopy/histology for Pap+ or VIA+	31	94 (85-98)	78 (74-82)	Verification bias
17	Sankaranarayanan *et al.* (2004)	54,981	25-65, Opportunistic,India /Africa (11studies)	Midwife	Colposcopy/histology	16 [R=:7-27]	77 (74-79) [R=:56-94]	86 (85-86) [R=:74-94]	Pooled analysis of 11 studies (R: range within the studies)

Some test characteristics of the table are not the ones reported in corresponding publications. Estimates have been computed when they were not provided or have been corrected to achieve comparability between studies. This correction was performed to take into account differences in study design or analysis, due to various factors: different threshold of test positivity, different disease definition, only a subset of the population used for characteristics estimation, improper computation method, etc.

This work was funded by the Bill & Melinda Gates Foundation through the Alliance for Cervical Cancer Prevention (ACCP)