Published in final edited form as:

Clin Cancer Res. 2011 November 15; 17(22): 6971–6972. doi:10.1158/1078-0432.CCR-11-1862.

DNA cytometry testing for cervical cancer screening: approaches and reporting standards for new technologies

Graham A Colditz, MD and John Crowley, PhD

Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine in St. Louis and Alvin J. Siteman Cancer Center

Graham A Colditz: colditzg@wustl.edu

In the journal a report by Tong suggested that DNA cytometry might be beneficial as a screening strategy compared to conventional cytology in developing settings (Tong, Shen et al. 2009). For their study, the authors complete a flow diagram showing that from 25,000 women evaluated for eligibility, 22,993 underwent randomization. Individual women were randomized to the order of testing; group 1 had cytometry followed by cytology, and group 2 had cytology followed by cytometry. All women at entry had a cervical scrape with Cervix brush and smears were classified according to the Bethesda System terminology. DNA cytometric assessments were also conducted at the same initial visit. In the primary report data were presented for all women regardless of the order of testing when rates of positive test results and of cancer were presented (in their Table 2). The primary outcome of the trial was diagnosis of cervical cancer through the screening process. Secondary outcomes included CIN 1 to 3 and death. Of the women almost all received both tests but 48 women received cytological testing only and 46 received DNA cytometry only. This use of both tests on almost all women adds difficulty to the interpretation of the trial results when the primary presentation ignores order of testing and the cancer endpoint was reported as a rate per all women receiving the specific test (21,471 DNA cytometry; 21,693 cytologic testing). The authors report one year follow-up of participants by telephone. It is unclear what, if any, data were collected through this follow-up.

As there is much interest in the evaluation of new technologies and in particular the use of technologies for cervical cancer screening (Arbyn, Ronco et al. 2009), it is important to consider the standards that are proposed for reporting and interpretation of studies evaluating these technologies.

In 2003, a report addressing standard for reporting of diagnostic accuracy (STARD) encouraged investigators to more thoroughly report the details of studies to improve interpretation for researchers and for clinicians and the public (Bossuyt, Reitsma et al. 2003). While numerous threats to internal and external validity of diagnostic tests exist, the primary goal of improving reporting is to help readers detect and bias in a study.

Important areas for reporting are set out in the STARD report (Bossuyt, Reitsma et al. 2003) and include: details on the participants, including details of recruitment; the index test and the reference standard. With regard to the test methods details of the test, criteria for positivity (including cut points), and the blinding (or not) of readers of the index test and the reference test to the results of the other test is necessary. For participants report the beginning and end dates, the clinical and demographic characteristics, and the time frame from index test to the reference standard. Cross tabulations of the index test against the results of the reference standard, are needed along with details of any missing data. Report estimates of diagnostic accuracy. Also report how indeterminate results, missing data, and outliers of the index test are handled. Finally if possible report an estimate of test reproducibility.

Within the context of the reported trial, Tong et al compare two tests but do not employ a gold standard as the comparator; the details of the population are somewhat lacking, and in particular the screening history (if any) of the women prior to implementing the program clouds consideration of the test as a true screening. Women were excluded if they self reported a history of cervical cancer.

While some of the appropriate data for reporting of diagnostic test evaluation were added through a response to correspondence (Tong, Shen et al. 2009; Garner, Guillaud et al. 2010), the underlying data from the reported trial fail to provide full details expected by the STARD criteria. If we compare this report against the STARD criteria in the context of screening we note that a diagnostic test evaluation would require that all subjects receive the gold standard or reference test. This may not be suitable in the context of screening programs when the majority of the population is free from disease. This is the classic problem of how we identify true negatives from false negatives when screening for cancer. Importantly, the goal of screening is the reduction of cancer mortality. Reducing advanced stage disease may also be used as a primary endpoint in some settings (typically when treatment of advanced disease has little clinical benefit). Introduction of a new test that increases the detection of early stage lesions makes incidence of premalignant and early stage lesions not particularly relevant as endpoints. For cervical cancer this is relevant as new technology may detect lesion that will not progress, hence not particularly important for the primary goal of reducing mortality. Furthermore, using colposcopy as an intermediate marker to detection of cervical lesion is problematic as the sensitivity of colposcopy varies with the number of biopsies. It is thus an imperfect referent standard and without details of number of biopsies per colposcopy does not limit the potential for bias in the work up of positive screen results.

A final concern in evaluation of screening technology is the use of cross-sectional performance vs. longitudinal performance and outcomes. The Tong et al trial appears to report incidence of cancer from the screening test only. This then is a cross-sectional evaluation of sensitivity and specificity. Improving sensitivity in a cross-sectional study does not imply that adding the test to a screening program will necessarily lead to reduction in incidence of aggressive or lethal cancers. Accordingly, Arbyn and colleagues recommend that evaluation of screening technologies focus on the reduction of cancer in subsequent screening rounds rather than settle for cross-sectional test performance (Arbyn, Ronco et al. 2009).

With improved reporting standards for diagnostic tests (Arbyn, Ronco et al. 2009), application of these criteria to improve reporting in obstetrics and gynecology (Selman, Morris et al. 2011) and improved design and evaluation of new technologies for screening, we may better inform practices to improve clinical outcomes and population health.

Literature Cited

- Arbyn M, Ronco G, et al. How to evaluate emerging technologies in cervical cancer screening? International journal of cancer. Journal international du cancer. 2009; 125(11):2489–2496. [PubMed: 19626591]
- Bossuyt PM, Reitsma JB, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ. 2003; 326(7379):41–44. [PubMed: 12511463]
- Garner DM, Guillaud MD, et al. DNA ploidy cytometry testing for cervical cancer screening in China - letter. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010; 16(13):3517. author reply 3517–3519. [PubMed: 20570931]
- Selman TJ, Morris RK, et al. The quality of reporting of primary test accuracy studies in obstetrics and gynaecology: application of the STARD criteria. BMC women's health. 2011; 11:8. [PubMed: 21429185]

Tong H, Shen R, et al. DNA ploidy cytometry testing for cervical cancer screening in China (DNACIC Trial): a prospective randomized, controlled trial. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2009; 15(20):6438–6445. [PubMed: 19825960]