Published in final edited form as:

Med Decis Making. 2012 ; 32(3): 400–427. doi:10.1177/0272989X11428516.

POLICY IMPLICATIONS OF ADJUSTING RANDOMIZED TRIAL DATA FOR ECONOMIC EVALUATIONS: A DEMONSTRATION FROM THE ASCUS-LSIL TRIAGE STUDY

Nicole G. Campos, Ph.D.^{1,2}, Philip E. Castle, Ph.D., M.P.H.³, Mark Schiffman, M.D., M.P.H.⁴, and Jane J. Kim, Ph.D.¹

¹Center for Health Decision Science, Harvard School of Public Health, Boston, MA

²Center of Excellence for Health Disparities Research: El Centro, School of Nursing and Health Studies, University of Miami, Coral Gables, FL, USA

³American Society for Clinical Pathology Institute, Washington, DC

⁴Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, DHHS, Rockville, MD

Abstract

Background—Although the randomized controlled trial (RCT) is widely considered the most reliable method for evaluation of health care interventions, challenges to both internal and external validity exist. Thus, the efficacy of an intervention in a trial setting does not necessarily represent the real-world performance that decision makers seek to inform comparative effectiveness studies and economic evaluations.

Methods—Using data from the ASCUS-LSIL Triage Study (ALTS), we performed a simplified economic evaluation of age-based management strategies to detect cervical intraepithelial neoplasia grade 3 (CIN3) among women who were referred to the study with low-grade squamous intraepithelial lesions (LSIL). We used data from the trial itself to adjust for 1) potential lead time bias and random error that led to variation in the observed prevalence of CIN3 by study arm, and 2) potential ascertainment bias among providers in the most aggressive management arm.

Results—We found that using unadjusted RCT data may result in counterintuitive cost-effectiveness results when random error and/or bias are present. Following adjustment, the rank order of management strategies changed for two of the three age groups we considered.

Conclusion—Decision analysts need to examine study design, available trial data and cost-effectiveness results closely in order to detect evidence of potential bias. Adjustment for random error and bias in RCTs may yield different policy conclusions relative to unadjusted trial data.

Keywords

economic evaluation; randomized trial; bias; cervical cancer; human papillomavirus

Reprints or correspondence: Nicole G. Campos, P.O. Box 248153, Coral Gables, FL 33124-3850, phone: 305-284-2895
gastin@post.harvard.edu.

Potential conflicts of interest. N.G.C.: No conflict. P.E.C. has received donations of HPV tests from Qiagen and serves on a data safety and monitoring committee for Merck, the manufacturer of Gardasil, for which he receives compensation. M.S.: No conflict. J.J.K.: No conflict.

INTRODUCTION

Although the randomized controlled trial (RCT) is considered the most reliable method for evaluation of health care interventions,^{1,2} the use of RCT data for estimating the health impact of interventions for economic evaluations may be problematic.^{3,4,5} Beyond threats to internal validity—including random error and bias—there are challenges to external validity as well. Due to strict exclusion criteria invoked in subject selection,³ and high intensity of testing and care under protocol,⁴ intervention efficacy in a trial setting does not necessarily represent the real-world performance that decision-makers seek to inform analyses of comparative effectiveness.

Eddy notes that, in the process of evaluating medical technologies, a common heuristic in dealing with threats to validity is to either 1) designate the bias as “acceptable” and take a study’s findings at “face value”, or 2) determine the bias is “unacceptable” and disregard the study.⁶ As an alternative to these extreme approaches, Eddy argues for a middle ground that involves estimating the magnitude of biases and adjusting for them in a formal way, relying on the adjusted results for decision-making purposes. Using this general framework, Rittenhouse has adjusted RCT data to demonstrate the impact of strict exclusion criteria on cost-effectiveness analysis.³ Other economic evaluations have adjusted RCT data to correct for the artificial conditions under which intervention effectiveness is measured.^{5,7}

We use data from the atypical squamous cells of undetermined significance (ASCUS) low-grade squamous intraepithelial lesion (LSIL) Triage Study (ALTS), a randomized multicenter trial sponsored by the National Cancer Institute (NCI) from 1997 to 2001, to demonstrate the pitfalls of using unadjusted RCT data in a decision analytic model. In this paper, we apply Eddy’s general framework to inform an economic evaluation of age-based management protocols for women with LSIL cervical cytology results. We present a simple method of correcting for random error and bias in study design using data from the trial itself, and show that such adjustments lead to different policy conclusions relative to unadjusted trial data.

The ASCUS-LSIL Triage Study (ALTS) (Appendix Table 1)

Approximately 15 carcinogenic types of human papillomavirus (HPV) cause virtually all cases of cervical cancer, but the vast majority of infections are benign and clear within 1–2 years of sexual acquisition.⁸ Widespread screening with cervical cytology can reduce the incidence of cervical cancer through detection and treatment of precancerous lesions.⁹ More than 50 million Pap smears are performed in the United States each year, and more than 5% are reported as abnormal.¹⁰ While health care providers generally agree that cytologically diagnosed high-grade squamous intraepithelial lesions (HSILs) should be further evaluated with colposcopy and biopsy, optimal management of approximately 3 million women with cytologic results of LSIL and ASC-US* are less straightforward.¹¹ An LSIL classification indicates an acute HPV infection.¹² Some LSIL heralds the presence of high-grade precancer, histologic cervical intraepithelial neoplasia grade 3 (CIN3) (Figure 1). ASC-US is the most common abnormality in the United States, forming the boundary between normal and abnormal cytologic interpretations. A sizable fraction of CIN3 and higher are detected by ASC-US cytology, due to the commonality of the result and the fact that some represent true abnormalities.¹² Colposcopically directed biopsy to assess disease severity is costly and likely leads to overtreatment of lesions that may spontaneously regress. Yet many clinicians are hesitant to pursue less aggressive follow-up measures for fear of missing CIN3.^{13,14}

*The name of the study, ALTS, includes reference to “ASCUS” as defined by the 1988 Bethesda System. However, in this manuscript we follow the 2001 Bethesda System and refer to the cytology result as “ASC-US”.¹¹

ALTS compared three methods of managing women with ASC-US or LSIL cytologic interpretations¹⁵ for the timely detection of CIN3. Although the treatment threshold in the trial and in current practice is CIN grade 2 and higher [CIN2+], CIN2 is not a reliable indication of precancer—it may be caused by non-carcinogenic HPV types and frequently regresses.⁸ CIN3 detection was chosen as the primary outcome of interest because histologically confirmed CIN3 is at high risk of progressing to invasive cancer and requires definitive treatment.¹⁶ The three arms of the trial included 1) immediate colposcopy (IC) and directed biopsy (all women were sent to colposcopy at enrollment); 2) conservative management (CM) (women received colposcopy only if enrollment or any follow-up cytology at 6 month intervals was HSIL or above [HSIL+]); and 3) HPV DNA testing with the Hybrid Capture 2 test (HC2) (Qiagen, Gaithersburg, MD) (colposcopy only if the enrollment HC2 test was positive for carcinogenic HPV or missing, or enrollment cytology was HSIL+). IC was considered to have optimal sensitivity (and thus safety against cancer risk) but might lead to costly overtreatment and its sequelae,¹⁴ whereas CM presumably allowed for some lesions to resolve spontaneously, but may have posed a greater cancer risk. Detection of carcinogenic HPV was considered as a means of triaging women to colposcopy based on high (carcinogenic HPV+) risk of precancer and cancer.¹⁷ HPV DNA testing could be a useful triage tool if carcinogenic HPV infections are not detected in the vast majority of women with ASC-US or LSIL, and if the detection of carcinogenic HPV is sensitive for CIN3.

Cost data were not collected as part of the study. Two published studies—one from a long-term societal perspective¹⁸ and one from the short-term perspective of ALTS¹⁹—have estimated the cost-effectiveness of management strategies for women in the United States with an ASC-US referral. Both found HPV triage to colposcopy an attractive option for women with ASC-US cytology results, regardless of age. To our knowledge, there are no published studies on the cost-effectiveness of LSIL management in the United States. As of 2009, clinical guidelines in the U.S. recommend that cervical cancer screening should begin at age 21,⁹ with referral to colposcopy and biopsy for those with LSIL results.²⁰ Given the decreasing prevalence of carcinogenic types of HPV as women age, even among those with LSIL cytology,^{21,22,23} the present analysis examines the cost-effectiveness of management strategies for LSIL women in ALTS according to age strata, to discern whether HPV DNA testing may be an attractive option for triaging older women to colposcopy. We demonstrate how cost-effectiveness results may change following adjustment for random error and bias in trial design.

METHODS

All women in ALTS received pelvic examinations, Pap smears, and HPV DNA testing at enrollment and every 6 months for two years. After the enrollment pelvic examination and specimen collection, women randomized to the IC arm received same-day colposcopy or, if unable to stay, were given an appointment to return for colposcopy (ideally within three weeks).¹⁵ Women randomized to the HPV and CM arms were not triaged to colposcopy until enrollment test results were available several weeks after the enrollment visit. Regardless of study arm, women with a cytology result of HSIL during follow-up were referred to colposcopy, or received safety notification. Women diagnosed as CIN2+ by the clinical center or CIN3+ by the quality control (QC) pathology group were treated by excision. At exit, colposcopy was performed on all women to capture any missed disease. The primary study end point was a worst QC pathology diagnosis of CIN3 throughout the two-year study.[†] The ALTS protocol was reviewed and approved by institutional review boards at the NCI and each participating clinical center. All study participants provided written informed consent.

The HPV arm for women referred with an LSIL cytology interpretation closed early after an interim analysis revealed that 83% were triaged to colposcopy based on a positive HPV DNA test result.²⁴ Investigators estimated the high prevalence of carcinogenic HPV infections would limit the usefulness of HPV DNA testing as a triage mechanism for women with LSIL, with the expected cost of HPV testing likely outweighing potential savings from avoiding colposcopy in 20–27% of women.²⁴ An interim safety analysis revealed that CM was not sufficiently sensitive for detection of CIN3 among women with an LSIL referral. Consequently, women in the LSIL CM arm who had not previously been triaged to colposcopy were referred to this procedure at their next visit as the result of a protocol modification in July 1999.²⁵

Study sample

We restricted the present analysis to the 1,572 women in ALTS with a referral interpretation of LSIL. We excluded those with missing QC pathology diagnoses (n=6) as uninformative and women with cancer detected at enrollment (n=4) as a pre-existing condition that was unrelated to the management strategy. We included one case of cancer in the CM arm that was not detected until 18 months into the study and was therefore considered a failure of the arm. Our sample included 669 women, 222 women, and 671 women in the IC, HPV, and CM arms, respectively. We stratified women into the following age tertiles: <21 years, 21–24 years, and ≥25 years.

Decision Model

Analytic overview—We modified a decision analytic model previously used to evaluate the short-term costs and effectiveness of management strategies for women with ASC-US.¹⁹ An overview of the model is shown in Figure 2.

We conducted this analysis from a payer perspective. Our measure of effectiveness for each management strategy was the number of clinical center diagnoses of CIN2+ (at enrollment for the IC and HPV arms; at enrollment, 6 month, and 12 month visits for the CM arm) among women with a worst 2-year QC pathology group diagnosis of CIN3 because the clinical center diagnosis of CIN2+ led to immediate treatment of women with the primary endpoint, as it does in real-world practice. Due to the short-term horizon of the trial, we did not discount costs or effectiveness measures, as would be recommended in a cost-effectiveness analysis from a societal perspective.²⁶ The relative performance of each management strategy was expressed as an incremental cost-effectiveness ratio (ICER), defined as the incremental cost divided by the incremental effectiveness of a strategy compared to the next most costly strategy, after eliminating strategies that were dominated (i.e., more costly and less effective than other strategies, or having a higher incremental cost-effectiveness ratio than more effective strategies). The ICER was expressed as the cost per case of CIN3 detected at the clinical centers.

Effectiveness Data—Primary data from ALTS were used to calculate the probability of events in the model. The proportions of women in each model health state—normal, CIN1, CIN2, and CIN3—were based on worst 2-year QC pathology group diagnosis. The probabilities of a positive triage test result and of a clinical center diagnosis of CIN2+ on biopsy were conditioned on health state. Women who were referred to but did not participate

[†]Due to the imperfect sensitivity of colposcopy, it is uncertain whether CIN3 detected late in the study may have been missed prevalent or newly incident cases. However, a review of all records for CIN3 cases suggested that many represented missed prevalent disease, rather than incident CIN3 (The ALTS Group 2003). Given the lack of a gold standard, the detection of CIN3 at any time during the two years of the trial (i.e., worst 2-year QC diagnosis of CIN3) is the best proxy available for the presence of precancer at enrollment.

in colposcopy at the specified visit did not incur costs of colposcopy and were not eligible for case detection. Women who missed one or more visits in the CM arm did not incur costs of Pap smear or colposcopy and were not eligible for case detection at the missed visit, but could incur such costs or benefit from case detection if they received a Pap smear and referral to colposcopy at a subsequent visit.

Table 1 displays the distribution of trial parameters that determine the effectiveness and cost-effectiveness of management strategies, by study arm and age tertile. We calculated exact binomial 95% confidence intervals for proportions. Fisher exact tests were used to assess associations between age tertiles (within study arm) and study arms (within age tertile). When $p < 0.05$, a Tukey-type multiple comparison test for proportions was used to discern which groups were significantly different from each other.²⁷

Although prevalence of CIN3 increased slightly with age but did not vary significantly between study arms or age tertiles, it was highest in the HPV arm and lowest in the CM arm, with the difference perhaps due to the lead-time detection of small prevalent lesions in the HPV arm that would have otherwise regressed.[‡] We note the relatively higher CIN3 prevalence among women in the oldest tertile of the HPV arm (22.1%), compared with the other arms and age tertiles (~13–16%), which we attribute in part to random error as a result of small sample size of the HPV arm (which closed early). As expected, colposcopic referrals were significantly different between all study arms across all age tertiles ($p < 0.0001$), with the lowest percentage of referrals in the CM arm. While performance of management strategy for detection of CIN3 did not vary significantly between study arms or age tertiles, sensitivity of colposcopy was higher in the CM and HPV arms than in the IC arm, especially in women the youngest tertile of age ($p = 0.008$). We hypothesize that the lower sensitivity in the IC arm may in part be due to colposcopists' perception of lower risk among women who were automatically referred at enrollment, as opposed to those referred several weeks later for abnormal triage test results (as in the HPV and CM arms). There was a significant association between study arm and positive predictive value of colposcopy in the youngest tertile ($p = 0.03$).

Costs—Costs are displayed in Table 2. We considered only direct medical costs that would be accrued to detect CIN3 using each management strategy in a clinical setting, and ignored costs that would have accrued due to protocol-driven follow-up. We used Medicare reimbursement rates to estimate costs.^{28,29}

Model Assumptions—We made the following assumptions: 1) only clinical center colposcopy results from the enrollment visit were considered in the IC and HPV arms, and from the enrollment, 6 month, and 12 month visits were considered in the CM arm; 2) women were censored from the model and could not accrue further costs or case detection after their initial clinical center colposcopy; 3) HPV DNA testing in the HPV arm was assumed to be reflex testing of liquid-based cytologic samples from the referral visit, and thus did not require an additional office visit; 4) women with missing HPV DNA test results were sent to colposcopy, as in the study; 5) testing positive in the CM arm was calculated separately for the enrollment, 6 month, and 12 month visits, conditional on participation and testing <HSIL at previous visits; and 6) women in the CM arm who underwent the protocol modification were censored from the analysis thereafter.

[‡]There appears to have been statistically significant regression of missed prevalent CIN2 in the CM arm relative to the IC arm over the two years of the trial (comparisons to the HPV arm were not possible due to small numbers) (The ALTS Study Group 2003). Despite modest gains in statistical power to detect differences in disease prevalence by including CIN2 as an endpoint of interest (Appendix Table 2), the tendency of CIN2 to regress makes it an unreliable indication of precancer.

Analysis—First we conducted a cost-effectiveness analysis of age-based management of women with LSIL cytology results using unadjusted ALTS data (i.e., all probabilities in the model were derived from ALTS data on corresponding study arm and age tertile) (*Step 1*). In a second step, we adjusted for potential lead time bias in CIN3 detection and random error by using the same prevalence distribution of model health states for each study arm (*Step 2a*) and independently adjusted for ascertainment bias by informing colposcopic performance in a subgroup of the IC arm with data from a clinically comparable subgroup of the HPV arm (*Step 2b*). Finally, we adjusted for both lead time bias/random error and ascertainment bias simultaneously (*Step 3*).

To arrive at a prevalence distribution of model health states for each age tertile (*Step 2a*), we calculated a weighted average of the prevalence of each state (i.e., normal, CIN1, CIN2, CIN3) across all three arms. This distribution should allow us to reflect age trends in the prevalence of health states, but will not favor one management strategy over another in cost-effectiveness analysis as the result of there being more cases of CIN3 detected in a particular arm over the course of the study.

To adjust for ascertainment bias among colposcopists (*Step 2b*), we estimated the extent to which greater colposcopic sensitivity in the HPV and CM arms was due to the clinical profile of women referred to colposcopy versus different management resulting from provider knowledge of study arm. Published literature suggests that colposcopy is more sensitive to detect CIN2+ among women with an HSIL+ cytology result than in women with <HSIL on cytology, due to the correlation between HSIL+ on cytology and size of a lesion.^{30,31} Data from ALTS suggest that HPV16 may cause more visual abnormalities than other HPV types, even after controlling for histologic diagnosis.³² We thus restricted CIN3 cases in each arm to those that were HC2+ (at enrollment in the IC and HPV arms, and at the visit of referral to colposcopy in the CM arm),[§] and stratified these cases based on cytology result at enrollment (<HSIL vs. HSIL+). (In the CM arm, clinical center colposcopy results were only available for women with HSIL+) (Table 3). While sensitivity of colposcopy to detect CIN3 was uniformly high across arms among women with HC2+ and HSIL+ at enrollment (all ages: Fisher exact $p=0.7$), sensitivity was lower in the IC arm relative to the HPV arm among women with HC2+ and <HSIL (all ages: Fisher exact $p=0.02$). We explored how management of these women (CIN3, HC2+, <HSIL) may have differed between the IC and HPV arms, but were unable to determine a mechanism for differences in colposcopic performance based on examination of provider type, number of biopsies taken,³³ and severity of clinical center colposcopic impressions (Appendix Tables 3–5).

Among women with similar clinical profiles and seemingly smaller lesions (CIN3, HC2+, <HSIL), we adjusted colposcopic performance in our model as follows (*Step 2b*). In the IC arm, we stratified women with CIN3 in the model based on whether they were HC2– or HC2+. Among the few women who were HC2– (youngest tertile: $n=0$; middle tertile: $n=3$; oldest tertile: $n=3$), we made no adjustment, as we had no better data to inform sensitivity of colposcopy in this subgroup. Among HC2+ women, we further stratified on whether they were HSIL+ and <HSIL at enrollment. For those with HSIL+ on enrollment cytology, we made no adjustment, given the similarity in colposcopic performance between arms in this subgroup. For women with <HSIL on enrollment cytology, we used colposcopic performance data from women with a comparable profile (HC2+, <HSIL) in the HPV arm.

[§]To be consistent with the protocol for the HPV arm, we assumed the $n=3$ women with CIN3 in the IC arm with missing enrollment HC2 results were defaulted positive and included their colposcopy results in our analysis of performance. In the CM arm, all women with CIN3 who were HSIL+ and sent to colposcopy were HC2+.

RESULTS

Step 1: Unadjusted ALTS data

Results from the unadjusted analysis are presented in Table 4. One repeat cytology visit (CM 1st visit) was the least costly and least effective strategy across age tertiles. HPV triage to colposcopy was the most effective strategy across tertiles and, contrary to expectations, consistently dominated other strategies. We explored the impact of two potential sources of bias that led to these counterintuitive cost-effectiveness results— variation in prevalence of CIN3 by arm and potential ascertainment bias at colposcopy— separately (*Steps 2a* and *2b*) and together (*Step 3*).

Step 2a: Adjustment for variation in prevalence of CIN3

When we used the same prevalence distribution of health states across study arms, ICERs for HPV triage to colposcopy became relatively less attractive than in *Step 1* (Table 4). Effectiveness of HPV triage decreased because prevalence of CIN3, which had been slightly higher in the HPV triage arm than the IC arm in all age tertiles (Table 1), declined. With fewer cases of CIN3 in the HPV arm, the cost per case detected increased. While HPV triage remained the most attractive strategy (i.e., most effective and had the lowest ICER) in women under 25, and had the lowest ICER among women 25 and over, IC was the most effective strategy among women 25 and over, yielding an ICER of \$3,260 per CIN3 detected compared to HPV testing.

Step 2b: Adjustment for potential ascertainment bias among colposcopists in the IC arm

When we adjusted performance of colposcopy among women in the IC arm with CIN3 who were HC2+ with <HSIL on enrollment cytology, IC became the most effective strategy in the youngest tertile, as the sensitivity of colposcopy improved. However, IC remained more costly and less effective than HPV triage in women over 21 (Table 4). HPV triage remained the strategy with the lowest ICER in all age tertiles, as with unadjusted trial data.

Step 3: Simultaneous adjustment for variation in prevalence of CIN3 and ascertainment bias

Compared to the unadjusted data, standardizing CIN3 prevalence and adjusting colposcopic performance in the IC arm resulted in IC becoming the most attractive strategy (i.e., most effective and lowest ICER) in women under 21 and 25 and over. HPV triage remained the most attractive strategy in women aged 21 to 24 years because of its superior effectiveness and low cost in this age group.

DISCUSSION

Our analysis demonstrates that economic evaluation results based on RCT data may change following adjustment for random error and biases in study design. We used ALTS data to evaluate the cost-effectiveness of strategies for CIN3 detection in women referred to the study with LSIL cytology results, to determine whether there might be a role for HPV triage to colposcopy at older ages when HPV prevalence declines. While the use of unadjusted ALTS data leads to the finding that HPV triage is the most effective strategy and has the lowest ICER in all age groups, adjustments for variation in the prevalence of CIN3 between arms and possible ascertainment bias by clinical center colposcopists suggest that there may not be a role for HPV triage among women with LSIL. Despite our efforts at adjustment, the small sample size in the HPV arm, due to early closure, limits our power to detect age trends and significant predictors of colposcopic sensitivity. It also limits our power to discern true differences in colposcopic performance between arms from differences due to statistical noise.

As an economic evaluation, our analysis is limited in scope and is not intended to inform policy. It was conducted from a payer perspective with a short time horizon that ends at colposcopy. We did not include costs of treatment, and did not capture the long-term societal economic, health, or quality of life impacts associated with positive colposcopy/biopsy interpretations (such as adverse events and long-term sequelae associated with treatment) or of missed precancers. Extending the time horizon is beyond the scope of the present analysis. We demonstrate the range of policy conclusions that can be drawn in economic analysis when RCT data are taken at “face value” versus when we adjust for potential biases in study design.

Our adjustment for variation in prevalence of CIN3 between arms was simple (*Step 2a*), and involved calculating a weighted average across arms. Theoretically, given a large sample and no difference in lead time between arms, prevalence of CIN3 should have been uniform across study arms. In women under 25 years, the adjustment had minimal impact, changing prevalence of CIN3 by only 1–2%. ICERs associated with HPV triage increased slightly, as the cost per case of CIN3 detected was spread over fewer cases. In women over 25, however, CIN3 prevalence fell by nearly 6% as the result of adjustment, making IC the most effective strategy.

Inferior performance of colposcopy among women in the IC arm relative to the HPV arm, even after controlling for HPV status and enrollment cytology results, was presumably due to ascertainment bias. We hypothesize that the immediacy of colposcopy in relation to the enrollment visit alerted colposcopists that women were assigned to the IC arm. Colposcopists may have perceived these women as having lower risk of precancer than those in the HPV and CM arms, who were referred to colposcopy based on abnormal test results. The apparent bias occurred primarily in women with small lesions (i.e., <HSIL on enrollment cytology, younger women), and may have led to less effective targeting of biopsies by providers. Following adjustment of colposcopic performance in the IC arm (*Step 2b*), IC was the most effective strategy in the youngest tertile, although HPV triage was more effective and had a lower cost-effectiveness ratio in women over 21 years.

Simultaneous adjustment for variation in CIN3 prevalence and ascertainment bias (*Step 3*) suggested that IC was the most effective strategy to detect CIN3, and had the lowest cost per case of CIN3 detected, among women with LSIL under 21 or 25 and older.

Although our adjustments to CIN3 prevalence were within the confidence intervals for each study arm, and might have been assessed in sensitivity analysis, we emphasize the importance of incorporating adjustments into baseline assumptions when the analyst is confident that random error or bias is present. As Rittenhouse notes, “[t]he baseline case takes on an aura of having a greater likelihood of being correct if for no other reason than it represents the assumptions that dominate any report of results.”³ Given the weight of baseline assumptions in decision making, these (rather than sensitivity analyses) should reflect our best sense of clinical realities.³

There are limitations associated with our adjustment of colposcopic sensitivity in the IC arm. We did not adjust for sensitivity among women with HC2+ and HSIL+, because there was little indication of ascertainment bias in this subgroup, as the IC arm performed comparably, if not better than, the CM arm in all age tertiles. In 21 to 24 year olds with HC2+ and HSIL+, colposcopy in the IC arm detected 70% of CIN3 cases that were referred, while 100% of referred CIN3 cases were detected in the HPV arm (Table 3). While this difference was not statistically significant, the magnitude was large enough to make HPV triage the most effective strategy in this age group. We also did not adjust performance of colposcopy among women with CIN3 in the IC arm who were HC2–, as there was no

comparable subgroup in the HPV and CM arms. Excluding these few women (n=3 in 21–24 year olds; n=3 in 25 years and older) did not change the rank order of strategies.

Our method of adjustment assumes that colposcopy in the IC arm under-performed as a result of providers' knowledge of group. We cannot rule out a scenario in which the HPV and CM arms over-performed as a result of review bias arising from providers' knowledge of abnormal test results. While it is unlikely that colposcopists were aware of particular HC2 or cytology results from enrollment, the lag time between enrollment and the colposcopy visit were an indication of abnormal triage results. If the discrepancy in colposcopic sensitivity between arms is due to review bias rather than ascertainment bias, it would be more appropriate to inform sensitivity in the HPV arm with data from the IC arm, instead of vice versa. An exploratory analysis in which we adjusted for both variation in CIN3 prevalence and review bias (by informing colposcopic sensitivity among <HSIL women in the HPV arm with data from <HSIL, HC2+ women in the IC arm) yielded a similar rank ordering of strategies as the model in *Step 3* of the present analysis (Appendix Table 6). Although ICERs were relatively higher due to the lower effectiveness of IC under this method of adjustment, IC was the most effective strategy and had the lowest ICER in the youngest and oldest tertiles, while HPV triage had the lowest ICER in the middle tertile. It is possible that both review and ascertainment bias were at work in ALTS, but we cannot discern the relative magnitude of either bias. We believe that under-performance in the IC arm was more influential than over-performance in the HPV arm, based on the surprisingly low sensitivity of IC in the trial.

In real-world clinical practice, availability (or lack) of triage test results may affect how colposcopy is done. The fact that ALTS data captures this tendency is not necessarily problematic from the standpoint of economic evaluation if sensitivity is influenced by triage test availability in clinical practice. However, if colposcopists in the trial behaved differently not only because they were aware of study arm or triage results, but because they were self-conscious about protocol-driven review of their evaluations, this is a type of Hawthorne effect and is problematic from the standpoint of economic evaluation.³⁴ It is not possible to distinguish real-world provider behavior from a Hawthorne effect using trial data.

The threshold of HC2 positivity at which HPV triage might be cost-effective in a population of women with LSIL cytology results is of great interest to policy makers. Unfortunately, due to the importance of colposcopic sensitivity in assessing cost-effectiveness, and the remaining uncertainty in this parameter in the IC and HPV arms despite our best efforts at adjustment, we are unable to address this issue with precision. We emphasize the need for data on colposcopic performance— how performance varies by provider type, the influence of procedural factors, and how colposcopists use information from triage tests in routine clinical practice— to inform comparative effectiveness studies and economic evaluations.

Rather than approaching ALTS data with the view that either 1) its limitations are acceptable and data should be taken at “face value” or 2) its limitations and biases are unacceptable and the data should be ignored for purposes of economic analysis, we sought a middle ground by adjusting for variation in prevalence of CIN3 and possible ascertainment bias using data from the trial itself. While there are more statistically rigorous ways to adjust for biases using Bayesian methods (for example, the Confidence Profile Method described by Eddy),³⁵ our goal was to present a method of adjustment that was straightforward and grounded in understanding of trial design, clinical presentation of disease, and test performance. Eddy himself laments that the complexity of his method hampers its use among many researchers who assess health technologies.³⁶

Adjustments for random error and potential bias may not always impact policy conclusions, and the present analysis does not invalidate previously published findings from ALTS. A re-analysis of women referred to ALTS with ASC-US revealed that even after adjusting for possible variation in CIN3 prevalence and ascertainment bias (Appendix Tables 7–9), HPV triage remained the strategy with the lowest ICER in all age quartiles, and the most effective strategy in women over age 25 (Appendix Table 10). This is in line with other cost-effectiveness studies (including one based on ALTS data) that have found HPV triage to be an attractive strategy for women with ASC-US.^{18,19} When evidence of bias is not pronounced, or a strategy is robustly attractive even as influential parameters are varied, adjusting for potential bias may not alter policy conclusions. The decision analyst must examine RCT design when the results of economic evaluation are sensitive to reasonable variation in trial parameters. Even when the unavailability of individual-level data prevents thorough examination of mechanisms for bias, the analyst can attempt crude adjustments (with varying levels of conservatism) based on published figures from the RCT (Appendix). If the optimal strategies shift when using reasonably adjusted vs. unadjusted data, the analyst will be aware that the foundation of the economic evaluation is tenuous.

We present our methods of adjustment to demonstrate the potential pitfalls of taking RCT data at face value, without consideration of biases in study design. Random error and biases may be present even in well-designed RCTs, and our analysis reveals that these may not always be obvious or readily confirmed, even when individual-level trial data are available. We assert that decision analysts need to examine study design, available trial data and cost-effectiveness results closely in order to detect evidence of potential bias. Even if bias cannot be confirmed, efforts should be made to adjust the data and assess the impact of potential bias on policy conclusions. Failure to adjust RCT data may yield counterintuitive results in economic evaluation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Cancer Institute (R01 CA93435), the National Institutes of Health/ National Cancer Institute Intramural Research Program, and National Cancer Institute, National Institutes of Health Department of Health and Human Services contracts No. CN-55153, CN-55154, CN-55155, CN-55156, CN-55157, CN-55158, CN-55159, and CN-55105.

References

1. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Health Care*. 2003; 12:47–52. [PubMed: 12571345]
2. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000; 342:1887–1892. [PubMed: 10861325]
3. Rittenhouse BE. The relevance of searching for effects under a clinical-trial lamppost: a key issue. *Med Decis Making*. 1995; 14:348–357. [PubMed: 8544678]
4. Rittenhouse BE. Exorcising protocol-induced spirits: making the clinical trial relevant for economics. *Med Decis Making*. 1997; 17:331–339. [PubMed: 9219194]
5. Drummond M. Experimental versus observational data in the economic evaluation of pharmaceuticals. *Med Decis Making*. 1998; 18:512–518.
6. Eddy, DM. Should we change the rules for evaluating medical technologies?. In: Gelijns, AC., editor. *Modern Methods of Clinical Investigation*. Washington, DC: National Academy Press; 1990.

7. O'Brien BJ, Anderson DR, Goeree R. Cost-effectiveness of enoxaparin vs. warfarin prophylaxis against deep-vein thrombosis after total hip replacement. *Can Med Assoc J.* 1994; 150:1083–1090. [PubMed: 8137188]
8. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet.* 2007; 370:890–907. [PubMed: 17826171]
9. American College of Obstetricians and Gynecologists. ACOG Practice Bulletin No 109: Cervical cytology screening. *Obstet Gynecol.* 2009; 114:1409–1420. [PubMed: 20134296]
10. National Cancer Institute Fact Sheet. [Last accessed on August 27, 2010] Pap test. Available at: <http://www.cancer.gov/cancertopics/factsheet/Detection/Pap-test>
11. Solomon D, Davey D, Kurman R, Moriarty A, O'Connor D, Prey M, et al. The 2001 Bethesda System: Terminology for reporting results of cervical cytology. *JAMA.* 2002; 287:2114–2119. [PubMed: 11966386]
12. Schiffman M, Wentzensen N, Wacholder S, Kinney W, Gage JC, Castle PE. Human papillomavirus testing in the prevention of cervical cancer. *J Natl Cancer Inst.* 2011; 103:368–383. [PubMed: 21282563]
13. Schiffman M, Kjaer SK. Natural history of anogenital human papillomavirus infection and neoplasia. *J Natl Cancer Inst Monogr.* 2003; 31:14–19. [PubMed: 12807940]
14. Schiffman M, Solomon D. Findings to date from the ASCUS-LSIL Triage Study (ALTS). *Arch Pathol Lab Med.* 2003; 127:946–949. [PubMed: 12873166]
15. Schiffman M, Adhisa ME. ASCUS-LSIL Triage Study – design, methods and characteristics of trial participants. *Acta Cytol.* 2000; 44(5):726–742. [PubMed: 11015972]
16. Solomon D, Schiffman M, Tarone R. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. *J Natl Cancer Inst.* 2001; 93:293–299. [PubMed: 11181776]
17. The ASCUS-LSIL Triage Study (ALTS) Group. Results of a randomized trial on the management of cytology interpretations of atypical squamous cells of undetermined significance. *Am J Obstet Gynecol.* 2003; 188:1383–1392. [PubMed: 12824967]
18. Kim JJ, Wright TC, Goldie SJ. Cost-effectiveness of alternative triage strategies for atypical squamous cells of undetermined significance. *JAMA.* 2002; 287:2382–2390. [PubMed: 11988059]
19. Kulasingam SL, Kim JJ, Lawrence WF, Mandelblatt JS, Myers ER, Schiffman M, et al. Cost-effectiveness analysis based on the Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesion Triage Study (ALTS). *J Natl Cancer Inst.* 2006; 98:92–100. [PubMed: 16418511]
20. American College of Obstetricians and Gynecologists. ACOG Practice Bulletin No. 99: Management of abnormal cervical cytology and histology. *Obstet Gynecol.* 2008; 112:1419–1444. [PubMed: 19037054]
21. Sherman ME, Schiffman M, Cox JT. Effects of age and human papilloma viral load on colposcopy triage: data from the randomized Atypical Squamous Cells of Undetermined Significance/Low-grade Squamous Intraepithelial Lesion Triage Study (ALTS). *J Natl Cancer Inst.* 2002; 94:102–107. [PubMed: 11792748]
22. Arbyn M, Martin-Hirsch P, Buntinx F, Van Ranst M, Paraskevaidis E, Dillner J. Triage of women with equivocal or low-grade cervical cytology results: a meta-analysis of the HPV test positivity rate. *J Cell Molec Med.* 2009; 13:648–659. [PubMed: 19166485]
23. Brismar-Wendel S, Froberg M, Hjerpe A, Andersson S, Johansson B. Age-specific prevalence of HPV genotypes in cervical cytology samples with equivocal or low-grade lesions. *Br J Cancer.* 2009; 101:511–517. [PubMed: 19623178]
24. Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesions Triage Study (ALTS) Group. Human papillomavirus testing for triage of women with cytologic evidence of low-grade squamous intraepithelial lesions: baseline data from a randomized trial. *J Natl Cancer Inst.* 2000; 92:397–402. [PubMed: 10700419]
25. The ASCUS-LSIL Triage Study (ALTS) Group. A randomized trial on the management of low-grade squamous intraepithelial lesion cytology interpretations. *Am J Obstet Gynecol.* 2003; 188:1393–1400. [PubMed: 12824968]

26. Gold, MR.; Siegel, JE.; Russell, LB.; Weinstein, MC., editors. Cost-effectiveness in health and medicine. New York, NY: Oxford University Press; 1996.
27. Elliott, AC.; Reisch, JS. [Accessed on February 18, 2011] Implementing a multiple comparison test for proportions in a 2xc crosstabulation in SAS®. Available at: <http://www2.sas.com/proceedings/sugi31/204-31.pdf>
28. Centers for Medicare and Medicaid Services. [Accessed on June 24, 2010] Clinical Diagnostic Laboratory Fee schedule. Apr. 2010 Available at: http://www.cms.gov/ClinicalLabFeesched/02_clinlab.asp
29. Centers for Medicare and Medicaid Services. [Accessed on June 24, 2010] Physician Fee Relative Value File. Jan. 2010 (revised). Available at: <http://www.cms.gov/physicianfeesched/pfsrvf/ItemDetail.asp?ItemID=CMS1230197>
30. Pretorius RG, Belinson JL, Zhang WH, Burchette RF, Elson P, Qiao YL. The colposcopic impression. Is it influenced by the colposcopist's knowledge of the findings on the referral Papanicolaou smear? J Reprod Med. 2001; 46:724–728. [PubMed: 11547646]
31. Pretorius RG, Zhang WH, Belinson JL, Huang MN, Wu LY, Zhang X, et al. Colposcopically directed biopsy, random cervical biopsy, and endocervical curettage in the diagnosis of cervical intraepithelial neoplasia II or worse. Am J Obstet Gynecol. 2004; 191:430–434. [PubMed: 15343217]
32. Jeronimo J, Massad S, Schiffman M. Visual appearance of the uterine cervix: correlation with human papillomavirus detection and type. Am J Obstet Gynecol. 2007; 197:47.e1–47.e8. [PubMed: 17618753]
33. Gage JC, Hanson VW, Abbey K, Dippery S, Gardner S, Kubota J, et al. Number of cervical biopsies and sensitivity of colposcopy. Obstet Gynecol. 2006; 108:264–272. [PubMed: 16880294]
34. Buck C, Donner A. The design of controlled experiments in the evaluation of non-therapeutic interventions. J Chron Dis. 1982; 35:531–538. [PubMed: 7085845]
35. Eddy DM. The Confidence Profile Method: A Bayesian method for assessing health technologies. Operations Research. 1989; 37:210–28. [PubMed: 10292450]
36. David, M. [Accessed on February 19, 2011] Eddy: Top 10 Projects. Available at: <http://www.davidmeddy.com/Top10projects.htm#The%20Confidence%20Profile%20Method>

Figure 1. Natural history of cervical carcinogenesis

Schiffman M, Kjaer SK. Natural history of anogenital human papillomavirus infection and neoplasia. *J Natl Cancer Inst Monogr* 2003;31:14–19, by permission of Oxford University Press.¹³ This figure depicts the natural history of cervical carcinogenesis. HPV incidence peaks around the time of sexual debut, and while most infections are benign and clear within 1 to 2 years of sexual transmission, some will persist. Persistent infections with any of the 10 to 15 carcinogenic HPV types are at risk of developing into precancer. Histologically confirmed CIN3 is at high risk of progressing to cancer and requires definitive treatment. HPV: human papillomavirus; CIN: cervical intraepithelial neoplasia.

Figure 2. Model overview

This figure displays the general structure of the decision tree model¹⁹ used to evaluate costs and number of cases of CIN3 detected among women with a referral cytology of LSIL in ALTS. Management strategies included immediate colposcopy (IC), HPV DNA testing, and conservative management (CM) with up to three repeat cytology tests at enrollment, 6 months, and 12 months. Model health state was based on the worst 2-year QC pathology group diagnosis. Triage test results were only relevant in the HPV DNA testing arm (Hybrid Capture 2 [HC2]) and CM (cytology) arms, and women were referred to colposcopy for a positive or missing HC2 test, or HSIL+, respectively. Colposcopy results were based on evaluation at clinical centers, with a treatment threshold of CIN2. The outcome was cost per case of CIN3 detected. CIN: cervical intraepithelial neoplasia; LSIL: low-grade intraepithelial lesion; ALTS: ASCUS/LSIL Triage Study; IC: immediate colposcopy; HPV: human papillomavirus; CM: conservative management; HC2: Hybrid Capture 2; QC: quality control; HSIL: high-grade squamous intraepithelial lesion.

	<21 years [†]			21-24 years [†]			25+ years [†]		
	Number of CIN3 cases referred	% of CIN3 referrals detected (95% CI)	Number of CIN3 cases referred	% of CIN3 referrals detected (95% CI)	Number of CIN3 cases referred	% of CIN3 referrals detected (95% CI)	Number of CIN3 cases referred	% of CIN3 referrals detected (95% CI)	
IC	23	43.5 (23.2, 65.5)	35	51.4 (34.0, 68.6)	42	57.1 (41.0, 72.3)			
HPV	10	80.0 (44.4, 97.5)	10	70.0 (34.8, 93.3)	17	58.8 (32.9, 81.6)			
CM (overall)	12	91.7 (61.5, 99.8)	19	78.95 (54.4, 93.95)	24	70.8 (48.9, 87.4)			
CM 1 st visit	6	83.3 (35.9, 99.6)	15	73.3 (44.9, 92.2)	17	70.6 (44.0, 89.7)			
CM 2 nd visit	3	100.0 (29.2, 100.0)	2	100.0 (15.8, 100.0)	6	83.3 (35.9, 99.6)			
CM 3 rd visit	3	100.0 (29.2, 100.0)	2	100.0 (15.8, 100.0)	1	0 (0, 97.5)			
P value [‡]		0.008		0.1		0.5			
Positive predictive value of colposcopy for detection of CIN3^{**}									
	Number of positive colposcopies	Positive predictive value (95% CI)	Number of positive colposcopies	Positive predictive value (95% CI)	Number of positive colposcopies	Positive predictive value (95% CI)			
IC	27	37.0 (19.4, 57.6)	39	46.2 (30.1, 62.8)	43	55.8 (39.9, 70.9)			
HPV	13	61.5 (31.6, 86.1)	11	63.6 (30.8, 89.1)	14	71.4 (41.9, 91.6)			
CM (overall)	14	78.6 (49.2, 95.3)	22	68.2 (45.1, 86.1)	28	60.7 (40.6, 78.5)			
CM 1 st visit	8	62.5 (24.5, 91.5)	18	61.1 (35.75, 82.7)	17	70.6 (44.0, 89.7)			
CM 2 nd visit	3	100 (29.2, 100)	2	100 (15.8, 100)	9	55.6 (21.2, 86.3)			
CM 3 rd visit	3	100 (29.2, 100)	2	100 (15.8, 100)	2	0 (0.0, 84.2)			
P value [‡]		0.03		0.2		0.6			

* We excluded n=2 (IC), n=1 (HPV), and n=1 (CM) cases of cancer detected at enrollment, although there was n=1 cancer case included * in the oldest tertile of the CM arm that was not detected until 18 months; this was a major failure of the arm. We report exact binomial confidence intervals. IC = immediate colposcopy arm; HPV = HPV DNA triage arm; CM = conservative management arm; CIN3 = cervical intraepithelial neoplasia grade 3.

[†] P values from Fisher exact test for comparison of age tertiles within each study arm were not significant.

[‡] P values from Fisher exact test for comparison between study arms. The CM arm (overall), as opposed to by visit, was compared to other arms. A Tukey-type multiple comparison test for proportions demonstrated that associations found for colposcopic referrals were significantly different between all three study arms, and sensitivity of colposcopy and positive predictive value were significantly different between the IC and CM arms in the youngest age tertile.

[§] Percentage referred is out of the total number of women in the arm, regardless of missed visits. For IC and HPV referrals, only enrollment visit referrals are considered. For CM referrals, the enrollment, 6 month, and 12 month visit referrals are considered. We present results for the CM arm overall and by visit; the number of women in the arm decreases by visit as we censor women with HSIL+ and/or colposcopy at previous visits. For HPV referrals, missing results were defaulted as positive and referred to colposcopy.

^{||} Performance of management strategy for detection of CIN3^{††} is defined as the probability of a clinical center histopathologic diagnosis of CIN2+ (at enrollment in the IC and HPV arms; at enrollment, 6 month, or 12 month visit in the CM arm) among women in an arm who have a worst 2-year QC pathology group diagnosis of CIN3, and thus conveys the combined sensitivity of the triage test (in the case

of the HPV and CM arms) and clinical center colposcopy at identifying women with CIN3. For the CM arm, we present results overall and by visit; the number of CIN3 cases decreases by visit as we censor women with HSIL+ and/or colposcopy at previous visits.

[†]Sensitivity of colposcopy for detection of CIN3[†] is defined as the probability of a clinical center histopathologic diagnosis of CIN2+ (at enrollment in the IC and HPV arms; at enrollment, 6 month, and 12 month visits in the CM arm) among women in an arm who have a worst 2-year QC pathology group diagnosis of CIN3 *that were referred to colposcopy*; this measure assesses the performance of colposcopy within an arm and does not account for how women got to colposcopy. For the CM arm, we present results overall and by visit; the number of CIN3 cases by visit represents the number referred due to HSIL+ at that visit.

^{**}Positive predictive value of colposcopy for detection of CIN3^{**} is defined as the probability of having a worst 2-year QC pathology group diagnosis of CIN3 among women who received a clinical center histopathologic diagnosis of CIN2+ (at enrollment in the IC and HPV arms; at enrollment, 6 month, or 12 month visits in the CM arm).

Table 2

Direct medical costs.*

Intervention	Cost (2010 U.S. \$)
Cytology	
Liquid-based Pap smear	29.02
Office visit	65.82
Physician review (for HSIL+ results) [†]	15.13
Human papillomavirus test	28.72
Colposcopy (no biopsy + office visit)	170.69
Colposcopy (with biopsy + office visit)	203.78

* Sources: 2010 Clinical Diagnostic Laboratory Fee Schedule; 2010 Physician Fee Schedule.^{29,29} HSIL: high-grade squamous intraepithelial lesion.

[†] While cytology specimens are typically evaluated by a cytotechnologist, with results of ASC-US+ requiring review by a pathologist in a clinical setting, we conservatively assumed that only results of HSIL+ would be subject to pathologist review. This assumption should bias results in favor of CM by lowering costs.¹⁹

Table 3

Performance of colposcopy by study arm, enrollment cytology, and age tertile.*

Sensitivity of colposcopy for detection of CIN3, given HSIL+ and HC2+ [†]										
<21 years					21–24 years			25+ years		
	Number of CIN3 cases referred	Sensitivity of colposcopy (95% CI)	Number of CIN3 cases referred	Sensitivity of colposcopy (95% CI)	Number of CIN3 cases referred	Sensitivity of colposcopy (95% CI)	Number of CIN3 cases referred	Sensitivity of colposcopy (95% CI)	Number of CIN3 cases referred	P value [‡]
IC	8	100.0 (63.1, 100.0)	10	70.0 (34.8, 93.3)	16	87.5 (61.7, 98.4)	0.2			
HPV	1	100.0 (2.5, 100.0)	2	100.0 (15.8, 100.0)	6	83.3 (35.9, 99.6)	1.0			
CM (overall)	12	91.7 (61.5, 99.8)	19	78.9 (54.4, 93.9)	24	70.8 (48.9, 87.4)	0.4			
P value [§]		1.0		0.8		0.5				
Sensitivity of colposcopy for detection of CIN3, given <HSIL and HC2+ at enrollment [†]										
	Number of CIN3 cases referred	Sensitivity of colposcopy (95% CI)	Number of CIN3 cases referred	Sensitivity of colposcopy (95% CI)	Number of CIN3 cases referred	Sensitivity of colposcopy (95% CI)	Number of CIN3 cases referred	Sensitivity of colposcopy (95% CI)	Number of CIN3 cases referred	P value [‡]
IC	15	13.3 (1.7, 40.4)	22	45.5 (24.4, 67.8)	23	34.8 (16.4, 57.3)	0.1			
HPV	9	77.8 (40.0, 97.2)	8	62.5 (24.5, 91.5)	11	45.5 (16.7, 76.6)	0.3			
P value [§]		0.003		0.7		0.7				
P value (<HSIL vs. HSIL+, IC arm)		<0.0001		0.3		0.001				
P value (<HSIL vs. HSIL+, HPV arm)		1.0		1.0		0.3				
Positive predictive value of colposcopy for detection of CIN3, given HSIL+ and HC2+ [†]										
	Number of positive colposcopies	Positive predictive value (95% CI)	Number of positive colposcopies	Positive predictive value (95% CI)	Number of positive colposcopies	Positive predictive value (95% CI)	Number of positive colposcopies	Positive predictive value (95% CI)	Number of positive colposcopies	P value [‡]
IC	13	61.5 (31.6, 86.1)	14	50.0 (23.0, 77.0)	21	66.7 (43.0, 85.4)	0.7			
HPV	3	33.3 (0.8, 90.6)	5	40.0 (5.3, 85.3)	7	71.4 (29.0, 96.3)	0.5			
CM (overall)	14	78.6 (49.2, 95.3)	22	68.2 (45.1, 86.1)	28	60.7 (40.6, 78.5)	0.5			
P value [§]		0.2		0.4		0.9				
Positive predictive value of colposcopy for detection of CIN3, given <HSIL and HC2+ at enrollment [†]										
	Number of positive colposcopies	Positive predictive value (95% CI)	Number of positive colposcopies	Positive predictive value (95% CI)	Number of positive colposcopies	Positive predictive value (95% CI)	Number of positive colposcopies	Positive predictive value (95% CI)	Number of positive colposcopies	P value [‡]
IC	14	14.3 (1.8, 42.8)	22	45.5 (24.4, 67.8)	20	40.0 (19.1, 63.9)	0.13			
HPV	10	70.0 (34.8, 93.3)	6	83.3 (35.9, 99.6)	7	71.4 (29.0, 96.3)	1.0			

Sensitivity of colposcopy for detection of CIN3, given HSIL+ and HC2+ [†]		
	<21 years	21–24 years
P value [§]	0.01	0.2
P value $\parallel$ (<HSIL vs. HSIL+, IC arm)	0.02	1.0
P value $\parallel$ (<HSIL vs. HSIL+, HPV arm)	0.5	0.2
		1.0

* Colposcopic performance in the IC and HPV arms is evaluated at the enrollment visit, and in the CM arm is evaluated at the enrollment, 6 month, and 12 month visits combined. We report exact binomial confidence intervals. HSIL+: high-grade squamous intraepithelial lesion on cytology; <HSIL: less than high-grade squamous intraepithelial lesion on cytology (i.e., Normal, ASC-US, ASC-H, LSIL); CIN3: cervical intraepithelial neoplasia grade 3; HC2+: Hybrid Capture 2 positive or missing at enrollment in the IC and HPV arms (n=3 missing in IC arm; n=1 missing in HPV arm among women with CIN3) or at the visit of colposcopic referral for HSIL+ in the CM arm (n=0 missing among women with CIN3).

[†]“Sensitivity of colposcopy for detection of CIN3” is defined as the probability of a clinical center histopathologic diagnosis of CIN2+ (at enrollment in the IC and HPV arms; at enrollment, 6 month, and 12 month visits in the CM arm) among women in an arm who have a worst 2-year QC pathology group diagnosis of CIN3 *that were referred to colposcopy*. For the CM arm, we present results overall and by visit; the number of CIN3 cases by visit represents the number referred due to HSIL+ at that visit. “Positive predictive value of colposcopy for detection of CIN3” is defined as the probability of having a worst 2-year QC pathology group diagnosis of CIN3 among women who received a histopathologic diagnosis of CIN2+ (at enrollment in the IC and HPV arms; at enrollment, 6 month, or 12 month visits in the CM arm) at a clinical center.

[§]Fisher’s exact test, comparing across age tertiles.

[§]Fisher’s exact test, comparing across study arms.

$\parallel$ Fisher’s exact test, comparing across cytology interpretations (<HSIL vs. HSIL+) within a study arm.

Table 4

Costs, effectiveness, and cost-effectiveness of each management strategy for detection of CIN3, for Steps 1–3.*

Step 1: Unadjusted ALTS data.									
<21 years					21–24 years				
Strategy	Costs [†]	Effects [‡]	ICER ^{\$}	Strategy	Costs [†]	Effects [‡]	ICER ^{\$}	Strategy	Costs [†]
CM (1 st visit)	112.96	24.1	--	CM (1 st visit)	126.89	50.2	--	CM (1 st visit)	120.62
CM (2 nd visit)	186.28	38.6	//	HPV	180.37	104.4	\$990	HPV	176.98
HPV	197.62	115.9	\$920	IC	198.34	77.3	//	IC	199.19
IC	198.07	59.1	//	CM (2 nd visit)	201.38	59.4	//	CM (2 nd visit)	203.85
CM (3 rd visit)	249.28	53.1	//	CM (3 rd visit)	260.11	68.5	//	CM (3 rd visit)	265.10
Step 2a: Adjustment for variation in prevalence of CIN3.									
<21 years					21–24 years				
Strategy	Costs [†]	Effects [‡]	ICER ^{\$}	Strategy	Costs [†]	Effects [‡]	ICER ^{\$}	Strategy	Costs [†]
CM (1 st visit)	113.12	25.0	--	CM (1 st visit)	128.33	56.0	--	CM (1 st visit)	121.75
CM (2 nd visit)	186.52	39.9	//	HPV	179.92	90.7	\$1,490	HPV	174.20
HPV	194.67	107.8	\$980	IC	197.12	73.3	//	IC	198.35
IC	196.92	58.6	//	CM (2 nd visit)	202.51	66.2	//	CM (2 nd visit)	204.92
CM (3 rd visit)	249.58	54.9	//	CM (3 rd visit)	261.08	76.4	//	CM (3 rd visit)	265.65
Step 2b: Adjustment for ascertainment bias (inferior performance in IC arm).									
<21 years					21–24 years				
Strategy	Costs [†]	Effects [‡]	ICER ^{\$}	Strategy	Costs [†]	Effects [‡]	ICER ^{\$}	Strategy	Costs [†]
CM (1 st visit)	112.96	24.1	--	CM (1 st visit)	126.89	50.2	--	CM (1 st visit)	120.62
CM (2 nd visit)	186.28	38.6	//	HPV	180.37	104.4	\$990	HPV	176.98
HPV	197.62	115.9	\$920	IC	198.48	93.4	//	IC	195.73
IC	198.46	116.4	\$1,680	CM (2 nd visit)	201.38	59.4	//	CM (2 nd visit)	203.85
CM (3 rd visit)	249.28	53.1	//	CM (3 rd visit)	260.11	68.5	//	CM (3 rd visit)	265.10
Step 3: Simultaneous adjustment for variation in prevalence of CIN3 and ascertainment bias.									
<21 years					21–24 years				
Strategy	Costs [†]	Effects [‡]	ICER ^{\$}	Strategy	Costs [†]	Effects [‡]	ICER ^{\$}	Strategy	Costs [†]
CM (1 st visit)	113.12	25.0	--	CM (1 st visit)	128.33	56.0	--	CM (1 st visit)	121.75
CM (2 nd visit)	186.28	38.6	//	HPV	180.37	104.4	\$990	HPV	176.98
HPV	197.62	115.9	\$920	IC	198.48	93.4	//	IC	195.73
IC	198.46	116.4	\$1,680	CM (2 nd visit)	201.38	59.4	//	CM (2 nd visit)	203.85
CM (3 rd visit)	249.28	53.1	//	CM (3 rd visit)	260.11	68.5	//	CM (3 rd visit)	265.10

Step 1: Unadjusted ALTS data.											
<21 years				21-24 years				25 years			
Strategy	Costs [†]	Effects [‡]	ICER [§]	Strategy	Costs [†]	Effects [‡]	ICER [§]	Strategy	Costs [†]	Effects [‡]	ICER [§]
CM (2 nd visit)	186.52	39.9	//	HPV	179.92	90.7	\$1,490	HPV	174.20	86.3	//
HPV	194.67	107.8	//	IC	197.25	88.6	//	IC	194.74	103.2	\$1,460
IC	197.30	115.3	\$930	CM (2 nd visit)	202.51	66.2	//	CM (2 nd visit)	204.92	75.3	//
CM (3 rd visit)	249.58	54.9	//	CM (3 rd visit)	261.08	76.4	//	CM (3 rd visit)	265.65	75.3	//

* Strategies are presented in order of increasing cost. ICER = incremental cost-effectiveness ratio; IC = immediate colposcopy arm; HPV = HPV DNA triage arm; CM = conservative management arm; CIN3 = cervical intraepithelial neoplasia grade 3.

[†] Costs per hypothetical woman in the model are reported in 2010 U.S. dollars.

[‡] Effects are reported as cases of CIN3+ detected per 1000 women in the model.

[§] ICERs are reported as the cost per case of CIN3+ detected (\$/case detected).

// Dominated strategy (i.e., was more costly and less effective than other strategies or had a higher incremental cost-effectiveness ratio than more effective strategies)