Vaginal Discharge Syndromic Management in Family Planning Programs: A Review of the Literature

Laurette Cucuzza

Introduction

One of the most important mandates from the International Conference on Population and Development, Cairo (1994) was the call for integration of family planning and sexually transmitted infection (STI) services. It was argued that this integration of services would allow more women in resource-poor countries access to health care. This increased access would be achieved through utilization of existing personnel and infrastructure and consequently would not increase cost.

The high cost of traditional laboratory testing for STIs and the relatively sophisticated clinical infrastructure required to provide these tests are problems that need to be overcome when integrating STI services into family planning programs in developing countries. To overcome these problems, the World Health Organization's "syndromic approach" to managing STIs originally designed for STD clinic clients, was widely adopted for use in family planning clinics (WHO, 1993). These algorithms, or flow charts, allow providers to follow groups of signs and symptoms (syndromes) to diagnose and treat various RTIs. For the sexually-transmitted RTIs that commonly affect women - gonorrhea (NG), chlamydia (CT) and trichomoniasis (TV) - the entry point of the algorithm is the client complaint of abnormal vaginal discharge. Vaginal discharge is also used to diagnose endogenous reproductive tract infections, such as bacterial vaginosis (BV) (Footnote A) and candidiasis (CA) (see table 1). The syndromic approach was thought to be simple, feasible and cost effective.

Critics of this approach, however, point out that the algorithm performs poorly in detecting gonorrheal and chlamydial infections. These infections are usually limited to the cervix and produce little or no discharge, resulting in many missed cases of the most serious infections. The algorithm also falsely identifies many women as having an STI, and recommends treating for both gonorrhea and chlamydia, as well as trichomoniasis, bacterial vaginosis and candidiasis, resulting in costly over-treatment.

Many modifications to the syndromic approach have been developed in response to these criticisms, including: the addition of clinical examinations, risk assessments, and diagnostic tools such as LED and microscopy. This paper reviews the literature on syndromic management of RTIs and assesses its usefulness among family planning populations in developing countries. We focus on several factors in our review of the algorithms discussed in the literature -

- Validity: It must truly measure the condition it is designed to identify
- Reliability: consistently giving the same results when used repeatedly
- Feasibility: simple enough to use in the field
- Acceptability: to both the provider and to the client
- Affordability: administered at a reasonable cost.

Validity is measured by sensitivity (the ability of the test to correctly identify actual cases of infection), specificity (the correct identification of those not infected), and positive predictive value (PPV), (the proportion of cases identified as infected which truly have the disease) (Footnote B).

Highly sensitive tests will identify most of the actual cases of infection, but also have a higher proportion of false positives, leading to over treatment and higher costs per case treated. Improved specificity will result in more individuals being correctly identified as not having disease, but also a greater number of false negatives, and more cases of infection going untreated. The ideal would be to have 100% sensitivity and 100% specificity, but the reality is that there is always a trade off between the two. One must choose the best combination based on the
priorities of the situation. Higher levels of disease prevalence in the population create higher ppv values (see Figure 1).

To determine the validity of the algorithms tested in these studies, and to overcome the weaknesses previously mentioned (missed cases and over-treatment), the most important measures are sensitivity and positive predictive value.

Methodology

This review was accomplished through a computer search of peer-reviewed and unpublished studies containing the key words "syndromic management", "vaginal discharge", "family planning", "integration of family planning and STD", "algorithm", "reproductive tract infections" and "assessment of syndromic management". The period covered was January 1990 - August 2000. The search was conducted on Medline, AIDSline, and Popline and was limited to articles in English. To examine issues of validity, we searched for studies that compared algorithms with laboratory tests as the "gold standard". We found 11 studies assessing the prevalence of RTIs and the effectiveness of syndromic management in family planning clinic settings for review (Costello et al., 1994; Ronsman et al., 1996; Gertig et al., 1997; Kapiga et al., 1998; Ryan et al., 1998; Schneider et al., 1998; Maggwa et al., 1999; Solo et al., 1999; Tyndall et al., 1999; Vishwanath et al., 2000; Ward et al., 2000).

Results

All 11 studies reviewed in this paper found somewhat higher than expected prevalences of RTIs in a population traditionally considered "low-risk" (family planning clients) (see Figure 2). Despite these unexpectedly high levels of prevalence, positive predictive values were relatively low. Prevalence of endogenous (non-sexually transmitted) RTIs was even higher, and correspondingly the PPVs of algorithms for identifying these infections were somewhat higher, although not all studies measured these infections.

Because of their physiology and anatomy, women suffer higher rates and graver consequences of sexually transmitted infections than men. The World Health Organization estimates that 60-70% of gonococcal and chlamydial infections among women are asymptomatic and will be missed by any algorithm based on symptoms alone (WHO, 2000). The studies reviewed in this paper found similar rates of asymptomatic infection. For example, researchers in Nakuru found that 77% of family planning clients diagnosed by lab tests with any RTI had no symptoms (Solo et al., 1999). In Tanzania, 62.2% of women with cervicitis (CT and/or NG) and 67.6% with vaginitis (TV, BV and/or CA) were asymptomatic (Kapiga et al., 1998).

Predictive Validity of Vaginal Discharge Algorithms:

Three studies (Ronmans et al., 1996; Kapiga et al., 1998; Maggwa et al., 1999) measured the effectiveness of algorithms based only on the woman's complaint of vaginal discharge (Footnote C). The sensitivity of these algorithms for diagnosing cervical infection (NG and/or CT) varied from 24% (CT only) to 46%. The PPV was low (<16%) in all studies. When used to diagnose vaginal infections these algorithms performed a little better - sensitivity ranged from 35% to 62.7% (for CA only) and PPV ranged from 12.2% (CA only) to 28%. This type of algorithm not only missed most of the cases of cervical infection it also falsely identified many women as being infected and performed only marginally better for vaginal infections. These studies illustrate the poor performance of algorithms based on symptoms only. In contrast, when the symptoms-only algorithm was applied to women seeking care for RTIs (vs. all family planning clients) the sensitivity increased to 72% for cervical infections and 69% for vaginal infections. However the PPV was still low - 9% and 34% respectively, meaning many women were misdiagnosed and treated unnecessarily.

One would expect the addition of a clinical examination, using a speculum and bimanual exam to discover signs of infection, to improve algorithm performance. Six studies (see Table 2) tested algorithms that used clinical examination. Three of these studies looked at algorithms that combined signs and symptoms while the others tested sign-only algorithms. Results from these studies were mixed. Researchers in Turkey found that the addition
of a clinical examination to the symptom-only algorithm actually decreased both sensitivity and PPV (CT only). Researchers in Zimbabwe found that the addition of a clinical exam to the symptom-only algorithm decreased sensitivity but increased specificity and PPV for both cervical and vaginal infections. In most of the studies, the specificity of the sign-based algorithm (with and without symptoms) was around 90% for both vaginal and cervical infections, showing a fairly good ability to accurately identify uninfected women. However, since the results of a clinical examination, like self-report of symptoms, are subjective, much depends on the skill and experience of the clinician. Skilled staff may be in short supply and training will be expensive.

Predictive Validity of Risk Assessment

Does risk assessment improve an algorithm's ability to identify sexually transmitted infections? The World Health Organization created syndromic management to better manage STIs in symptomatic clients and to find cases in asymptomatic and low-symptomatic populations (WHO, 1993). Including risk assessment of behavioral, historical, and demographic characteristics has been suggested as a way to improve the case finding ability of the algorithm. The guidelines emphasize that the risk questions should be adapted to local prevalence of disease and risk factors, in order to increase its predictive value and to make it more acceptable to local populations.

Six studies looked at risk in isolation (Costello et al., 1994; Ronsman et al., 1996; Gertig et al., 1997; Solo et al., 1999; Tyndall et al., 1999; Ward et al., 2000), either asking risk questions or applying an algorithm calculated from logistic regression analysis of demographic and behavioral risks obtained from interviews. Eleven separate algorithms were tested with varying risk factors and decision criteria. Sensitivity ranged from 0% (CT only) (Ronsmans et al, 1996) to 60% (Tyndall et al., 1999). Positive predictive values were low. Only two algorithms had PPVs greater than 25%: 33.9% in Jamaica and 60% in Nakuru, Kenya.

The variability of these measures demonstrates the problem with risk factors; where they are not sensitive, they are probably not appropriate/applicable to the population. Algorithms based on risk alone performed better when the risks were determined for the population in question.

Two studies (Ronsmans et al, 1996; Ryan et al., 1998) tested the original WHO risk assessment model for cervical infection,(Footnote D) with entry point of vaginal discharge symptom and standard risk questions (Footnote E). Both found low PPVs (<12%). Moroccan scientists found slightly improved PPVs of around 22% for vaginitis. Researchers measured the addition of clinical signs from pelvic examination to the original, risk-inclusive WHO algorithm in three of the studies, and found widely varying results in sensitivity for cervical infections (possibly due to the subjectivity of clinical exams). They ranged from 13.5% in Tanzania to 85.7% in Moroccan family planning clients (Ronsmans et al, 1996; Kapiga et al., 1998; Ryan et al., 1998). Positive predictive values, however, were uniformly low (<16%). Moroccan researchers found it highly sensitive for BV and TV (90.8%), but not specific (7.5%), and PPV was 26.8%.

Researchers in three studies examined algorithms using adapted risk assessments, in addition to a clinical exam (Costello et al. 1994; Gertig et al., 1997; Tyndall et al., 1999). Results for cervical infections varied widely, identifying just 2% of cases in Nairobi, Kenya to 76% of gonorrhea infection in Dar es Salaam. Two of these studies also tested the algorithms for the vaginal infection, trichomoniasis (TV) and again, results varied considerably.

Researchers in India using an algorithm combining symptoms, risk factors, and a clinical exam found ppvs of 38% for bacterial vaginosis and trichomoniasis combined, and 88% for candidiasis, and concluded that this may be useful for management of vaginal infections.

Since hierarchical algorithms, based on symptoms and signs, perform so poorly in identifying cases of cervical infection, researchers tried creating non-hierarchical weighted-risk algorithms. In non-hierarchical algorithms, each risk factor is assigned a numerical score, or weight, based upon its association with cervical infection, derived from logistic regression analysis. Three studies (Gertig et al., 1997; Schneider et al., 1998; Tyndall et al., 1999) using this method, found that sensitivity and specificity varied depending on cut off levels used, but no PPV was greater than 48% (WHO, 1993), and many were around 5%.
Tyndall (1999) and Ward (2000) adapted the WHO algorithm with weighted scoring and both found PPVs of about 25%. Researchers in two studies (Gertig et al., 1997; Kapiga et al., 1998) applied an algorithm that had been devised for use with pregnant women and prostitutes in Zaire (Vuylsteke et al., 1993). But in this family planning population, PPVs never reached higher than 7% (for NG only) and 22% (for CT & NG). These results emphasize the need for situation-specific risk questions, as they could not be generalized to a family planning population in a different country. Removal of the vaginal discharge symptom/sign as an entry point, and supplementation with risk, enhances algorithm performance somewhat, but not enough to use for case finding. Some investigators thought that the addition of simple diagnostic tools, such as a leukocyte esterase dipstick (LED) or basic wet-mount microscopy, might improve predictive values of algorithms and virtually all of the studies noted a need for such tools. LED, which has been useful in screening men with urethritis (Tyndall et al., 1994), measures the level of white blood cell enzymes in urine indicative of infection. Researchers in Jamaica devised two weighted-scoring algorithms using LED; both were more specific than sensitive, with PPVs of about 23% for cervical infections (Ward et al., 2000).

Scientists in Kenya (Tyndall et al., 1999) tested several models, including risks, LED levels and/or microscopy, with various combinations and cut off levels. The algorithms with microscopy reached a high of 50% PPV. The ones using microscopy and LED combined, reached a PPV as high as 67%, but a woman would have to have all six risk factors, which is very unlikely in "low risk" family planning populations.

Thus, as is consistent with findings of other studies (Knud-Hansen et al., 1991; Vuylsteke et al., 1993), the addition of these tools did not increase sensitivity or predictive power of algorithms enough for utility in finding/managing cervical infections. Researchers in India (Vishwanath et al., 2000) found that using microscopy improved the sensitivity, specificity and positive predictive value of the algorithm for the detection of trichomonas to 100% The addition of microscopy had little effect on the specificity and PPV for candidiasis, which were at 98%, and the sensitivity increased to 57%. This indicates that it may be more useful in finding/managing vaginal infections, if areas have access to the equipment and providers are well trained. Aside from intrinsic weaknesses discussed above, risk assessment may not be culturally appropriate in countries with restraints against extramarital sex, such as in Morocco (Ryan et al., 1998). Researchers found that few women would admit to sexual risk factors when asked, and many interviewers found it very difficult to ask such questions.

Three of the studies examined whether participants felt themselves at risk of STI/HIV (Kapiga et al., 1998; Maggwa et al., 1999; Tyndall et al., 1999). Less than fifty percent of women in Zimbabwe perceived themselves at risk of disease. In the Nakuru study, only 17.1% of family planning clients felt at higher risk for infection, mostly due to the behavior of their partners. Moroccan investigators found women were at increased risk of cervical infection if they perceived risks in their partner's behavior, though few reported risk behaviors in themselves. Low perception of risk may contribute to poor performance of risk assessment.

Feasibility

Another problem with the syndromic approach is that providers do not always follow the guidelines. In Zimbabwe (Maggwa et al., 1999), only 65% of those women with symptoms and signs of vaginal discharge syndrome were diagnosed using the algorithm and treated according to national guidelines. Jamaican researchers point out that providers find the use of weighted risk scoring "complicated and bothersome" (Ward et al., 2000). Initially, in the Nakuru study (Solo et al, 1999), providers were reluctant to use the risk checklist and complained that it took too long, though use improved over time. In a related study on improving quality of STI case management in primary health care centers in South Africa, focus group discussions among providers revealed that many lack training in syndromic management, morale was low, and negative attitudes toward STI patients hindered compliance with protocols (Harrison et al, 1998).

Cost Effectiveness
As demonstrated by the widely varying, but generally low, sensitivities and positive predictive values generated by these studies, one can see that over-treatment and complications of missed cases of RTIs would be costly. The opportunity costs of provider's time being wasted on useless treatment is also large. Their time could be better spent using a preventive public health approach, providing counseling, education and condom promotion that have been proven more cost-effective (Shelton, 1999).

Very few studies have been done on cost and cost-effectiveness of syndromic management in family planning situations (Footnote G; Hawkes et al.). The study in India (Vishwanath et al., 2000) figured the cost of drugs in applying the algorithm, which came to US$2 per case, but in reality, the cost was US$4.25 per infection correctly treated. The Zimbabwe study (Maggwa et al., 1999) included a cost-effectiveness analysis, looking at cost per woman and comparing that to the estimated expenditure for all health care per capita in the country. Investigators calculated costs per family planning client and compared four RTI diagnostic algorithms. Syndromic approach for clients seeking RTI services cost the least and used only 5.3% of government per capita health spending, but 75% of RTIs were missed, and 56% of women were over-treated. Laboratory testing was not feasible at a cost of US$25.77, 54.5% of the estimated US$47.12 government per capita spending on health care. Over-treatment has several drawbacks, including increased costs for drugs, and the potential increase in drug-resistant strains of disease, requiring newer, and possibly more expensive drug therapies. Women may also suffer side effects from antibiotics.

Discussion

All eleven studies found vaginal discharge syndromic management of STIs not effective in the context of family planning programs, especially for cervical infections. Reasons for this ineffectiveness include: the large number of asymptomatic infections, a lack of correlation between vaginal discharge symptom and STIs, and the lack of risk factors that apply to family planning clients. Though prevalence of RTIs/STIs was surprisingly high for family planning clinic populations, syndromic management, as a diagnostic tool, lacked sensitivity and positive predictive value, leading to missed diagnoses and over-treatment, especially of cervical infections. Modifications, such as physical examination, risk assessment, and simple diagnostic tools, only slightly improved performance. A low perception of personal risk and lack of reported risk behaviors among family planning clients may be one reason that risk assessments haven't proven useful. Though it was adopted to reduce cost of diagnosis and management of RTIs, the syndromic approach is not a cost-effective tool. Though syndromic management is not effective for dealing with cervical infection, it may be appropriate for management of vaginal infection, as it tended to produce higher sensitivity, specificity and positive predictive value for TV, BV, and CA. A large number of women suffer from vaginal infections, as prevalence rates for TV, BV and CA in these studies demonstrate.

Finally, all of these studies point to the need for more research to produce inexpensive, accurate and easy-to-use diagnostic tools to identify and manage STIs/RTIs. Some progress has been made in the development of a dip-stick type test for the diagnosis of gonorrheal and chlamydial infections, but it may yet take some time to produce and validate (Mabey, 2000). The results from the reviewed studies suggest that syndromic management should be dropped from family planning programs for detecting gonorrhoea and chlamydia, but suggest that it may be appropriate for diagnosing vaginal infections. Until a reliable, cost-effective tool is available, family planning programs should focus on prevention through increased emphasis on condom distribution and use. Paradoxically, perhaps the most effective way to find and manage cervical infections among family planning clients is by focusing on syndromic management of men's STIs, and treatment of them and their (possibly asymptomatic) partners. Several studies have shown that urethral discharge syndromic management of chlamydia and gonorrhoea in males is both effective and cost effective (Footnote H; Alary et al., 1998; Dalabetta, Gerbase, and Holmes, 1998; Djajakusumah et al., 1998; Moherdau et al., 1998).

Acknowledgements

I would like to thank Dr. Johannes Van Dam, Jim Foreit, Dr.PH, Cynthia Green, Ph.D. & Kris Lantis of the Population Council for technical assistance, editorial assistance and readings of previous drafts.
Footnotes

a. Bacterial vaginosis, though not transmitted sexually, has been linked to pelvic inflammatory disease (PID) and increased likelihood of transmission of HIV, and therefore should be considered potentially as serious as STIs.

b. An example of positive predictive value would be if a test found 100 family planning clients positive for a chlamydia and had a ppv of 25 %, it would mean that it correctly identified 25 women truly having the disease, but also falsely identified 75 women not having the disease as positive.

c. Kapiga et al., 1998, used vaginal discharge and/or dysuria as entry point for the algorithm.

d. Researchers in Turkey only looked at chlamydia infection.

e. WHO suggested risk factors are: partner symptomatic, single, age <25, new partner last 3 mos., >1 partner.

f. This ppv is for NG, CT & TV, whereas the others were only for NG & CT.

g. A similar study in Bangladesh, involving maternal and child health and family planning clients, compared the costs of treating women according to the WHO algorithm and an adapted, speculum-based algorithm, neither of which was cost effective. Under the WHO algorithm, all cases of RTIs/STIs were treated at a cost of US$1.22 per symptomatic woman, or US$3.61 per true case. But 87% of these costs were squandered on over-treatment. The speculum-based algorithm resulted in a lower overall cost of US$0.38 per woman with symptoms or US$2.75 per true case. Over-treatment cost was lower, at 36% of this expenditure; however, it missed all cervical infections, and most STI. When compared to the estimated government expenditure of less than US$4.00 per capita, syndromic management of RTIs/STIs was clearly not cost effective. (Hawkes et al. 1999).

h. A study in Indonesia compared cost effectiveness of the WHO algorithm, the WHO algorithm with microscopy, and "gold standard" laboratory testing, and found the cost per patient to be US$2.56, US$2.80 and US$18.70, respectively. Considering the positive predictive value of 97%, virtually no money was wasted on over-treatment, and as sensitivity was 100%, no cases were missed (Djajakusumah et al. 1998).

References


