Published in final edited form as:

Nurs Res. 2010 ; 59(1): 2–10. doi:10.1097/NNR.0b013e3181c3b9dd.

Accuracy of Vaginal Symptom Self-Diagnosis Algorithms for Deployed Military Women

Nancy A. Ryan-Wenger, PhD, CPNP, FAAN [Director of Nursing Research],
Nationwide Children's Hospital and Professor Emerita, The Ohio State University College of Nursing, Columbus, Ohio

Jeremy L. Neal, PhD, CNM [Assistant Professor],
The Ohio State University College of Nursing, Columbus, Ohio

Ashley S. Jones, RN, BSN [Doctoral Student], and
The Ohio State University College of Nursing, Columbus, Ohio

Nancy K. Lowe, PhD, CNM, FACNM, FAAN [Professor and Chair]
Division of Women, Children & Family Health, College of Nursing, University of Colorado Denver, Denver, Colorado

Abstract

Background—Deployed military women have an increased risk for development of vaginitis due to extreme temperatures, primitive sanitation, hygiene and laundry facilities, and unavailable or unacceptable health care resources. The Women in the Military Self-Diagnosis (WMSD) and Treatment Kit was developed as a field-expedient solution to this problem.

Objectives—The primary study aims were to evaluate the accuracy of women's self-diagnosis of vaginal symptoms and 8 diagnostic algorithms, and to predict potential self-medication omission and commission error rates.

Method—Participants included 546 active duty, deployable Army (43.3%) and Navy (53.6%) women with vaginal symptoms who sought health care at troop medical clinics on base. In the clinic lavatory, women conducted a self-diagnosis using a sterile cotton swab to obtain vaginal fluid, a FemExam® card to measure positive or negative pH and amines, and the investigator-developed WMSD Decision-Making guide. Potential self-diagnoses were "bacterial infection" (bacterial vaginosis and/or trichomonas vaginitis - BV/TV), "yeast infection" (candida vaginitis - CV), "no infection/normal", or "unclear." The Affirm VPIII® laboratory reference standard was used to detect clinically significant amounts of vaginal fluid DNA for organisms associated with BV, TV and CV.

Results—Women's self-diagnostic accuracy was 56% for BV/TV and 69.2% for CV. False positives would have led to a self-medication commission error rate of 20.3% for BV/TV and 8% for CV. Potential self-medication omission error rates due to false negatives were 23.7% for BV/TV and 24.8% for CV. The positive predictive power (PPV) of diagnostic algorithms for BV/TV ranged from 0 to 78.1%, and for CV was 41.7%.

Discussion—The algorithms were based on clinical diagnostic standards. The non-specific nature of vaginal symptoms, mixed infections, and a faulty device intended to measure vaginal pH and amines explain why none of the algorithms reached the goal of 95% accuracy. The next prototype of the WMSD Kit will not include non-specific vaginal signs and symptoms in favor of recently

Correspondence: Nancy A. Ryan-Wenger, PhD, CPNP, FAAN, Director of Nursing Research, Nationwide Children's Hospital, 700 Children's Drive, Columbus, OH 43205, Nancy.RyanWenger@NationwideChildrens.org.

Disclaimer: The views expressed here are those of the authors and not the official policy or position of the United States Military.

available point-of-care devices that identify antigens or enzymes of the causative BV, TV, and CV organisms.

Keywords

vaginitis; self care; military personnel

In the civilian environment, vulvovaginal pain, itching, burning, and vaginal discharge are the most common symptoms reported by women, resulting in 4.46 million office visits in 2006 (Centers for Disease Control and Prevention [CDC], 2006). If left untreated, vaginal symptoms can significantly interfere with women's physiologic functioning, quality of life, comfort, and concentration. The population of interest in this study was deployable military women. According to the Women in Military Service for America (WIMSA, 2006), 344,418 women were serving in 90% of all occupational areas of military as of July 17, 2006. The Defense Department Advisory Committee on Women in the Services (DACOWITS, 2007) reported that in June 2007, women comprised 10% ($n = 17,613$) of the deployed force in support of Operations Iraqi Freedom (Iraq) and Enduring Freedom (Afghanistan). Deployments are not likely to decrease in the next few years.

Deployment settings are typically austere, characterized by extreme temperatures, primitive sanitary conditions, and limited hygiene and laundry facilities. These factors increase military women's risk for development of vaginitis. Increased risk is compounded by inadequate management of these conditions due to unavailable or unacceptable health care resources for women. Ryan-Wenger and Lowe (2000) surveyed 1,537 military women about their symptoms of genitourinary infections and health care experiences in their home duty stations and during deployments. Of the 841 women who had been deployed, 87% ($n=732$) reported that they experienced vaginal symptoms such as itching, discharge, or foul odor at some time during deployment. Because of these symptoms nearly half of the women (48%) noted a decrease in the quality of their work performance, and 24% lost from a few hours to more than a day of work time. These consequences are of particular concern to the military because optimal health and functioning of all soldiers is critical to unit safety and success of the military mission. In focus groups conducted by DACOWITS, in our survey, and in a phenomenological study of soldier care, women evaluated deployment health care services for women as inadequate, citing lack of confidence in the knowledge and skills of the provider, lack of privacy, and lack of confidentiality (DACOWITS, 2007; Jennings, 2005; Ryan-Wenger & Lowe, 2000). In Ryan-Wenger and Lowe's survey, women stated that the nearest available health care provider during deployment was likely to be a medic. The problem is that medics are trained to save lives in combat; they are not trained to provide primary care for women. Twenty-five percent of the women said that they would not see that provider for treatment of genitourinary symptoms.

DACOWITS (2007) reported that the Defense Department is unlikely to increase the availability of specialists in women's health in deployment settings. We propose that a viable solution to the problem is a field-expedient kit for self-diagnosis and self-treatment of common genitourinary symptoms. Such kits must be easy to understand and use, small enough to carry in a uniform pocket, waterproof, contain as few items as possible, require minimal manipulation of items, produce a minimal amount of waste products, yield a reasonably accurate self-diagnosis, and include appropriate single-dose oral anti-infectives and other products to alleviate symptoms. An accurate and reliable self-diagnosis and self-treatment kit for vaginitis would provide women with self-care options and free deployed military women to focus fully on their missions. This article reports the results from a clinical trial to test the accuracy of the second prototype of the Women in the Military Self-Diagnosis (WMSD) Kit. The specific aims of this study were to:

- Compared to laboratory reference standards:

- a. Evaluate the accuracy of military women's vaginal self-diagnoses with the WMSD Kit.
 - b. Evaluate the accuracy of the eight WMSD diagnostic algorithms.
- Estimate the potential for self-treatment errors based on women's self-diagnosis accuracy.
 - Obtain military women's evaluation of the WMSD Kit.
 - Evaluate the inter-rater agreement between participants' and research advanced practice nurses' results of pH and amines testing using the FemExam® Card.

Background

The three most common vaginal infections are bacterial vaginosis (BV, 40%–50%), candida vaginitis (CV, 20%–25%), and trichomoniasis (TV, 15%–20%) (Brown, 2004; Sobel, 2005). (See Table, Supplemental Digital Content 1, which summarizes the etiology, signs, symptoms and laboratory tests for the three vaginal infections, <http://links.lww.com/Axxx>). The challenge for development of a vaginitis self-diagnosis kit is to select the combination of signs and symptoms that are likely to yield the most accurate diagnoses.

Normal vaginal flora includes lactobacilli that maintain an acid pH which prevent overgrowth of bacteria and parasites (Donders, 2007). A deficiency of lactobacilli is implicated in the overgrowth of bacteria that cause BV. In the laboratory, BV is diagnosed according to a scoring system described by Nugent, Krohn, & Hillier (1991). The Nugent system uses Gram staining and microscopic observation of lactobacilli, *Gardnerella vaginalis* and *Mobiluncus* morphotypes. On a 0 to 10 scoring system, results are categorized as: normal flora (0 to 3), intermediate flora (4 to 6), and BV (7 to 10). In the clinical setting the diagnosis of BV is based on observation in vaginal secretions of at least 3 of 4 Amsel criteria, which include (a) pH >4.5, (b) positive whiff test (fishy odor when vaginal fluids are mixed with potassium hydroxide causing the release of amines), (c) thin and homogeneous vaginal discharge, and (d) observation of clue cells with wet mount microscopy (Amsel et al., 1983; Forsum, Hallén, & Larsson, 2005; Sobel, 2005). Both Nugent and Amsel systems rely heavily on the diagnostic acumen of the clinician. Forsum, et al. found that presence of clue cells is definitively diagnostic of BV but demonstrated that a negative amine test is not necessarily indicative of the *absence* of BV. In a resource-poor setting in rural Azerbaijan, investigators concluded that, compared to Gram stain, pH and amines alone were adequately sensitive and specific for clinical diagnosis (ranges from 59% to 92%) (Posner, Kerimova, Aliyeva, & Duerr, 2005). However, in their calculations, the investigators eliminated the group of women with intermediate Gram stain scores from the analysis, thus true sensitivity and specificity were overestimated. In a study of women from a gynecologic clinic, 64 of 422 women had BV according to microscopy, and a positive pH of vaginal fluids alone identified 70.6% (n = 45) of the BV cases (Thinkhamrop, Lumbianon, Thongdrjai, Chongsomchai, & Pakarasang, 1999).

Symptomatic CV is perhaps the most distressing of the three common vaginal infections. Candida induced inflammation of the vulvovaginal area is manifested by varying intensities of swelling, redness, pruritis, dyspareunia, external dysuria, and thick curdy discharge (CDC, 2006; White & Vanthuyne, 2006). CV is described by some experts as “a syndrome rather than an infection” (White & Vanthuyne, p. iv28), which suggests that a wide range of signs and symptoms are consistent with the diagnosis of CV. About 75% of women will develop CV during their lifetime according to the CDC (2006). Ideally, diagnosis should be based on symptoms and Gram stain or wet mount microscopy of vaginal fluid from practitioner or self-taken vaginal swabs (White & Vanthuyne).

At 5 million new cases diagnosed annually, TV is more prevalent than chlamydia (3 million new cases/year) and gonorrhea (650,000 new cases/year), yet only the latter two infections are routinely screened in most clinics in the United States (Soper, 2004). Sobel (2005) and Brown (2004) argue that observation of *T. vaginalis* organism is essential to the diagnosis of TV because the clinical signs and symptoms are unreliable, and as many as 50% of cases are asymptomatic.

Despite these diagnostic standards, studies show that clinicians often misdiagnose vaginal infections. For example, in one study, 197 vaginal samples were analyzed by culture, Gram stain, microscopy and DNA hybridization with Affirm™ VPIII to derive a diagnosis of BV, TV and/or CV (Schwartz, Taras, Rusch, & Rusch, 2006). Compared to laboratory diagnoses, physicians misdiagnosed CV in 77.1% of 109 cases, BV in 61.3% of 80 cases, and 87.5% of 8 mixed infections. One reason for such high levels of inaccuracy is that many providers do not use the common office-based tests that are recommended to achieve a diagnosis. This point is illustrated by a study of diagnostic procedures used by physicians with 52 women who made 150 visits to a vaginitis clinic (Wiesenfeld & Macio, 1999). Microscopic assessment was done in 63% of the visits, and whiff and pH tests were conducted in only 3% of visits. In another study, 556 nurse practitioners and 608 physicians reported their diagnostic practices on a web-based survey (Anderson & Karasz, 2005). An average of 79% of these providers indicated that they “often or always” examined women with vaginal symptoms, 47% conducted whiff tests, and only 33.5% conducted pH tests on vaginal fluid. The accuracy of their diagnoses was not evaluated.

Some clinics use telephone triage protocols to make decisions about diagnosis and treatment based on patient report alone. A study comparing diagnoses of BV, TV and/or CV by microscopy with diagnoses by telephone nurses and practitioners showed poor agreement (Allen-Davis, Beck, Parker, Ellis, & Polley, 2002). Kappa inter-rater reliability statistics for telephone nurses ranged from -0.03 to 0.13, and for practitioners ranged from 0.29 to 0.39. In a study by Lowe, Neal, & Ryan-Wenger (2009), experienced research advanced practice registered nurses (APRN) used a full armamentarium of clinical techniques to make their clinical diagnoses, including an hour or more of time if needed, clinical history, pelvic examination, wet mount microscopy, pH and amines via FemExam®, pH via nitrazine paper, and amines via whiff test. Compared to DNA probe for clinically significant levels of *G. vaginalis*, *T. vaginalis*, and *candida* species, the APRN’s clinical diagnoses had a positive predictive value of 79% for BV, 99% for TV, and 67% for CV. Mixed infections were accurately diagnosed in 52% of cases.

Because trained professionals, microscopy, and DNA analysis are not available in military deployment environments, a self-diagnosis and self-treatment approach is needed for women’s management of these common symptoms. Self-diagnosis materials must be small, portable, useful under austere conditions, and readily accessible. An earlier prototype of the WMSD Kit included women’s observations of vaginal discharge odor, color and consistency, experience of vaginal itching, pH of vaginal fluids measured by nitrazine paper and a color chart, and odor from a whiff test after dipping a swab of vaginal fluid in a small vial of potassium hydroxide (Lowe & Ryan-Wenger, 2000). This information from 72 symptomatic women was compared with the same information obtained by APRNs. The women were not required to distinguish between BV and TV in their self-diagnoses because treatment with oral metronidazole in deployment situations would be the same for both BV and TV. Sixty-eight percent of the women’s self-diagnoses matched the APRN diagnoses. Women’s interpretations of vaginal fluid color and consistency had positive predictive power (PPV) of $\leq 52\%$ and therefore were not good predictors of BV/TV or CV. The symptom of vaginal itching had a PPV of 62% for diagnosis of CV, and the PPV for absence of itching was 68% for diagnosis of BV. Nitrazine pH and whiff tests were cumbersome for the women, but yielded higher PPVs of 65% to 84%.

This report focuses on the performance of a new prototype of the WMSD Kit in which the measurement of pH, amines and the symptom of vaginal itching for self-diagnosis of BV/TV and CV were retained, and results were compared to laboratory standards. The parent study involved evaluating military women's use of a self-diagnosis kit and diagnostic algorithms for vaginal and urinary tract infection signs and symptoms (1 R01 NR07662-01A1, PI: N. A. Ryan-Wenger). This report focuses on evaluation of women's use of the vaginitis portion of the kit.

Multiple theories provided the conceptual framework for the study. Self-care theory (Orem, 2001) was the theoretical basis for development of the WMSD Kit. Self-care includes perception and monitoring of symptoms, self-diagnosis of symptoms, decision-making about symptom management, and evaluation of symptom management, all of which are common, but largely immeasurable, aspects of daily life. Epidemiological theory and methods were used to evaluate the accuracy of the diagnostic algorithms (Brownson & Pettiti, 2006). Public health and molecular epidemiologic sciences contributed to understanding how combinations of specific agents (microorganisms), hosts (women) and environmental factors (e.g., access to hygiene facilities, sexual activity) result in vaginal infections. Epidemiologic statistical methods, including prevalence, predictive values, sensitivity, and specificity provided the framework for evaluating the effectiveness of diagnostic tools (Brownson & Pettiti).

Methods

Sample

The target population was deployable active duty military women experiencing vaginal symptoms. All women at the four sites who reported to an Army or Navy troop medical clinic with vaginal symptoms were eligible to participate. Pregnant women were excluded because they are not deployable for up to 3 months after delivery, and could not receive some of the single-dose oral medications that were an integral part of the parent study. Women who were menstruating or had sexual intercourse within the last 24 hours were excluded because blood and semen invalidate the measurement of vaginal pH and amines.

Sample size for a diagnostic study is based on the prevalence of a condition (a population proportion) and the desired levels of accuracy of the diagnostic test (Daniel, 2004). We wanted to achieve 95% positive predictive values (PPV), 95% negative predictive values (NPV), and precision of plus or minus 3%. The formula: $n = z^2pq/d^2$ was employed, where z represents the standard score of 1.96 at 0.05 level of significance, $p = 0.95$ (represents the NPV and PPV goals), q equals $1 - p$ (0.05), and d represents precision of the estimator ($d = 0.03$). The resulting sample sizes are then adjusted for the prevalence of BV/TV and CV with the formula $N = n/\text{prevalence}$. In the absence of prevalence data from military troop medical clinics, data from a previous clinical study of civilian women (Lowe & Ryan-Wenger, 2000) was used. Based on prevalence of 38% for BV/TV and 45% for CV, at least 534 participants with vaginal symptoms were required.

Institutional Review Board approvals were obtained from all agencies and universities involved in the study. The parent study enrolled 715 active duty Army and Navy women from four military sites who were experiencing vaginal or urinary symptoms. This report focuses on 546 women who presented with vaginal symptoms.

Procedure

APRNs for this study were certified women's health nurse practitioners credentialed by the military hospitals to provide care for women with genitourinary symptoms in the troop medical clinics. Women experiencing genitourinary symptoms were recruited for participation via posters in the clinic lavatories and flyers placed at the reception desks, and by civilian APRNs

who approached women in the clinic waiting areas. Care was taken not to give the impression of coercion. Women could choose to see the regular provider or the research APRN. After obtaining written consent and approval to use their health information, research APRNs taught the women how to use the WMSD Kit materials. In a clinic lavatory stall that resembles the size of a typical latrine in deployment settings, the women followed directions in the Kit for vaginal fluid testing. Women marked their signs, symptoms, and self-diagnoses on the WMSD Decision-Making Guide and sealed the Guide in an envelope. They then were seen by a research APRN who followed a clinical protocol that included a clinical interview, pelvic examination, measurement of vaginal pH and amines with a FemExam card, standard nitrazine paper and whiff test, and wet mount microscopy of vaginal fluids. Tests for chlamydia and gonorrhea were standard practice in these clinics whenever women were evaluated for genitourinary symptoms. Research APRNs made diagnoses based on their clinical judgment and treated the participants with appropriate single-dose oral medications and palliative products. Clinic laboratory staff conducted the DNA testing, and APRNs were blind to the DNA results. The women's reactions to the WMSD Kit were solicited at the end of their clinic visit and at a follow-up appointment 2 to 5 days later.

Instruments

The *WMSD Kit* contained 4 items for vaginal self-diagnosis: a sterile cotton swab, 2 FemExam® cards, the WMSD Decision-making Guide©, and a yellow marker. The Guide© instructed women to insert the swab (PurWraps®, Harwood Products, Guilford, ME) about 2 inches into the vagina to avoid touching the cervix and twirl the swab to completely moisten it. Studies show that women have successfully obtained their own vaginal swab specimens for clinical diagnosis and research on normal vaginal flora, BV, TV, CV, Chlamydia, gonorrhea and human papilloma virus (Ferris, et al., 2006; Hoebe, Radamaker, Brouwers, Ter Waarbeek, & Bergen, 2006; Lowe & Ryan-Wenger, 2000). The FemExam® (CooperSurgical, 1998) is a credit card-sized, qualitative colorimetric product with two small black-ringed yellow circles for measurement of pH and amines. The vaginal swab is rotated on the two circles and after two minutes, the results within the two circles are read as a blue-green PLUS sign (+) or MINUS sign (−) to indicate pH greater than or less than 4.7 and positive or negative amines. Performance studies showed that FemExam® pH compared to nitrazine paper yielded 85.1% sensitivity, 86.6% specificity, and 85.8% accuracy. Amine test results compared to the whiff test yielded 84.7% sensitivity, 91.4% specificity, and 89.3% accuracy. The FemExam® was designed for use by professionals, and is not available over-the-counter, but a reproducibility study by the company compared findings of 4 first time users (no training), 2 experienced, trained health care providers, and 3 trained users. Of the 333 pH and amine samples, 99% were correctly read regardless of training or experience (CooperSurgical, 1998). To insure that both the participant and the research APRN used FemExam® cards from the same manufacturing lot, each Kit contained 2 cards with the same lot number to eliminate the potential for manufacturing control differences between cards as a source of measurement error. Near the end of data collection for the study reported herein, FemExam® cards were sold as the QuickVue® Advance card by Quidel Corporation (San Diego, CA).

The *WMSD Decision-making Guide*©, developed by the investigators, is a 3.5" × 7" envelope of heavy card stock with strategically-placed "windows" and an inner card that slides up and down (Invention Disclosure #97ID35F, The Ohio State University). The text on the card, written at an 8th grade reading level, guides participants through the self-testing and decision-making process for self-diagnosis of vaginal symptoms on one side, and urinary symptoms on the other side (not addressed in this report). Specific combinations of signs and symptoms are revealed in a window near the top, the corresponding self-diagnoses in the center window, and treatment recommendations in the bottom window. The decision-making algorithms in the Guide© including signs, symptoms and corresponding diagnoses were based upon published

clinical guidelines and prior research by the investigators. For the purposes of this study, the treatment recommendation for all self-diagnoses was, “See your health care provider,” meaning the research APRN. The *Vaginal Symptoms* side of the Guide instructed participants on how to collect a vaginal fluid specimen with a swab and conduct a pH and amine test on the Fem Exam®. Any possible combination of + and – pH, + and – amines, and vaginal itching (yes or no) could be selected. Corresponding diagnoses included bacterial infection (BV and/or TV), yeast infection (CV), no infection, and unclear. For the purposes of data collection, women were instructed to mark their findings in each window with a yellow highlighter.

Laboratory Reference Standard—Clinically significant amounts of DNA of *G. vaginalis*, *T. vaginalis* and *Candida* species detected by the Affirm™ VPIII Microbial Identification Test (Becton-Dickinson, 1996) served as the laboratory reference standard for diagnosis of BV, TV and CV. Research APRNs and clinic laboratory staff were trained to use this device by company representatives. Vaginal swabs obtained during pelvic examinations were processed within four hours by mixing specimens with a lysis solution, incubating them at 85°C for ten minutes, and then mixing them with a buffer solution. The Affirm™ device dips specimens into wells of liquid reagents that cause hybridization of nucleic acid analytes to color development probes and capture probes. The device does not require calibration; test results are indicated on a 1×2 inch Affirm™ test card with 5 colorimetric beads: 2 for positive and negative control results, and separate DNA indicator beads for *T. vaginalis*, *G. vaginalis* and *Candida* species. The presence of DNA is confirmed if the indicator beads change from white to blue. Tests are valid if the negative control bead remains white and the positive control bead remains blue. Becton-Dickinson reported that compared to wet mount microscopy, Affirm™ sensitivities and specificities were 95–98% and 100% for *G. vaginalis*, 90–93% and 99.9% for *T. vaginalis*, and 82% and 98.4% for *Candida* species. Test results are not adversely affected by vaginal lubricants, douches, menses, or spermicides. Three different first time users (APRNs) achieved 100% agreement and accuracy on identification of 24 coded samples (Becton-Dickinson).

Analysis

Analyses of women’s self-diagnoses compared to laboratory reference standards were calculated from 2×2 contingency tables shown in Box 1. Formulae embedded in an Excel spreadsheet were used to calculate accuracy scores from contingency table frequencies (Brownson & Pettiti, 2006). Self-diagnosis accuracy for BV/TV and CV was defined as the proportion of women whose positive and negative self-diagnoses matched the laboratory reference standard. Positive predictive values (PPV) indicate the proportion of women who correctly self-diagnosed an infection, and negative predictive values (NPV) reflect the proportion of women who correctly said, “no,” they did not have an infection. The women’s *potential* for self-medication errors was determined by the proportion of false positives (commission errors) and the proportion of false negatives (omission errors). Kappa coefficients were used to measure inter-rater agreement of pH and amine results between the women and their APRNs.

Box 1**2x2 contingency table and formulae used to calculate accuracy of a diagnostic test**

		Diagnosis per laboratory result		
		+	-	
Self-diagnosis	+	TP	FP	TP + FP
	-	FN	TN	FN + TN
		TP + FN	FP + TN	TP + FP + FN + TN ¹

¹TP = true positive, FP = false positive, FN = false negative, TN = true negative

Accuracy = (TP + TN)/(TP + FP + FN + TN)

Positive Predictive Value (PPV) = TP/(TP + FP)

Negative Predictive Value (NPV) = TN/(FN + TN)

Potential Self-Medication Commission errors = FP/(TP + FP + FN + TN)

Potential Self-Medication Omission errors = FN/(TP + FP + FN + TN)

Results

The 546 participants' ages ranged from 18 to 50 years, with an average age of 25.7 years (SD = 5.8). Other demographic characteristics are listed in Table 1. Notably, 372 (65.6%) women had previous deployment experience. During their last deployment, the women's living quarters were tents (n = 86, 23.1%), ships (n = 195, 52.4%) and other structures (n = 90, 24.2%). More than one-fourth of the women reported that they had vaginal symptoms during their last deployment (n = 99, 26.6%).

The accuracy of military women's vaginal self-diagnoses compared to laboratory reference standards is summarized in Table 2. A total of 302 women self-diagnosed BV and/or TV and Affirm™ diagnosed 320 cases. However, only 193 of these cases matched; there were nearly as many false positives (n=109) as false negatives (n=127). The potential for self-medication commission and omission errors was 20.3% and 23.7% among these symptomatic women. Sixty-nine women self-diagnosed CV while Affirm™ indicated 148 cases, yet only 26 of them matched. The few false positives (n=43) minimized the potential for self-medication commission errors (8%), but the number of false negatives and potential for self-medication omission errors were three-fold higher.

An evaluation of the algorithms used to derive self-diagnoses for vaginal symptoms is shown in Table 3. Self-diagnosis of CV and "no vaginal infection" had the highest accuracies at slightly above 70%. The five algorithms for a BV/TV diagnosis yielded accuracies ranging from 34.4% to 46.3% (Table 3). With respect to mixed infections other than BV/TV, CV in combination with BV and/or TV was relatively common (n = 72, 13.4%), but only a few women

were positive for Chlamydia ($n = 3$) or gonorrhea ($n = 16$). Contamination of the vaginal specimen by chemicals could alter the pH and amine test results. The research APRN's clinical notes indicated that 7 women reported douching within the past 24 hours with vinegar and water, tea tree oil, or two commercial douches that contain povidone iodine, vinegar, and numerous other chemicals.

Of 539 women who responded to questions about their experience with the WMSD Kit after their examination by the research APRN, most women reported that the directions in the kit were clear and easy to follow ($n = 503$, 93.3%) and that the materials in the kit were very easy to use ($n = 454$, 84.2%). At the two-week follow-up visit with the research APRN, 477 women responded to the questionnaire. Ninety-nine percent ($n = 472$) of the women said that they would like to have this kit during deployment, and 421 (88.1%) said that they would definitely use it, 36 (7.5%) would probably use it, and 17 (3.6%) would definitely or probably seek health care instead.

When the women's results were compared to their research APRN's pH and amine findings using the FemExam® card, kappa coefficients were only 0.25 for amines and 0.38 for pH. Different manufacturing batches of cards did not explain the inter-rater discrepancies because the women had two cards with the same lot number in their Kits, one of which they gave to their research APRN. Post-hoc investigation was needed to determine if the women simply misread their cards, or if the cards themselves were faulty. The APRNs recorded results from both FemExam® and Amsel criteria. Post-hoc reliabilities were low between pH from FemExam® and pH from nitrazine paper ($\text{kappa} = 0.45$, $p < .001$), and between amines from FemExam® and amine odor on whiff tests ($\text{kappa} = 0.43$, $p < .001$). Therefore, we concluded that defects of the FemExam® cards were responsible for the low observed kappa coefficients.

Discussion

The WMSD diagnostic algorithms did not adequately diagnose BV, TV or CV, most likely due to clinical variations in the presentation of these vaginal infections, as well as recently identified flaws in the FemExam® card. The low observed accuracy rates could have originated from several sources, including participants' misunderstanding of the directions on how to use the WMSD Kit components, contamination of the vaginal specimens collected by the participants, inadequacy of the diagnostic algorithms for BV/TV and CV, or inaccuracy of the clinical assessment devices.

Low educational level is often a source of error in studies that rely on participants to interpret reading material. However all of the women in this study had the equivalent of at least a high school education because it is a military entry requirement, and many women had some college education because college courses are readily available on nearly all military bases. Verbal instructions by the research APRN and written instructions on how to collect a vaginal swab and how to use the FemExam® card and the WMSD Decision-Making Guide could have been sources of error, yet most women reported that the items in the Kit were easy to use and that the instructions were easy to follow. Consumers' ability to understand written instructions that accompany self-testing devices was evaluated in a multi-site study of 206 women. They were tested on their understanding of how to use an over-the-counter vaginal pH self-testing device after reading the package insert (Roy, Caillouette, Faden, & Roy, 2005). Correct answers were given by 91% to 99% of the women to 15 of 16 items on a test based on package insert information. The remaining item was correctly answered by 64% of the women ("Is more than 30% of vaginal itching due to bacterial infection?"). Roy, et al.'s study supports our belief that misunderstanding the kit directions was not a significant source of error in our participants' self-diagnoses.

Self-obtained vaginal specimens may have been contaminated with menstrual blood, semen or feminine hygiene products. When the research APRN's clinical examination revealed blood or semen, those women were removed from the data set. The number of women who douched with chemicals was too small to negatively influence the findings. The swab may have reached fluids from the cervix, which are more acidic than the vaginal wall, but this would be a rare occurrence. We conclude that contamination of the vaginal fluid specimen was not a significant source of error. It was possible that the amount of vaginal fluid on the swab may have been insufficient to adequately moisten the reagents on the FemExam® card, but we did not have data to test this hypothesis. Some women may have selected the wrong combination of pH, amine and itching results on the WMSD Decision-Making Guide, which would lead to an incorrect self-diagnosis. However, very few women reported that they had difficulty understanding how to use the Guide and other materials in the Kit.

Only one algorithm was considered to be diagnostic of CV, i.e. the combination of negative pH, negative amines, and itching. A problem is that many women with CV who do *not* experience itching would not be correctly diagnosed with CV because the algorithm of negative pH, negative amines, and *no* itching was reserved for the diagnosis of "normal" or no infection. This algorithm was not a major source of error however, since only 6 women with "normal" signs and symptoms had CV according to DNA. Co-infections may blur expected distinctions between BV, TV and CV (Romanik, Ekiel, Friedek, & Martirosian, 2005). Another potential source of error in self-diagnosis of BV is that the standard for BV was identification of *G. vaginalis* DNA by Affirm.™ A few other organisms are implicated in BV, but overgrowth of *G. vaginalis* is the most common. It is unlikely that a significant number of BV cases were missed due to other organisms.

Some signs and symptoms such as positive pH or vaginal itching are clinically non-specific and may be observed in BV, TV and/or CV (Anderson, Klink, & Cohrsen, 2004). In women of menopausal age, positive pH scores could be indicative of BV or of low serum estradiol (Caillouette, Sharp, Zimmerman, & Roy, 1997). In our sample, low serum estradiol was not a likely source of error since only 7 women were over age 45 and 2 women were age 50. Women with either a positive pH or a positive amine could have had "intermediate flora" according to the Nugent system, which would place them *between* "BV" and "no infection." Women would have self-diagnosed BV from the WMSD Guide even though Affirm™ would be negative for a clinically significant amount of *G. vaginalis*. Intermediate flora may indicate a transition from normal to BV such that clinical signs and symptoms are not interpretable (Forsum, et al., 2005). On the other hand, it may be that intermediate flora represents the opposite – a transition from BV to normal flora (Donders, 2007).

Because our desired criteria of PPV and NPV $\geq 95\%$ were not reached, we considered potential sources of the observed lower than desired accuracy rates. Conceptually, errors may have occurred due to participants' misunderstanding of the directions for use of the WMSD Kit components, contamination of the vaginal specimens collected by the participants, inadequacy of the diagnostic algorithms for BV/TV and CV, or inaccuracy of the clinical assessment devices.

Limitations

The FemExam® card was purported to yield highly reliable and valid pH and amine levels in vaginal fluid, but it proved to be the weak link in this study. Anecdotally, many women and two of the four research APRNs reported that the plus and minus signs on the FemExam® cards were not very distinctive and therefore difficult to interpret. As noted from the inter-rater agreement analysis, faulty pH and amine test cards were a major source of error in the women's self-diagnoses. Similar problems with the readability and accuracy of the FemExam® were reported by Gutman, Peipert, Weitzen, & Blume (2005). A related issue is that the Amsel

criteria for BV is vaginal pH > 4.5, but the FemExam® reads positive when pH is > 4.7, resulting in a potential for false negative self-diagnoses for BV. Near the end of our data collection, the FemExam® product was discontinued and later sold as QuickVue® by Quidel®. This device also was discontinued, and similar products are no longer available.

Power analysis required at least 534 subjects, therefore, our sample size of 546 women was adequate to detect accuracy, PPV and NPV rates of 95% and precision of 3% if such rates existed. Our results suggest that these rates would be unachievable with any sample size because even in the best of circumstances, vaginal pH, amines and itching are non-specific in single and mixed infections. The women were not randomly selected; rather, they self-selected to attend a troop medical clinic due to symptoms, and were invited to participate. Thus, generalizability of the results is limited to military women such as these.

Our results and other evidence of the relatively low diagnostic accuracy of experienced clinicians who may or may not follow practice guidelines for diagnostic decision-making (Allen-Davis, et al., 2002; Anderson & Karasz, 2005; Lowe et al., 2009; Schwiertz, et al., 2006; Weisenfeld & Marco, 1999) suggest that a goal of $\geq 95\%$ for self-diagnosis accuracy, PPV and NPV likely is unachievable even with accurate devices to measure pH and amines. Lower accuracy levels must be tolerated because of the non-specific nature of signs and symptoms of vaginitis. Our goal for the WMSD Kit is to minimize false positive results that will result in self-medication commission errors. False negative self-diagnoses that lead to self-medication errors of omission are less serious since women can retest the next day if symptoms persist. The military women in this study want a WMSD self-diagnosis and treatment kit issued to them when they are deployed. Further research is needed to achieve clinically acceptable levels of diagnostic accuracy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by a grant from the National Institutes of Health, National Institute of Nursing Research: # 1 R01 NR07662-01A1 (Nancy A. Ryan-Wenger, Principal Investigator).

References

- Allen-Davis JT, Beck A, Parker R, Ellis JL, Polley D. Assessment of vulvovaginal complaints: Accuracy of telephone triage and in-office diagnosis. *Obstetrics and Gynecology* 2002;99:18–22. [PubMed: 11777504]
- Amsel R, Totten PA, Spiegel CA, Chen KC, Eschenbach D, Holmes KK. Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. *American Journal of Medicine* 1983;74(1):14–22. [PubMed: 6600371]
- Anderson, MR.; Karasz, A. How do clinicians manage vaginal complaints? An internet survey. *Medscape General Medicine*. 2005. Retrieved on July 5, 2008 from www.medscape.com/viewarticle/505432
- Anderson MR, Klink K, Cohn A. Review: Vaginal signs and symptoms perform poorly in diagnosing vaginal candidiasis, bacterial vaginosis, and vaginal trichomoniasis. *JAMA* 2004;291:1368–1379. [PubMed: 15026404]
- Becton-Dickinson. Affirm VPIII microbial identification test. Mississauga; Ontario: 1996.
- Brown D Jr. Clinical variability of bacterial vaginosis and trichomoniasis. *Journal of Reproductive Medicine* 2004;49:781–786. [PubMed: 15568400]
- Brownson, RC.; Pettiti, DB. *Applied epidemiology: Theory to practice*. 2. Oxford, UK: Oxford University Press; 2006.

- Caillouette JC, Sharp CF, Zimmerman GJ, Roy S. Vaginal pH as a marker for bacterial pathogens and menopausal status. *American Journal of Obstetrics and Gynecology* 1997;176:1270–1277. [PubMed: 9215184]
- Centers for Disease Control and Prevention. Sexually transmitted diseases treatment guidelines - 2006. Atlanta, GA: 2006.
- CooperSurgical. FemExam pH and Amines TestCard. Shelton, CT: CooperSurgical; 1998.
- Daniel, WW. Biostatistics: a foundation for analysis in the health sciences. 6. Hoboken, NJ: Wiley & Sons; 2004.
- Defense Department Advisory Committee on Women in the Services. 2007 Report. 2007. Retrieved June 17, 2008, from http://www.dtic.mil/dacowits/annual_reports/DACOWITS07_Report.pdf
- Donders GGG. Definition and classification of abnormal vaginal flora. *Best Practice and Research in Clinical Obstetrics and Gynaecology* 2007;21(3):355–373.
- Ferris DG, Francis SL, Dickman ED, Miler-Miles K, Waller JL, McClendon N. Variability of vaginal pH determination by patients and clinicians. *Journal of the American Board of Family Medicine* 2006;19(4):368–373. [PubMed: 16809651]
- Forsum R, Hallén A, Larsson PG. Bacterial vaginosis - a laboratory and clinical diagnostics enigma. *APMIS* 2005;13:153–161. [PubMed: 15799757]
- Gutman RE, Peipert JF, Weitzen S, Blume J. Evaluation of clinical methods for diagnosing bacterial vaginosis. *Obstetrics and Gynecology* 2005;105:551–556. [PubMed: 15738023]
- Hoebe CJPA, Rademaker CW, Brouwers EEHG, Ter Waarbeek HLG, Bergen JEAM. Acceptability of self-taken vaginal swabs and first-catch urine samples for the diagnosis of urogenital chlamydia trachomatis and Neisseria gonorrhoeae with an amplified DNA assay in young women attending a public health sexually transmitted disease clinic. *Sexually Transmitted Diseases* 2006;33(8):491–495. [PubMed: 16547452]
- Jennings BM. Soldiers' experiences with military health care. *Military Medicine* 2005;170:999–1004. [PubMed: 16491935]
- Lowe NK, Neal JL, Ryan-Wenger NA. Accuracy of the clinical diagnosis of vaginitis compared with a DNA probe laboratory standard. *Obstetrics & Gynecology* 2009;113(1):89–95. [PubMed: 19104364]
- Lowe NK, Ryan-Wenger NA. A clinical test of women's self-diagnosis of genitourinary infections. *Clinical Nursing Research* 2000;9(2):144–160. [PubMed: 12162239]
- Nugent RP, Krohn MA, Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *Journal of Clinical Microbiology* 1991;29(2):297–301. [PubMed: 1706728]
- Orem, D. Nursing Concepts of Practice. 6. St. Louis: Mosby; 2001.
- Posner KSF, Kerimova J, Aliyeva F, Duerr A. Strategies for diagnosis of bacterial vaginosis in a resource-poor setting. *International Journal of STD and AIDS* 2005;16:52–55. [PubMed: 15705274]
- Romanik M, Ekiel A, Friedek D, Martirosian G. Can chlamydial cervicitis influence diagnosis of bacterial vaginosis? *Journal of Clinical Microbiology* 2005;43(9):4914–4915. [PubMed: 16145176]
- Roy S, Caillouette JC, Faden JS, Roy T. The role of an over-the-counter vaginal pH self-test device package insert: Can subjects learn what the device is for and how to use it? *American Journal of Obstetrics and Gynecology* 2005;192:1963–1969. [PubMed: 15970861]
- Ryan-Wenger NA, Lowe NK. Military women's perspectives on health care during deployment. *Women's Health Issues* 2000;10:333–343.
- Schwartz A, Taras D, Rusch K, Rusch V. Throwing the dice for the diagnosis of vaginal complaints? *Annals of Clinical Microbiology and Antimicrobials* 2006;5:4–10. [PubMed: 16503990]
- Sobel JD. What's new in bacterial vaginosis and trichomoniasis? *Infectious Disease Clinics of North America* 2005;19:387–406. [PubMed: 15963878]
- Soper D. Trichomoniasis: Under control or undercontrolled? *American Journal of Obstetrics and Gynecology* 2004;190:281–290. [PubMed: 14749674]
- Thinkhamrop J, Lumbianon P, Thongdrjai P, Chongsomchai C, Pakarasang M. Vaginal fluid pH as a screening test for vaginitis. *International Journal of Gynecology and Obstetrics* 1999;66:143–148. [PubMed: 10468337]

- White DJ, Vanthuyne A. Vulvovaginal candidiasis. *Sexually Transmitted Infections* 2006;82(Suppl IV):iv28–iv30. [PubMed: 17151050]
- Wiesenfeld HC, Macio I. The infrequent use of office-based diagnostic tests for vaginitis. *American Journal of Obstetrics and Gynecology* 1999;18:39–41. [PubMed: 10411793]
- Women in Military Service for America. Statistics on women in the military. 2006. Retrieved on January 13, 2009 from <http://www.womensmemorial.org/Press/stats.html>

Table 1Demographic characteristics of the participants (N = 546^a)

Variables	n	%
Ethnicity		
Hispanic or Latino	97	17.9
Race		
American Indian or Alaska Native	7	1.3
Native Hawaiian or Pacific Islander	5	0.9
Asian	11	2.0
Black or African American	234	43.0
White	173	31.9
More than one race or other	53	10.1
Not reported	30	5.5
Education		
High school graduate or GED	125	23.1
Some college	313	57.7
Associate degree	40	7.4
College graduate	47	8.6
Post graduate	17	3.1
Marital status		
Married or living with significant other	217	40.0
Never married and not living with significant other	216	39.8
Divorced or separated	104	19.1
Widowed	2	0.4
Site		
Naval Station San Diego	307	56.2
Fort Hood	109	20.0
Fort Carson	97	17.8
Fort Bliss	33	6.0
Rank		
Enlisted (E1–E3)	145	26.7
Non-Commissioned Officer (E4–E6)	358	65.9
Junior Officer (O1–O3)	37	6.8
Senior Officer (O4–O6)	3	0.6

^a Some categories do not add to 546 due to missing data.

Table 2

Accuracy of women's vaginal self-diagnoses compared to laboratory reference standard for bacterial vaginosis (BV), trichomonas vaginitis (TV), and candida vaginitis (CV).

	Affirm™ DNA for BV and/or TV				Accuracy rate	Positive Predictive Value	Medication commission error rate	Medication omission error rate
		YES	NO	n				
Women's self- diagnosis of "bacterial infection"	YES	193 True +	109 False +	302	56%	63.9%	20.3%	23.7%
	NO	127 False –	107 True –	355		Negative Predictive Value		
		320	216	536		30.1%		
		Affirm™ DNA for CV				Positive Predictive Value	Medication commission error rate	Medication omission error rate
Women's self- diagnosis of "yeast infection"		YES	NO	n	Accuracy rate			
	YES	26 True +	43 False +	69	69.2%	37.7%	8%	24.8%
	NO	122 False–	345 True–	467		Negative Predictive Value		
		148	388	536		64.4%		

Table 3

WMSD¹ Kit vaginal self-diagnosis algorithms compared to expected diagnoses validated by Affirm VPIIITM DNA Probe Analysis

Algorithms			n	WMSD Self-Diagnosis	Expected Diagnosis	Accuracy rate	Positive Predictive Value	Negative Predictive Value
pH	amines	itching						
+	+	yes	32	Bacterial infection	BV and/or TV	44.0	78.1	41.8
+	+	no	50	Bacterial infection	BV and/or TV	46.3	78.9	42.7
+	-	yes	162	Bacterial infection	BV and/or TV	40.4	0	40.4
+	-	no	92	Unclear	BV, TV and/or CV	34.4	72.8	26.1
-	-	yes	73	Yeast infection	CV	70.3	41.7	74.9
-	-	no	51	No infection	None	73.6	49.0	76.2
-	+	yes	37	Bacterial infection	BV and/or TV	41.5	56.8	40.3
-	+	no	20	Bacterial infection	BV and/or TV	41.3	60.0	40.6

¹ WMSD = Women in the Military Self-Diagnosis