

Randomized Field Trial of Vaginal Douching, Pelvic Inflammatory Disease and Pregnancy

Kenneth J. Rothman, Donnie P. Funch, Tanya Alfredson,
Joanne Brady, and Nancy A. Dreyer

Background. Several case-control studies have reported that women who use vaginal douche products are at increased risk for pelvic inflammatory disease. Women who douche regularly may do so for reasons related to their risk of acquiring a sexually transmitted infection, introducing confounding that is difficult to control in non-experimental studies.

Methods. We conducted a multicenter randomized field trial with a 1-year follow-up period. The study comprised 1827 women age 18–34, with no current indication of pelvic inflammatory disease, who were regular users of a douche product and who had been treated recently for a sexually transmitted bacterial infection or bacterial vaginosis. Women were randomly assigned to use either a newly designed and marketed douche product or a soft cloth tow-elette, and were resupplied with product at each bimonthly follow-up visit. We measured the occurrence of pelvic in-

flammatory disease using a combination of clinical and laboratory indicators. We also recorded pregnancy occurrence among participants.

Results. The risk of PID among women assigned to use the douche product, relative to that among women assigned to use the wipe product, was 1.05 (95% confidence interval = 0.57–1.9). Using an alternative, less sensitive definition of PID gave a risk ratio of 1.26 (0.62–2.6). The probability of becoming pregnant was 15% lower among women assigned to use a douche product, and 33% lower among women who douched more frequently (ratio = 0.67; 0.42–1.08).

Conclusions. There was little or no indication of a greater risk of PID among women assigned to use the douche product. Douching may be related to a lower probability that a woman becomes pregnant.

(EPIDEMIOLOGY 2003;14:340–348)

Key words: pelvic inflammatory disease, douching, sexually transmitted diseases, epidemiology, pregnancy.

Several epidemiologic studies,^{1,2,3,4} but not all,⁵ have reported a positive association between vaginal douching and the risk of pelvic inflammatory disease (PID). In a meta-analysis, Zhang and colleagues⁶ estimated that douching was related to an increase in risk for pelvic inflammatory disease of 73%. All reported studies were case-control studies, in which other risk factors for PID, including those related to sexual behav-

ior and sexually transmitted infections (STIs), were controlled using information obtained by interview.

In most cases, PID is thought to result from a complication of an STI, especially gonorrhea and chlamydia.⁷ In turn, the major risk factors for STI, such as number of sex partners, a change in sex partners, age at first intercourse, frequency of intercourse and type of contraception, are strong risk factors for PID.⁸ Vaginal douching is a plausible risk factor because it is possible that douching acts to flush infectious organisms up the genital tract into the uterus and fallopian tubes, or that the douching solution alters vaginal flora and promotes the possibility of an ascending infection. Alternatively, it is possible that the association between douching and PID is secondary to confounding by sexual activity, because douching is used for personal hygiene, often in relation to sexual intercourse.⁹ We conducted a multicenter randomized field trial to evaluate the unconfounded effect of douching on risk of PID. While the trial was underway, Baird *et al.*¹⁰ reported that douching was related to a reduced probability of pregnancy,

From Ingenix Epidemiology and Safety Research (formerly Epidemiology Resources, Inc.), Auburndale, MA.

Address correspondence to: Kenneth J. Rothman, Department of Epidemiology, Boston University School of Public Health, 715 Albany Street, Boston, MA 02118; krothman@aol.com.

This study was supported by the Consumer Products Division of SmithKline Beecham, Inc. (now Glaxo-SmithKline).

The authors were contractually granted scientific control of the study, including the study design, data analysis and interpretation, and final wording of any resulting manuscript.

Submitted 3 July 2002; final version accepted 17 January 2003.

Copyright © 2003 by Lippincott Williams & Wilkins, Inc.

TABLE 1. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
<ol style="list-style-type: none"> 1. Women age 18–34 years in good overall health (based on the investigator's judgment) 2. History of douching an average of at least once a month during the past 6 months 3. History of a gynecologic infection (not including genital warts or herpes) that was treated with an antibiotic from 10 days to 6 months before enrollment (based on self-report) 4. Negative urine pregnancy test at the screening visit 5. Intent to avoid pregnancy during the study, evinced by the current use of any contraceptive method other than abstinence or withdrawal 6. Competence in written and spoken English or Spanish (at select sites), willingness to comply with the study procedures and willingness to sign the study consent form 	<ol style="list-style-type: none"> 1. Current PID diagnosis or treatment, or self-reported history of PID within the previous 6 months 2. History of hysterectomy, bilateral tubal occlusion or bilateral tubal surgery (including ligation) 3. Known or suspected sensitivity to any of the study products 4. Pregnant or less than 2 months postpartum 5. History of pelvic or abdominal surgery or termination of pregnancy within the previous 45 days 6. History of ectopic pregnancy, endometriosis, cervical or uterine cancer, or any unusual gynecologic abnormality apart from PID 7. Symptomatic infection with human immunodeficiency virus 8. Current drug or alcohol abuse 9. Participation in any other clinical study during the past 30 days or prior enrollment in this study 10. Presence of any condition that would, in the opinion of the investigator, preclude compliance with study procedures

so we added pregnancy as a secondary outcome to the trial.

Methods

We compared the risk of PID among participating women who were assigned either a vaginal douche product or soft cloth towelettes (wipes). Ethical constraints dictated that all women recruited for the study should be women who were regularly using vaginal douches before enrollment into the study. Women who enrolled in the study agreed to be randomly assigned to one of the two product-use groups and to return for bimonthly visits for a 1-year period. The douche product was different from previously marketed douche products in two ways: the applicator nozzle was redesigned with larger holes and deeper channels along the sides, to promote flow of the douching solution out of the vagina rather than up the genital tract, and the douche solution was formulated with a citric acid/sodium citrate buffer to a pH of about 4.2. The old product was discontinued and replaced in stores by the new product before the trial began. The towelettes are moistened cloths in individually sealed packages and are intended for external vaginal use.

Recruitment of Sites and Participants

International Pharmaceutical Research (IPR), a company contracted for site recruitment and management, found 59 sites in 24 states that proceeded through IRB approval and enrollment of at least one participant between November, 1995 and July, 1999. Sites included clinics providing primary or specialized care and research sites. Sites that failed to recruit at least 10 participants were administratively closed after a trial period. Sites that had less than 60% retention of participants during the year of follow-up were notified, and, in some cases, later closed to additional recruitment if they failed to improve their retention rates. Women were paid about

\$25 for each visit and received free study product during the study.

Inclusion and Exclusion Criteria

The study was designed to estimate the effect of douching among women who were at high risk of PID. Therefore, to be eligible a woman had to have been treated with an antibiotic for a sexually transmitted bacterial gynecologic infection during the previous 6 months. If a woman presented with an active infection, she was first treated for the infection and considered for enrollment after the treatment was completed. Before enrollment, all women were administered a pregnancy test (which had to be negative) and were interviewed using a structured questionnaire to determine study eligibility. Table 1 lists the inclusion and exclusion criteria.

Modifications to Eligibility Criteria

As the study progressed, recruitment difficulties prompted modification of the eligibility criteria. Initially, the upper age limit was 29 years, but the range was expanded to 34 mid-way through recruitment. Also, the type of qualifying infection was expanded to include bacterial vaginosis.

Product Assignment and Follow-Up

Women who met the study criteria were assigned to receive either the douche or wipe product. The assignment was based on a pseudo-random algorithm that was programmed into hand-held graphic calculators and then distributed to all study sites. The algorithm requested seed information for each participant that included her birth date and part of her social security number and telephone number. The calculator then displayed the treatment assignment, using a calculation that was kept secret during the study. In reality the telephone number and social security number were decoys. The assignment was made by converting the birth

date into a Julian day number and assigning the participant according to the odd-even parity of the date number. This system enabled us to check the validity of the random assignment for every participant, because the assignment always corresponded to the parity of the birth date. It would have been difficult to subvert the group assignment without falsifying the birth date, and then only with knowledge about the assignment calculation.

Each participant was given a 2-month supply of the product (up to 16 bottles of douche or 48 wipes, depending on need). Those who were assigned to the wipe group were instructed to refrain from using any douche product during the study. Those assigned to the douche group were instructed to use only the study douche product, and to continue their usual frequency and pattern of use. Participants returned for bi-monthly visits to receive additional product.

Data Collection

At the initial visit, eligible women who gave informed consent received a general physical examination and completed a questionnaire about general demographics and risk factors for sexually transmitted diseases, as well as information about douching practices. After random assignment, women were given a 3-month diary to record product use and the timing of menses, and a sheet to record any prescription drugs that they took. Women were seen at 2-month intervals, when completed documents were collected and new ones dispensed. At these visits, women were interviewed regarding their sexual activity, changes in the number of sex partners, sexually transmitted infections, other concurrent illnesses, symptoms that might correspond to PID in the interval since the last visit, and their use of any feminine hygiene product other than their assigned study product.

Assessment of PID and Pregnancy

The study was designed so that women would, as far as was possible, receive their usual medical care. A woman who became ill and appeared at the site between visits was asked the same series of questions that she would normally be asked during a standard visit. She received an examination if one was indicated, and if the site was a clinic she received treatment for any condition that required it. Unless this visit was within a few weeks of the next scheduled visit, she was requested to return for the regularly appointed visit and this visit was noted as an interim visit in her record. Each participant was instructed that if she went elsewhere for treatment, she should inform the caregivers that she was participating in a trial and show them her "Dear Doctor" card. The card explained the procedures to be followed if the woman had symptoms consistent with PID. A copy of

TABLE 2. Clinical and Laboratory Indicators of Pelvic Inflammatory Disease

I. Abdominal/pelvic tenderness
A. Lower abdominal tenderness
B. Cervical motion tenderness
C. Adnexal tenderness (may be unilateral)
D. Uterine tenderness
II. Other clinical symptoms and signs
A. Temperature $\geq 38^{\circ}\text{C}$ (100.4°F)
B. Mucopurulent vaginal discharge
C. Mass consistent with inflammatory process present on bimanual pelvic examination
III. Laboratory tests
A. Erythrocyte sedimentation rate >15 mm/hr
B. Evidence of the presence of <i>Neisseria gonorrhoea</i> or <i>Chlamydia trachomatis</i> in the endocervix. Gram stain reveals Gram-negative intracellular diplococci. Monoclonal antibody for <i>Chlamydia T</i>
C. Presence of >5 white blood cells per oil-immersion field in Gram stain of endocervical discharge or white blood count $\geq 10,300/\text{mm}^3$
D. Purulent material obtained by culdocentesis
E. Laparoscopy with evidence of tubal edema, erythema and purulent exudate and possible adhesions
F. Mass consistent with inflammatory etiology detected by sonography

the PID assessment form was printed on the card and could be completed and returned to the study site.

Women who had questions about their health or were experiencing problems were instructed to come for an interim visit. In addition, study participants were questioned about possible PID symptoms at each regular visit. Positive responses to any two questions about PID symptoms resulted in referral to a physician for a more thorough diagnostic assessment. We took the time of occurrence of PID to be the date of diagnosis. Although there was no way to keep participants or site coordinators blinded to the product assignment, information about the product assignment of study participants was not revealed to the site physician.

We classified the clinical and laboratory indicators for PID into three groups (Table 2). The site physician noted all relevant indicators on the PID assessment form. For most analyses we considered a woman to have PID if she presented with two or more indicators from group I and at least one indicator from group II or group III. This case definition is adapted from the Centers for Disease Control case definition for PID.¹¹

All pregnancies were ascertained from self-reports, either in person at a regular clinic visit or by telephone. If a woman reported becoming pregnant or having had a miscarriage before the final study visit, she was discontinued from the study as of the reported date of conception, or, lacking that information, as of 30 days after the last clinic visit. If a participant reported no menses for a 2-month period, she was questioned to rule out pregnancy.

Exclusion of Philadelphia Sites

Several problems arose in connection with the 10 sites in the Philadelphia area, prompting us to exclude

them from analysis. This decision was made after data collection was completed but before any assignment codes were revealed and before any data were analyzed.

The first problem, discovered by chance, was that 43 women enrolled in the study at more than one Philadelphia site. Upon this discovery, steps were introduced to prevent further double enrollment. Outside of Philadelphia, sites were sufficiently distant from one another so that double enrollment was unlikely. The second problem came to light after we had searched Pennsylvania vital statistics data to find how many study participants were lost to follow-up because they became pregnant. All women from the Pennsylvania sites were included in the search. From the Pennsylvania birth files we learned that 31 women who enrolled at Philadelphia sites appeared to have delivered a baby while they were actively enrolled in the study. We discovered this problem after follow-up had ended for these sites and the sites had been closed. We were never able to determine how these pregnancies went undetected or unreported by the site coordinators. This problem in Philadelphia led us to conduct similar vital records searches in two other states (Texas and Louisiana) where such searches were feasible, but there was no indication that the same problem existed outside of Philadelphia. The third problem with the Philadelphia sites was a near-complete absence of PID among the women in the study. Among women from other sites in the study, there were 41 cases of PID among 1885 women, whereas among the Philadelphia sites combined there was only one case of PID among 1518 women.

Given these anomalies in the data from Philadelphia, we decided to exclude these sites from the study. This exclusion involved little loss of information: if the Philadelphia-area sites had been included in the study, their inclusion would not have materially affected any of the comparisons that we report between the douche and wipe groups, because only one case occurred among these participants.

Data Analysis

The primary analyses concerned an effect of douching on the risk of PID. We conducted both intent-to-treat analyses and on-protocol analyses. In our intent-to-treat analyses, we compared the PID experience of all participants assigned to the douche and wipe groups regardless of their compliance with the study protocol. In the on-protocol analyses, we restricted the follow-up information for each participant to periods of follow-up that corresponded to periods of compliance with the study protocol. For these analyses, we discontinued from follow-up all women in the wipe group who were noncompliant because they reported using a douche product during their follow-up; no women in the douche group reported using a douche product other than the study

product. Noncompliant women were discontinued from follow-up 30 days after the visit date preceding the report of noncompliance. In the on-protocol analysis, we also discontinued the follow-up of any woman who had a gap of more than 150 days between consecutive clinic visits. The time of discontinuation was 30 days after the clinic visit preceding the gap. Most women who failed to complete the study were discontinued 30 days beyond their last actual visit, 30 days being half of the average length between scheduled visits. Women who had a definitive outcome marking the end of their participation (eg, pregnancy) were discontinued as of the date of the outcome or the date of the visit when the outcome was reported, if no earlier date was available.

We used both stratified analysis and proportional hazards regression to evaluate and control for potential confounding, and a life table analysis to generate cumulative incidence curves. To assess the influence of the diagnostic criteria for PID, we conducted some analyses using a less sensitive definition of PID, which required at least one positive laboratory test in addition to tenderness at two or more sites.

Results

A total of 3403 women were enrolled into the study. Of these, 1518 women (45%) were enrolled at Philadelphia area sites and 1885 (55%) at other locations. As we excluded the Philadelphia sites from the study before undertaking the analysis, our analysis is restricted to the 1885 women from sites outside of Philadelphia (Figure 1). Among the 1885 women enrolled, eight gave either insufficient or unreliable information and were not randomized to one of the study groups; the remaining 1877 women were randomly assigned, 895 to the douche group and 982 to the wipe group. We discovered that 45 of these women (22 in the douche group and 23 in the wipe group) were enrolled in error, because the site had inappropriately relaxed one or more of the entry criteria, and so we excluded these women from all analyses. We also excluded 5 women (3 in the douche group and 2 in the wipe group) who were dispensed the wrong product. None of the women who were dropped from the analyses experienced PID. There remained 870 women in the douche group and 957 in the wipe group.

Eighteen women in the wipe group indicated they had used a douche product at some point during the follow-up. These women were included in the intent-to-treat analyses; for the on-protocol analyses, their follow-up was truncated 30 days after their most recent clinic visit preceding the protocol violation. An additional 13 women from the douche group had their time truncated in the same manner for the on-protocol analyses because of a gap of over 2 months with no douche use. We did not truncate follow-up for the 24 women in

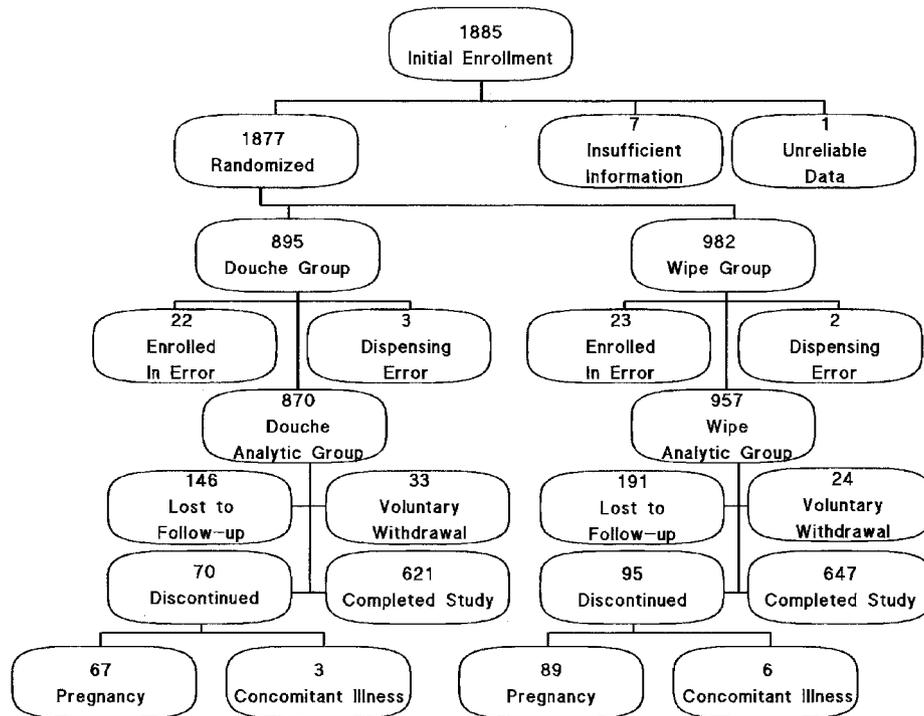


FIGURE 1. Participant flow diagram (excluding Philadelphia sites; see text).

the wipe group who had a gap of over 2 months in their use of wipes.

We assessed the reliability of the entry questionnaire information by comparing the entry information on the frequency of douching with the douching practices recorded during the study. We found that women who at the entry interview reported douching 1 to 3 times per month had a median frequency of douching during the follow-up of 2.2 times per month, based on their diaries. Women who reported douching more than 3 times per month at entry during the study had a median frequency of douching of 4.2 times per month during the follow-up.

The distribution of study participants by product group and by various demographic factors and selected risk factors for PID is given in Table 3. Overall, the two product assignment groups had similar distributions for all the variables that we examined, including those not shown in Table 3. The mean time between the occurrence of the infection that qualified a woman for entry into the study and the start of the study was 71 days for women assigned to the wipe group and 76 days for women assigned to the douche group.

About 19% of women were lost to follow-up during the 12-month study period. These women are included in the denominators of the intent-to-treat analyses and in the on-protocol analyses until the time that they were lost. Losses to follow-up did not alter the balance in PID risk factors between the study groups (data not shown).

Pelvic Inflammatory Disease

We first compared women assigned to the two product groups with respect to their PID outcome. There were 41 cases of PID among the women followed, 20 of which occurred in the douche group and 21 in the wipe group. The overall crude risk of PID among women in the douche group was 2.3%, and for women in the wipe group it was 2.2%, resulting in a crude risk ratio of 1.05 with a 95% confidence interval (CI) of 0.57–1.9. Table 4 gives a summary life table from which cumulative risks can be calculated. The latter are graphed in Figure 2. Consistent with the similarity of the crude risks, the two curves are roughly similar and cross one another at several points. We found no evidence of important confounding, but nevertheless we fit a proportional hazards regression with terms for the product assignment and several potential confounders measured at enrollment (Table 5). From this model we estimated a relative risk (RR) of 1.08 for douching, with a CI of 0.58–2.0.

We repeated the analyses after restricting attention to the experience of women who, based on their bimonthly interviews and diaries, kept to the intended protocol. These analyses gave nearly identical results to the intent-to-treat analyses, with a RR of 1.01 and a CI of 0.55–1.9.

To determine whether there was an effect of douching that was confined to those women who douched frequently, we compared the rate of PID for women who

TABLE 3. Distribution of Participants in the Douche and Wipe Groups for Selected Variables Ascertained from the Entry Questionnaire

Variable	Douche (N = 870)		Wipe (N = 957)	
	No.	%	No.	%
Age at entry				
18–19 Years	122	14	151	16
20–24 Years	355	41	403	42
25–29 Years	253	29	255	27
30–34 Years	140	16	148	15
Marital status				
Single	705	81	784	82
Married or living as married	165	19	172	18
Race				
Black	623	72	706	74
White	187	21	174	18
Other	60	7	76	8
Education				
High school or less	478	55	520	55
College or more	386	45	430	45
Annual income				
<\$15,000	630	74	693	74
\$15,000–\$29,999	181	21	199	21
\$30,000+	35	4	41	4
Qualifying infection				
Gonococcus/chlamydia	251	29	282	29
Other	619	71	675	71
Number of sex partners in 2 months before entry				
0	26	3	21	2
1	720	83	830	87
2+	124	14	104	11
Age at first intercourse				
8–15 Years	281	32	315	33
16–19 Years	541	62	582	61
20–29 Years	44	5	58	6
New sex partner in 2 months before entry				
No	786	90	872	91
Yes	84	10	84	9
Frequency of douching				
1–3/Month	789	91	876	92
4+/Month	81	9	79	8
Age of first douche				
9–16 Years	380	44	406	43
17–19 Years	346	40	370	39
20–33 Years	138	16	175	18
Barrier contraceptive use in 2 months before entry				
No	217	25	245	26
Yes	653	75	711	74

douched more than the median frequency (2.38 times/month), chosen as an arbitrary cutpoint, with that of women who douched less than or equal to the median frequency (Table 6). We obtained the frequency from the diary submitted at each 2-month visit and examined the rate of PID in the next 2-month interval in relation to that reported frequency. We found a large difference in the rates for more frequent and less frequent users, but the direction of the difference did not support the theory that douching increased the risk for PID. The rate was 5.5 cases per 100 woman-years among women who reported less frequent douching, which is 3.7 times greater than the rate of 1.5 per 100 woman-years among women who reported more frequent douching.

We repeated the main analyses for PID after restricting the case group to those cases who met the less

sensitive definition for PID. Of the 41 PID cases included in the above analyses, 30 had at least one positive laboratory test (16 in the douche group and 14 in the wipe group), giving a risk ratio of 1.26 (CI = 0.62–2.6). This effect estimate is slightly greater than the estimate based on PID cases meeting the primary case definition, although still substantially lower than the meta-analysis of case-control studies. Using the less sensitive definition of PID, we found that women who douched less frequently were at 6.8 times the risk of PID during the subsequent 2-month interval than women who douched more frequently (data not shown), a relation inverse to what would be expected if douching caused PID. We also analyzed the data excluding the four cases of PID that were not diagnosed at the study site; there was little difference in the findings after these exclusions.

Pregnancy

There were 161 women who became pregnant during the study follow-up period. Of these, 156 women were discontinued from follow-up because they were pregnant (Figure 1); the remaining five pregnant women were discovered to be pregnant at their study termination visit. The crude proportion becoming pregnant in the douche group was 8.0%, compared with 9.5% in the wipe group. The ratio of these proportions was 0.85 (CI = 0.63–1.14). To explore the possibility that the small difference in the probability of pregnancy between the groups was attributable to the use of the douche product, we examined the probability of pregnancy among women who were more frequent and less frequent users of the assigned study product (Table 7). The only subjects among the four groups that had a substantially different probability of pregnancy were women assigned to the douche group who used the product more frequently than the median amount of use; for these women, the probability of pregnancy was one-third lower than the probability of pregnancy among women who used the wipe product. We also estimated these probabilities in a proportional hazards analysis that included terms for age, frequency of sexual intercourse and race along with product assignment, and obtained similar results.

In Figure 3 we show the cumulative probability of becoming pregnant for women in the wipe group and for two subgroups of women in the douche group, according to the frequency of their use of the douche product. The women who used the douche product more frequently have a cumulative probability of pregnancy that does not cross the curves for the wipe group or the group that used the douche product less than the median frequency of use. Women who were using their douche product less frequently than the median frequency had nearly the same probability of pregnancy as women assigned to the wipe group, whereas the gap between the curve for

TABLE 4. Life Table for PID Occurrence Among Women in the Douche and Wipe Groups

Time (Days)	Douche				Wipe			
	At Risk	PID	Withdrawn	Cumulative Risk	At Risk	PID	Withdrawn	Cumulative Risk
0-30	870	2	1	0.0023	957	1	0	0.0010
31-60	867	1	71	0.0035	956	3	103	0.0044
61-90	795	3	32	0.0073	850	1	45	0.0056
91-120	760	2	16	0.0100	804	0	15	0.0056
121-150	742	4	26	0.0154	789	4	35	0.0107
151-180	712	0	13	0.0154	750	2	16	0.0134
181-210	699	0	35	0.0154	732	3	26	0.0175
211-240	664	2	12	0.0184	703	1	9	0.0189
241-270	650	2	20	0.0215	693	0	32	0.0189
271-300	628	0	6	0.0215	661	0	16	0.0189
301-330	622	1	16	0.0231	645	2	22	0.0220
331-360	605	2	109	0.0266	621	1	103	0.0237
361-390	494	1	458	0.0303	517	3	471	0.0341

women who were more frequent douche users and the other two curves widens with continued follow-up.

Discussion

The study was planned with the expectation of 100 or more cases of PID among the participants; less than half of this number was observed. Part of the shortfall in cases relates to the unexplained absence of cases in the Philadelphia sites, which was one reason that we excluded these sites before undertaking the analysis. The shortfall is also partly attributable to a secular decline in PID incidence during the decade in which the study was planned and carried out.¹² The smaller number of cases than expected resulted in wider confidence intervals than we had hoped to achieve in assessing the effect of vaginal douching.

Among the 1877 women in the analysis, the random assignment produced a split of 895 (47.7%) assigned to the douche group and 982 (52.3%) to the wipe group.

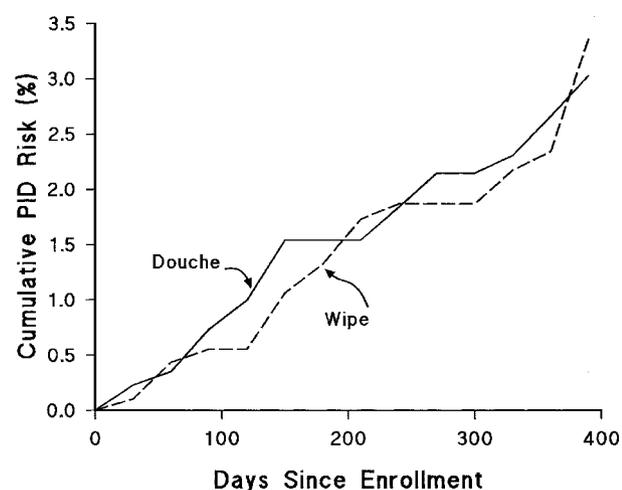


FIGURE 2. Cumulative risk of PID by product assignment: life table estimation of the cumulative risk of PID for the douche and the wipe groups.

Given this number of women, a deviation from an even split of this much or more should occur about 4.5% of the time, a value that is a bit unusual, but not strikingly so. The split would have been 1,691 vs 1,704 (49.8% vs 50.2%) had the Philadelphia sites been retained in the study.

The crude data show little difference between douche and wipe groups with regard to the risk of PID. In Figure 2 there is a period of about 100–150 days after randomization during which the risk for the douche group appears to be greater than for the wipe group. It is conceivable that there is a period of increased risk among women who douche that lags the qualifying gynecologic infection by several months. We consider this possibility remote, however, for the following reasons: (1) the qualifying infection could have occurred up to 6 months before enrollment, and therefore is unlikely to lead to such a focused difference between the study groups during an interval that begins about 100 days after randomization and lasts about 2 months; (2) few events determine the difference between the curves during that interval; (3) the gap between the curves does not persist; and (4) previous research indicates that the steepest increase in risk for PID after an STI is within the first 60 days.¹³

What could be the explanation for the higher risk of PID for women who douche less frequently? It is difficult

TABLE 5. Proportional Hazards Regression of PID Occurrence (Intent-to-Treat Analysis)

Variable	RR	95% CI
Douche/wipe	1.08	0.58–2.0
Age at entry (yrs)	0.84	0.77–0.92
Black/non-black	1.46	0.61–3.5
Qualifying infection with gonococcus	1.07	0.40–2.9
Qualifying infection with chlamydia	1.22	0.59–2.5
Past frequency of douching	1.31	0.46–3.7
Frequency of sexual intercourse (per month)	0.99	0.94–1.04
New sex partner	0.82	0.27–2.5
Number of sex partners	1.13	0.49–2.6
Married or living as married	1.19	0.43–3.3
Barrier contraceptive use at entry	1.58	0.63–3.9

TABLE 6. Rate of PID According to Frequency of Douching During the Previous 2 Months*

	Frequency of Douching		
	Less than or Equal to Median	More than Median	Wipe Group
Cases of PID	14	4	19
Person-years	252.3	263.3	533.5
Incidence rate (per 100 person-years)	5.5	1.5	3.6

* This analysis, based on women who reported adhering to the study protocol, shows PID rates during the inter-visit period after each questionnaire report, and thus does not include some cases and person-time relating to non-adherence or to periods for which no questionnaire information was available.

to see how douching could cause PID if infrequent douching elevates the risk more than frequent douching. One possibility is that this is only a chance difference. If it is real, it may represent reverse causality: women who douche infrequently may be more likely to douche in response to symptoms of a STI or may douche only after intercourse. Whatever the explanation, if the difference is real it likely signals the effect of some risk factor that is associated with douching behavior (specifically, less frequent douching behavior) rather than an effect of the douche product itself.

The validity of the study findings rests largely on the compliance of women with the study protocol. To the extent that women in the wipe group continued to use a douche product during the study follow-up, or that women in the douche group discontinued their douching practice, the contrast between the assigned groups is weakened and any effect of douching would be underestimated. We addressed this issue in part with the on-protocol analysis, which gave essentially the same results as the intent-to-treat analysis. The intent-to-treat analysis has the theoretical advantage of comparability achieved through random assignment at the cost of some underestimation of the effect. That cost is less consequential for efficacy trials but is more problematic in a safety trial such as this. The on-protocol analysis theoretically reduces the underestimation of effect from lack of compliance, but that advantage comes at the cost of

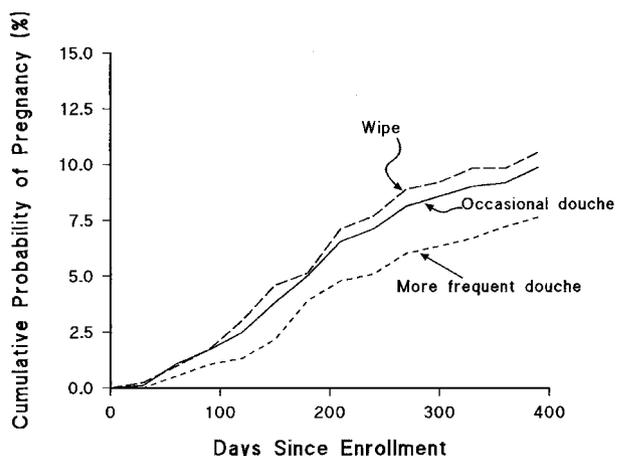


FIGURE 3. Cumulative probability of pregnancy for women assigned to the wipe group, and for women assigned to the douche group, according to their frequency of use (more frequent douche = douching more frequently than the median frequency; occasional douche = douching less frequently than the median frequency).

departing from a purely randomized comparison. In this study, the two analyses produced nearly identical results.

Our primary measure of compliance was the information about product use that was reported at each visit. This self-reported compliance falls far short of an ideal compliance measure, whatever that might be for a study of douching activity. Site monitors reported that wipes were well accepted by women assigned to receive them, some of whom requested additional product between visits. This observation, together with the high frequency of use, suggests that the wipes were an acceptable hygiene alternative for many of the women assigned to that group. Another indicator of compliance is a woman's willingness to continue in the study through the follow-up period. Women received an average of about \$25 per visit to continue, an amount modest enough so that it was unlikely to be a strong motivator by itself to return to the clinic. Women also received free study product at each return visit, which added further incentive for women who liked the product that they were assigned. Thus, the women who remained in the study

TABLE 7. Probability of Pregnancy by Reported Frequency of Use of Study Product (Restricted to Women Who Adhered to the Study Protocol)

	Frequency of Use			
	Less than or Equal to Median Frequency		More than Median Frequency	
	Douche	Wipe*	Douche	Wipe*
Pregnancies	37	42	25	40
Total participants	386	413	376	404
Probability of pregnancy (%)	9.6	10.2	6.6	9.9
Risk ratio and (95% CI)	0.94 (0.62-1.4)		0.67 (0.42-1.08)	

*Reference category.

and had a good record of return visits presumably represent women who were more compliant with the study protocol.

In summary, our data indicate little or no greater risk of PID among women assigned to use the douche product. We also found a modest decrease in the probability that a woman assigned to use the douche product would become pregnant, a decrease that was more apparent among women who douched more frequently. These findings relate specifically to the douche product used in this study. This product, which was newly engineered in accordance with ergonomic studies, is in wide use today but differs from earlier douche designs and those used by other manufacturers. Thus, earlier studies that showed a positive relation between douching and PID may have correctly assessed an increase in risk that occurred with douche products differing from the one studied here. Alternatively, the earlier results may have reflected residual confounding from risk factors for PID that were incompletely controlled because of misclassification. It also remains possible that we underestimated a real increase in risk for PID among women assigned to use the douche product because of some combination of chance and imperfect compliance with the study protocol, or because the follow-up period of the study was insufficient for an effect to become evident.

The results for pregnancy corroborated a finding already in the literature. In a study of women who were attempting to conceive, Baird *et al.*¹⁰ reported 30% lower fertility among women practicing vaginal douching than among women who did not practice douching. They found little change in the douching effect according to whether or not the women used their douche product around the time of sexual intercourse, a finding that is either inconsistent with a real effect of douching on the probability of conception, or at least inconsistent with mechanisms of reduced fertility that work directly on sperm access to the ovum. The difference they reported is greater than the 15% difference in the probability of becoming pregnant that we found in our overall data, but similar to the difference that we found for more frequent users of the douche product.

In our study the participants at the outset asserted that they did not intend to become pregnant during the next 12 months, in contrast to the study of Baird *et al.*¹⁰ in which the women were trying to conceive. Their study population differed from ours in other respects: their population was mostly white and better educated than average, and, most notably, in our study every woman reported having had a gynecologic infection before enrolling. It is not obvious how these differences

would influence the assessment of an effect of douching on conception, but behavioral and social factors have strong influences on the probability of a woman becoming pregnant. In our study, confounding could explain the association between douching and pregnancy: for example, women who are more fastidious about hygiene may have chosen to use a douche product more frequently and also may have used contraceptive agents such as condoms more assiduously. This theory would not explain the finding of Baird *et al.*, however. Despite differences in study design and populations, the similar findings from the two studies lend support to the theory that douching may be related to the probability of pregnancy.

We benefitted from the contributions of many people. Paul Starkey, Keith Callahan, Susan Clement and Peter Fratarcangelo made essential contributions to the conception and the design of the study. Peter Fratarcangelo, Penny Blinder, Helmut Albrecht, Sistine Chen, Theresa DeSantis, Betty Abebe and Eric Johnson contributed to the data collection and monitoring process. James McGregor and Noel Weiss assisted in providing external scientific advice at various stages in the project.

References

1. Wølnner-Hanssen P, Eschenbach DA, Paavonen J, *et al.* Association between vaginal douching and acute pelvic inflammatory disease. *JAMA* 1990;263:1936–1941.
2. Neumann HH, DeCherney A. Douching and pelvic inflammatory disease. *N Engl J Med* 1976;295:789.
3. Scholes D, Daling JR, Stergachis A, Weiss NS, Wang SP, Grayston JT. Vaginal douching as a risk factor for acute pelvic inflammatory disease. *Obstet Gynecol* 1993;81:601–606.
4. Ness RB, Soper DE, Holley RL, *et al.* Douching and endometritis. Results from the PID Evaluation and Clinical Health (PEACH) study. *Sex Transm Dis* 2001;28:240–245.
5. Jossens MO, Eskenazi B, Schachter J, Sweet RL. Risk factors for pelvic inflammatory disease. A case control study. *Sex Transm Dis* 1996;23:239–247.
6. Zhang J, Thomas AG, Leybovich E. Vaginal douching and adverse health effects: a meta-analysis. *Am J Public Health* 1997;87:1207–1211.
7. McCormack WM. Pelvic inflammatory disease. *N Engl J Med* 1994;330:115–119.
8. Grodstein F, Rothman KJ. Epidemiology of pelvic inflammatory disease. *Epidemiology* 1994;5:234–242.
9. Aral SO, Mosher WD, Cates W. Self-reported pelvic inflammatory disease in the U.S., 1988. *JAMA* 1991;266:2570–2573.
10. Baird DD, Weinberg CR, Voigt LF, Daling JR. Vaginal douching and reduced fertility. *Am J Public Health* 1996;86:844–850.
11. Case definitions for infectious conditions under public health surveillance. *MMWR Morb Mortal Wkly Rep* 1997;46:1–55.
12. Centers for Disease Control and Prevention. *STD Surveillance 2000 Report*. Available at: <http://www.cdc.gov/std/stats00/2000SFWomen&Inf.htm>.
13. Rothman KJ, Lanza L, Lal A, Peskin EG, Dreyer NA. Incidence of pelvic inflammatory disease among women treated for gonorrhoea and chlamydia. *Pharmacoepidemiol Drug Safety* 1996;5:409–414.