TECHNICAL PAPERS ON
HEALTH AND BEHAVIOR MEASUREMENT

TECHNICAL PAPER 29

Zenilman's Anomaly Reconsidered: Fallible Reports, *Ceteris Paribus*, Fragile Condoms, and Other Hypotheses

Charles F. Turner and Heather G. Miller

Reference Citation

Turner, C.F., and H.G. Miller. (1997) Zenilman's anomaly reconsidered: Fallible reports, *ceteris paribus,* and other hypotheses. *Sexually Transmitted Diseases* 24(9):522-527.

Zenilman's Anomaly Reconsidered

Fallible Reports, *Ceteris Paribus*, and Other Hypotheses

CHARLES F. TURNER, PHD, AND HEATHER G. MILLER, PHD

From the Program in Health and Behavior Measurement, Research Triangle Institute, Washington, DC

Background and Objectives: **In the January–February, 1995 issue of *Sexually Transmitted Diseases*, Zenilman and colleagues reported a null association between incident sexually transmitted diseases (STDs) and self-reported condom use. That anomalous finding generated a flurry of letters to the editor, some of which were quite heated. This article reconsiders the Zenilman team's results.**

Study Design: **New statistical analyses were conducted to test two hypotheses that sought to account for the null association: (1) deviation from study protocol, and (2) differential risks of acquiring an incident STD among segments of the study population that varied by reported level of condom use.**

Results: **No support was found for hypotheses concerning deviation from study protocol and differential risk of acquiring an incident STD by level of condom use. Indeed, for respondents who reported multiple sexual partners, the analyses found *increased rates of infection among those who reported more consistent condom use*.**

Conclusions: **Two of the most promising hypotheses for explaining Zenilman's anomalous findings are unsupported by reanalysis of the available empirical evidence. It is still possible that respondents who reported that they used condoms consistently differed from self-reported nonusers or inconsistent users in some way that altered their risk of acquiring an STD and thus obscured the protective effects of properly used condoms. Nonetheless, as Zenilman and others suggest, fallibility in self-reports of condom use remains the primary suspect as the cause of these anomalous results. Such fallibility may be particularly pronounced when self-reported behavioral data are collected in contexts that include strong educational campaigns or other norm-setting interventions.**

IN A STUDY OF PATIENTS recruited from a sexually transmitted disease (STD) clinic, Zenilman and colleagues[1] found no association between the consistency of self-reported condom use and the likelihood of acquiring an incident STD.* When those findings were first announced in May, 1993 at a National Institutes of Health conference on the role of condoms in reproductive health, they ignited considerable controversy. The uproar flared anew with the publication of the study in the January–February, 1995 issue of *Sexually Transmitted Diseases*.[2–5]

Because Zenilman and coworkers used interviewer-administered questionnaires to collect behavioral data in their study, biased reporting of condom use is a likely (although unproven) cause of their anomalous results. Past research[6–10] has provided persuasive evidence that self-reports of behavior can be quite fallible, particularly reports of sensitive activities like illegal drug use and sexual and contraceptive behaviors. Those studies suggest that requiring respondents to divulge such information to interviewers (rather than having them complete self-administered interviews) can introduce dramatic measurement biases.

Our suspicions in that regard are fueled by the fact that the reporting of condom use in the Zenilman team's research took place in a clinic environment in which condom use was being promoted. Given the clinic's strong encouragement of condom use, subjects might have felt an understandable tension about providing the socially desirable answer versus the accurate answer to the clinician's question—with the result that their reports might not be fully accurate.

Before accepting error in measures of self-reported condom use as the explanation for these anomalous results, we considered at length five alternative explana-

Supported in part by grant R01-HD/AG31067-03 from the National Institute of Child Health and Human Development and the National Institute on Aging.

Reprint requests: Charles F. Turner, PhD, Program in Health and Behavior Measurement, Research Triangle Institute, 1615 M Street, N.W., Suite 740, Washington, DC 20036.

Received for publication February 20, 1997, revised May 16, 1997, and accepted May 22, 1997.

*By the term *likelihood of acquiring an incident STD*, we mean the likelihood that a subject will either be reinfected with an STD after treatment with an efficacious, single-dose therapy at intake into the study or will become infected after being determined to be disease free at intake. The STDs considered in this study were gonorrhea, chlamydial infection, syphilis, and trichomoniasis.

tions. Elsewhere[11] we consider and find unconvincing three such hypotheses:

- Deficiencies in research procedures obscured the association between condom use and incident STDs.
- The null association occurred because a substantial number of people who became infected during the study were treated elsewhere.
- The reports of condom use were accurate, but high rates of mechanical failure or incorrect usage perturbed the association between condom use and incident STDs.

In this article we present new empirical evidence derived from reanalyses of the original data. The analyses tested two further hypotheses:

- Deviation from study protocol may have induced selective *sample attrition* that compromised the study's ability to detect the association between condom use and incident STDs. (Self-selection biases associated with attrition in a study sample are well-known threats to the validity of inferences drawn from prospective observational studies.)
- The assumption of *ceteris paribus* was violated in that people reporting different levels of condom use also had different levels of objective risk of infection. That could have occurred, for example, if the risk of exposure to *infected* partners was not the same for subjects who reported different levels of condom use. Theoretically, there could also be differences in susceptibility (for reasons other than condom use), but that explanation seems unlikely to explain the Zenilman team's findings.

Methods

Data

We analyzed data extracted from the complete dataset of responses by 598 subjects that Zenilman and colleagues presented in *Sexually Transmitted Diseases* in 1995. The extracted dataset supported analyses of the following categorical variables:

C: Condom Use—Self-reported frequency of use during vaginal and anal intercourse. The variable had three response categories: always, sometimes, or never.

I: Infection—Assessed incident infections with *Neisseria gonorrhoeae, Chlamydia trachomatis, Treponema pallidum, and Trichomonas vaginalis*. Subjects were screened at time 1 and treated if infected. Subjects were rescreened for those STDs at time 2. This variable had two categories: infected or not infected at time 2.

S: Prior STDs—The self-reported number of STDs diagnosed before time 1. This variable was dichotomized as more than the median number of STDs reported by participants of the same gender versus less than the median for that gender.

P: Number of Sexual Partners in Past 30 Days—The self-reported number of sexual partners during the 30 days before the interview. This variable was dichotomized as one partner versus two or more partners.

G: Gender of Subject—Male or female.

N: Number of New Sexual Partners in Past 30 Days—The self-reported number of new sexual partners during the 30 days before the interview, dichotomized as none versus one or more.

T: Timing of Second Interview—A proxy variable for attrition that was based on adherence to the protocol for the time 2 interview and clinical assessment. This variable was dichotomized as (1) the second interview was completed within 2 weeks of the scheduled appointment versus (2) the second interview was completed more than 2 weeks before or after the scheduled date.

Proxy Measures

Deviation From the Study Protocol and Selective Attrition. In prospective clinical studies, it is not unusual for a subset of subjects who participate in the early phases of research to fail to return in follow-up phases—or to return at times other than those scheduled. Investigators are necessarily concerned about biases from such events being insinuated into a study's findings. A particular worry in the current study was the possibility that subjects with an incident STD might have had a different likelihood of completing the time 2 reinterview, because the interview could have been completed when subjects came in for treatment of symptoms rather than for their reinterview as scheduled in the study protocol. To test for such effects, we used the timing of the reinterview as a control variable in our analyses.

Differential Risk of Infection. We cannot measure directly all of the factors that increase a subject's risk of acquiring an incident infection. For example, we can ask subjects how many new sexual partners they have had in a given period, but they are unlikely to know with any degree of certainty whether a new partner was infected. And even if current and recent partners could be recruited as study participants, investigators would prob-

ably never have access to the universe of all subjects' partners.

With no information on a partner's infectivity status—obviously a central determinant of the risk a subject faces—we are left to our own devices to construct a reasonable surrogate for the extent to which the "pools" of sexual partners available to different groups vary in the risks they pose. The most reasonable surrogate indicator for infection in sexual partners is likely to be the subject's own past history of STDs. This surrogate indicator is not without problems. In particular, we must presume that the historical data do not reflect historical differences in condom use. It is, however, the best we can do with the available data. The variable representing the reported number of lifetime STDs was dichotomized at a point that coincided with the gender-specific medians for the sample. We thus defined groups that historically had more (vs. fewer) STDs than the average man (or woman) in the sample. As a further control for differences in sexual exposures, we constructed two dichotomous variables reflecting the subjects' number of partners in the past 30 days (variable 1: zero or one partner vs. two or more partners; variable 2: zero *new* partners vs. one or more *new* partners).

Statistical Analyses

Our analyses fit a series of hierarchical log-linear models[12,13] to test whether the proposed hypotheses might account for the Zenilman team's anomalous findings. Our intent was to discover whether a particular way of viewing the data might reveal the expected pattern of association between condom use and incident STDs for some theoretically meaningful subset of subjects.

Results

Deviation From the Protocol and Selective Attrition

As Table 1 shows, the rates of reinfection varied substantially between male subjects who returned within 2 weeks of their appointment and male subjects who did not. Yet neither subgroup's pattern of results conformed to expectations. For both groups, the rates of reinfection were approximately equivalent for men who reported "always" and "never" using condoms.

The data for women who returned within 2 weeks of their scheduled appointment suggested a trend in the appropriate direction. For that subgroup, the rates of reinfection declined from 30% for those who reported never using condoms to 24% for those who sometimes used them and 19% for those who reported always using them. Although such results might seem encouraging, they are not statistically significant, nor is a similarly suggestive trend seen in the data provided by men. Thus, on the one hand, the empirical evidence does not rule out attrition or deviation from the study protocol as a potential contributor to the anomalous findings. On the other, it does not lend strong support to such a notion.

Differences in Risk of Sexually Transmitted Disease Infection

A subject's likelihood of acquiring an STD during a given interval can be conceptualized as a function of five variables: (1) the infectiousness of the pathogen, (2) the susceptibility of the subject, (3) the number of new sexual partners a subject has during the interval (and the extent of exposure to each), (4) the probability that a new partner is infected with the pathogen, and (5) the infectiousness of the partner. In theory, the use of condoms should diminish a subject's susceptibility, which in turn should be reflected in a reduced likelihood of infection. However, that proposition does not necessarily hold in the case of the Zenilman team's data. Subjects were not randomly assigned to different levels of condom use. Thus, other STD-related factors (such as number of new partners or STD rates in the population from which those partners were selected) would not necessarily be equivalent across all subjects reporting the same level of condom use. Differences in one or more of those factors could obscure the protective effects of condom use.

Under that alternative explanation, the verbal reports of condom use obtained with interviewer-administered questionnaires might be (approximately) accurate, but a null association of condom use and incidence of STD infection could arise because the risk of exposure to infected partners was not the same for all subjects. To test that explanation, we controlled for the total number of partners and the number of new partners, and we used a subject's history of STDs as a surrogate measure of the likelihood that his or her partners were infected (Table 2). The four-way tables for each control variable (infection by condom use by gender by STD history or by number of partners) include simple tests for the presence of an association between condom use and reinfection rates for the subpopulations defined by each row of the tables.

Although the four-way tables display much information of interest—including a demonstration that simple explanations of this sort are inadequate to account for Zenilman's anomaly—there is little suggestion and no statistically reliable evidence that the likelihood of contracting an incident STD is associated with reported condom use for any of these subpopulations. It is possible, however, that higher-order interactions are occurring and obscuring the expected association.

TABLE 1. Percentage of Subjects Who Were Reinfected at the Time of Reinterview, by Gender, Self-Reported Condom Use, and Timing of Follow-Up Interview*

Timing of Follow-Up	Gender	Percentage Reinfected			P Linear Association	P Categorical Association
		Never Used Condoms	Sometimes Used Condoms	Always Used Condoms		
2+ weeks outside of scheduled appointment	Male	19.6 (97)	34.9 (43)	18.9 (37)	0.67	0.13
Within 2 weeks of scheduled appointment	Male	10.5 (86)	4.0 (25)	11.4 (35)	0.97	0.51
2+ weeks outside of scheduled appointment	Female	29.9 (87)	24.2 (33)	18.5 (27)	0.22	0.46
Within 2 weeks of scheduled appointment	Female	23.5 (81)	21.7 (23)	29.2 (24)	0.63	0.81

*Whether the follow-up interview was completed within 2 weeks of the scheduled appointment.
P values are for the Mantel-Haenszel test for linear association (*df* = 1) and the likelihood ratio chi-square test for categorical association (*df* = 2). Base *Ns* are shown in parentheses.

To provide a more comprehensive and statistically adequate test for such interactions, we fit a full set of hierarchical log-linear models to the five-way tables: {ICGP$_1$T}; {ICGP$_2$T}; and {ICPST}—where I represents infection with an STD at follow-up, C is reported condom use, G is gender, P$_1$ is number of sexual partners in the past 30 days, P$_2$ is acquisition of a new sexual partner in the past 30 days, and T is timing of the follow-up interview (whether it was within 2 weeks of the scheduled appointment; see Table 1). For each of the tables, we fit 14 log-linear models; the models began with a base model that included terms to allow explicitly for any extant associations between infection and all other variables except condom use, and between condom use and all other variables except infection. For example, for the table {ICPST}, our base model included terms to fit

TABLE 2. Percentage of Subjects Who Were Reinfected at the Time of Reinterview, by Gender, Self-Reported Condom Use, and Control Variables*

Control Variables	Gender	Percentage Reinfected			P Linear Association	P Categorical Association
		Never Used Condoms	Sometimes Used Condoms	Always Used Condoms		
Partners in past 30 days						
0–1	Male	17.1 (129)	25.0 (48)	13.7 (51)	0.83	0.33
2+	Male	11.1 (54)	20.0 (20)	19.0 (21)	0.31	0.52
0–1	Female	27.2 (92)	20.0 (20)	19.2 (26)	0.35	0.61
2+	Female	26.3 (76)	25.0 (36)	28.0 (25)	0.92	0.97
New partners in past 30 days						
1 or more	Male	27.8 (36)	28.6 (21)	26.7 (15)	0.96	0.99
None	Male	12.2 (147)	21.3 (47)	12.5 (56)	0.70	0.31
1 or more	Female	50.0 (10)	57.1 (7)	50.0 (4)	0.94	0.95
None	Female	24.8 (157)	18.4 (49)	21.7 (46)	0.51	0.62
Prior history of STDs						
More than median for gender	Male	11.1 (54)	20.0 (20)	19.0 (21)	0.31	0.52
Less than median for gender	Male	17.1 (129)	25.0 (48)	13.7 (51)	0.83	0.33
More than median for gender	Female	26.3 (76)	25.0 (36)	28.0 (25)	0.92	0.96
Less than median for gender	Female	27.2 (92)	20.0 (20)	19.2 (26)	0.35	0.61

*Control variables are (1) total no. of sexual partners in past 30 days, (2) no. of new sexual partners in past 30 days, and (3) prior history of STDs. *P* values are for the Mantel-Haenszel test for linear association and the likelihood ratio chi-square test for categorical association. Base *Ns* are shown in parentheses.

TABLE 3. Percentage of Subjects Who Were Reinfected at the Time of Reinterview, by Self-Reported Condom Use and Numbers of Sexual Partners in Past 30 Days

Partners in past 30 days	Percentage Reinfected: Never Used Condoms	Sometimes Used Condoms	Always Used Condoms	*P* Linear Association	*P* Categorical Association
0–1	20.8 (298)	14.8 (88)	16.4 (110)	0.33	0.22
2+	20.8 (53)	44.4 (36)	38.5 (13)	0.05	0.05

P values are for the Mantel-Haenszel test for linear association ($df = 1$) and the likelihood ratio chi-square test for categorical association ($df = 2$). Base *Ns* are shown in parentheses.

the marginals {IPST} and {CPST}. We took that approach because our purpose was to discover associations between condom use and infection that might have been moderated by one or more other variables. In fitting the variables, we treated the three-category condom use variable as a categorical variable to allow us to detect nonmonotonic, higher-order interactions.

The results of our model fitting and tests of the significance of model parameters would fill several pages, but in the end, we found relatively little in the way of an explanatory higher-order interaction. Rather than report the details of those nonfindings, we summarize the few significant or borderline results that we detected with this exercise:

- No analysis detected an interaction effect that yielded the expected protective association between consistent condom use and diminished risk of infection with an incident STD for any of the subpopulations (e.g., people with multiple partners, people with histories of more frequent STDs).
- A significant interaction effect *in the reverse direction* was found for some subpopulations. Considering the five-way table of infection by condom use by number of partners by past STDs by timing of follow-up, or {ICPST}, we found a significant three-way interaction, [ICP], for the three-way subtable of infection by condom use by number of partners (likelihood ratio chi-square, $L^2 = 8.8$, $df = 2$, $P < 0.05$). The three-way tabulation {ICP} is displayed in Table 3. For the 496 subjects who reported only one sexual partner in the past 30 days, the table shows a weak and statistically insignificant ($P = 0.22$) trend for increasing condom use to be associated with the decreased likelihood of having an incident STD. But for the 102 subjects who reported two or more partners during the past 30 days, the table shows a statistically significant trend in the opposite direction. For subjects with two or more partners in the prior 30 days, the rates of incident STD infections were markedly higher among subjects who reported sometimes or always using condoms (44.4% and 38.5%, respectively), compared with subjects who reported never using them (20.8%). Although the sample of people with two or more partners is small, the trend is statistically significant ($P = 0.05$).
- The same three-way interaction, [ICP], was found to be significant in the fitting of log-linear models to the five-way table {ICPGT}—infection by condom use by number of partners by gender by timing of follow-up (likelihood ratio chi-square, $L^2 = 8.0$, $df = 2$, $P < 0.05$).

Discussion

We share the concern of other commentators about the interpretation of the Zenilman team's anomalous results. However, to the extent that additional evidence may be elicited through secondary analyses, we find little suggestion of a simple, satisfying explanation that would make these findings less puzzling—and less troubling. Indeed, our empirical analyses indicate that for the subpopulation of subjects who reported multiple partners, the anomaly included not merely a failure to find an association but rather the finding of a significant association in the wrong direction. In other words, *higher* rates of incident STDs were found among subjects who reported *more consistent* condom use.

We believe that these anomalous results are due to a combination of fallible reporting by respondents and violations of *ceteris paribus*. Thus, across *subpopulations reporting different levels* of condom use, we suspect that systematic differences existed in both the extent of bias in reports of condom use and the risk of exposure to an STD.

Because these data were collected in an observational rather than an experimental study design, levels of condom use were not randomly distributed across subjects. Consequently, an infinite number of differences could be postulated. Consider the results that would arise from "strategic" condom use coupled with fallibility in reports of use. Assume, for example, that subjects who re-

ported consistent condom use had not actually used condoms correctly in 100% of their sexual encounters, whereas those who reported never using condoms had, in fact, never used them. Assume as well that respondents could assess, albeit with substantial inaccuracy, the relative risk posed by different sexual partners. (Such assessments could reflect judgments based on several factors: the nature of the relationship—for instance, spouses, steady partners, casual partners, and so forth; the extent, if any, of the partner's sexual exposure to other potentially infected partners; or other factors that were correlated with the actual likelihood of the partner's being infected.) If subjects attempted to use condoms "strategically" with "riskier" partners, then subjects who reported "always" using condoms could be drawing their partners from a pool that posed a much higher risk of infection than the pool that other subjects were drawing from. In the presence of such fallible reporting and errors in assessing the relative risk posed by different partners, a negative association between condom use and infection rates could easily arise.†

The fallibility of self-reports that we posit is consistent with the Zenilman team's own interpretation. Our finding of a negative association between condom use and incident infection rates for people with multiple partners indicates that other factors are at work as well.

We believe that Zenilman and colleagues have performed a valuable service by conducting and publishing their work and by enduring the torrent of commentary that it generated. As readers ponder the implications of this research and our reconsideration of it, we hope that they will advance beyond simplistic notions of "validated behavioral measurements." Demonstrating a strong association between self-reports of behavior and other, independent evidence of that same behavior in one measurement context does not ensure that the same results will generalize to other measurement contexts. Neither does it indicate that those behavioral measurements have been made without error.

Understanding the conditions under which measurements yield the expected associations and those under which they do not is part of the burden of scientific measurement. Although challenges abound in measuring behavior through self-reports—and, in particular, measuring private behavior that does not usually allow for independent corroboration—those challenges are not unique and the territory is not entirely uncharted.[14–16] Furthermore, failures of measurement in those domains would be consistent with the long history of measurement fallibility in other areas of science.[17] As McNemar[18] noted more than 50 years ago: "All measurement is befuddled with error. About this the scientist can and does do something: he ascertains the possible extent of the error, determines whether it is constant (biasing) or variable, or both, and ever strives to improve his instruments and his techniques."

†In the extreme, consider a sample heavily seeded with (1) subjects who accurately reported never using a condom because they were appropriately confident that their partners had not been exposed to STDs transmitted from other partners and (2) subjects who reported "always" using condoms (but actually used them with less-than-complete consistency) and who had good reason to believe that their partners were likely to be infected. With proper selection of the extent of infection among the pool of partners for the latter group and the rate of inconsistency of their use of condoms, one could produce large, negative correlations between reporting of condom use and the likelihood of infection.

References

1. Zenilman JM, Weisman CS, Rompalo AM, et al. Condom use to prevent STDs: The validity of self-reported condom use. Sex Transm Dis 1995; 22:15–21.
2. Galavotti C, Cabral R, Beeker C. Condom use to prevent STDs. Sex Transm Dis 1996; 23:77–79.
3. Schacter J. Editorial response. Sex Transm Dis 1996; 23:82.
4. Weir SS, Feldblum PJ. Condom use to prevent STDs [letter]. Sex Transm Dis 1996; 23:76–77.
5. Zenilman JM, Ellish N, Rompalo AM, Weisman CS, Hook EW, Celentano D. Condom use to prevent STDs [letter]. Sex Transm Dis 1996; 23:79–82.
6. Jones EF, Forrest JD. Underreporting of abortion in surveys of U.S. women: 1976 to 1988. Demography 1992; 29:113–126 (see especially Table 1).
7. Aquilino W. Interview mode effects in surveys of drug and alcohol use: A field experiment. Public Opinion Quarterly 1994; 58:210–240.
8. Turner CF, Lessler JT, Devore JW. Effects of mode of administration and wording on reporting of drug use. In: Turner CF, Lessler JT, Gfroerer JC, eds. Survey Measurement of Drug Use: Methodological Issues. DHHS Publication No. 92-1929. Washington, DC: Government Printing Office, 1992:177-219.
9. Turner CF, Ku L, Sonenstein FL, Pleck JH. Impact of audio-CASI on bias in reporting of male-male sexual contacts. In Warnecke R, ed. Health Survey Research Methods. Hyattsville, MD: National Center for Health Statistics, 1996:171-176.
10. Turner CF, Danella R, Rogers S. Sexual behavior in the United States, 1930–1990: Trends and methodological problems. Sex Transm Dis 1995; 22:173–190.
11. Turner CF, Miller HG. Zenilman's paradox reconsidered. Technical Papers in Health and Behavior Measurement, No. 29. Washington, DC: Research Triangle Institute, 1996.
12. Goodman LA. The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables. Journal of the American Statistical Association 1968; 63:1091–1131.
13. Goodman LA. Analyzing Qualitative/Categorical Data. Cambridge, MA: Abt Associates, 1978.
14. Bradburn NM, Sudman S, and Associates. Improving Interview Methods and Questionnaire Design. San Francisco: Jossey Bass, 1979.
15. Turner CF, Martin E. Surveying Subjective Phenomena. 2 vols. New York: Russell Sage, 1984.
16. Wiley JA, Turner CF, Sorensen JL, et al. Methodological issues in AIDS surveys. In: Miller HG, Turner CF, Moses LE, eds. AIDS: The Second Decade. Washington, DC: National Academy Press, 1990:359-471.
17. Turner CF, Martin E. Surveying Subjective Phenomena. Vol. 1, Chapter 1. New York: Russell Sage, 1984:3-21.
18. McNemar Q. Opinion–attitude methodology. Psychol Bull 1946; 43:289–374.