

continues to be supportive of gay affirmative therapy and gay rights.

Much of the criticism of Spitzer's study is likely to focus on sample bias because many of his subjects were referred by religiously conservative organizations which promote the possibility of change; however, as astutely noted by Rosik (2003), this may actually be a strength of the research given that the clear majority of the studies on homosexuality and change used convenience samples solicited through gay-affirming organizations and media. Rosik (2003) further notes that "this suggests a skew in the existing literature as a whole, the degree of which can only be determined through a closer examination of individuals such as those in Spitzer's study" (p. 18).

A substantial majority of the participants in the Spitzer study valued and agreed with their traditional religious faiths, which view homosexual relationships as non-life-giving and outside of the will of God. Such value systems should not be dismissed, judged by the therapist to be improper, or overridden through therapeutic manipulation. Haldeman (2000) offers an interesting perspective in this regard:

A corollary issue for many [clients] is a sense of religious or spiritual identity that is sometimes as deeply felt as is sexual orientation. For some it is easier, and less emotionally disruptive, to contemplate changing sexual orientation, than to disengage from a religious way of life that is seen as completely central to the individual's sense of self and purpose. However we may view this choice or the psychological underpinnings thereof, do we have the right to deny such an individual treatment that may help him to adapt in the way he has decided is right for him? I would say that we do not. (p. 3)

As LeVay (2000) explains it:

First, science itself cannot render judgments about human worth or about what constitutes normality or disease. These are value judgments that individuals must make for themselves, while possibly taking scientific findings into account. Second, I believe that we should as far as possible, respect people's personal autonomy, even if that includes what I would call misguided desires such as the desire to change one's sexual orientation. (p. 12)

Spitzer's research has demonstrated that, contrary to the prevailing climate, the data on homosexuality are far from complete. Ethicality would suggest that the suppression of data and discouragement of further scientific research should not be tolerated. With appropriate guidelines in place (institutional review boards), it is not only ethical but well within the purview of science to encourage the study of issues such as change from homosexuality. The well-intentioned caretakers of our national organiza-

tions slide down a slippery slope when advocating what amounts to a virtual censorship of scientific investigation of politically unpopular views. It is ironic that Spitzer—the original architect of the 1973 decision to remove homosexuality from the DSM—is once again, going against the prevailing winds of his time and advocating the avoidance of that slippery slope.

A Methodological Critique of Spitzer's Research on Reparative Therapy

Helena M. Carlson, Ph.D.

Department of Psychology, Lewis and Clark College, Portland, Oregon 97219; e-mail: carlsonh@earthlink.net.

In Spitzer's study of the effectiveness of reparative therapy in changing sexual orientation, he reports that gay men and lesbians indicate that they have made major changes in their sexual orientation from homosexual to heterosexual. He also notes that even those who made only limited change in sexual orientation still found the therapy beneficial.

The criteria for acceptance into the study required subjects to have had a predominantly homosexual orientation before entering reparative therapy. They should be able to report that after reparative therapy they have sustained for at least 5 years some change toward a heterosexual orientation. The acceptance criteria also required participants to report in a telephone interview on their sexual behavior for the year before they entered reparative therapy and then also report on their behavior in the year before the current research interview. Spitzer reports that there was, on average, an interval of 12 years between the time of entry into reparative therapy and the telephone interview. This places a heavy burden on memory and Spitzer acknowledges there is greater fallibility in such long range memory.

Participants in the study come from a very narrow stratification of the population: 97% were Christian, 95% were Caucasian, the mean age for males was 42 years, the mean age for females was 44 years, 76% of the males were married, and 47% of the women were married. Some participants were directors of ex-gay ministries and some had publicly spoken favorably of efforts to change sexual orientation, often at their church. Thus, this is a population of highly religious, White, Protestant, middle aged, and middle class men and women. There is little evidence that they are representative of a diverse gay community.

Participants in this study were asked in a telephone interview to report on their sexual fantasies, masturbation fantasies, lustful looks, use of gay pornography, homosexual thoughts, and overt sexual behavior. A key question is the credibility of the participants' self-report. It should be

recognized that 93% of participants reported that religion was extremely or very important to them.

No consent form was administered and participants' identity was known to the interviewer. Ethical guidelines for informed consent for research issued by the American Psychological Association (2002) require not only protection of confidentiality for participants but that a consent form should clearly state that participants may withdraw from research at any time. Without a consent form, it is possible that participants were wary that their confidentiality would not be protected. The total reliance on self-report in this study can be disturbing when one considers that fundamentalist religious beliefs tend to be strongly opposed toward any acceptance of homosexuality.

Martin (2000) has pointed out that when the research topic is an emotion-laden issue, then individuals might not wish their true feelings to be known, particularly when these feelings differ from socially accepted practices in their community. It seems that this highly religious Christian sample would be particularly vulnerable to feelings of shame and embarrassment if they had to report that they had engaged in condemned behavior. There is a significant risk for self-deception and even lying in highly religious participants when responding to questions about sexual behavior that is strongly condemned by their religion. These participants would also be highly motivated to providing supportive data for the possibility of change of sexual orientation.

Another methodological concern is that all the telephone interviews were done by the investigator alone. This raises the methodological issue of interviewer bias. An interviewer can subtly influence respondent's answers by inadvertently indicating approval or disapproval (Cozby, 2001). Interviewers may also bring their own expectations to their interviews and that can bias their interpretation of responses. Although no research method is absolutely free of bias, the interview is more open to bias than most other research methods (Sommer & Sommer, 2002). It would have been better to have trained and used other interviewers, preferably those blind to the hypothesis of the study. A research assistant did independently rate audio recordings of 43 of the 200 interviews and Spitzer reports they achieved .98 interrater reliability based on this sampling. It would be helpful if one had some details on the background of the research assistant in order to evaluate more fully the interrater reliability.

Another issue of concern was the diversity and background of the therapies or counseling offered. Although all were described as reparative or conversion, with the goal to change sexual orientation from homosexual to heterosexual, the therapists came from different educational and

training backgrounds. The majority (43%) were from ex-gay religious ministries, primarily Protestant, who focused on conversion to heterosexuality; 23% came from a group of primarily psychoanalytic mental health professionals and lay people with the same focus; 9% were recruited as participants by their former therapists; and 25% were a variety of sexual reorientation counselors, including social workers, ministers, and lay people.

Despite the diverse educational and training background, Spitzer lumps all types of counseling together as reparative therapy. Since religious approaches may well be in the form of prayer, it is difficult to see how this form of counseling can be combined in data from the treatments used by trained psychoanalysts. This presents confusion since it is unclear to which particular type of therapy any reported changes can be attributed.

Spitzer used a numerical scale of sexual attraction to determine whether participants had a predominantly homosexual or heterosexual orientation before entering therapy and to assess any changes in sexual orientation by comparing scores on this measure after they had received reparative therapy. He defined participants in the study as predominantly homosexual if they scored at least 60 on the scale of sexual attraction before seeking therapy. He also required before acceptance in the study that participants report a change of at least 10 points, lasting at least 5 years, toward the heterosexual end of the scale.

It is difficult to assess data from this Sexual Attraction Scale, which appears to have been designed for this study. This is a 100 point scale (where 0 = *exclusively heterosexual* and 100 = *exclusively homosexual*). Spitzer defined a score ≥ 20 as homosexual (see Table I) and used the same score on the Sexual Orientation Self-Identity Scale. It appears that, for example, participants who scored 25 on this scale will be recognized as homosexual and similar to participants who scored 100 on this scale.

Spitzer reported that married couples were mailed copies of the Dyadic Adjustment Scale (Spanier, 1976). They were asked to complete this questionnaire independently of their partner and to mail it in. Spitzer reported a 72% response rate from married participants and that, on average, subjects reported the same degree of marital adjustment as the instrument's normative reference group. It is difficult to assess data from this measure because no description was given of it, no complete reference for it was given, and no validity cited. There was also no control over when and how the respondents actually completed the measure.

There was no control group in this research, although admittedly that would be hard to obtain. This means that causality cannot be demonstrated. Spitzer acknowledged that there are fundamental methodological problems with

the research but also claims that it provides support for the possibility of reparative therapy to change sexual orientation from homosexual to heterosexual. Spitzer cited some nonsexual benefits from this therapy. He noted that participants reported that after therapy they had a greater sense of masculinity in males and femininity in females. This needs more clarification in light of the many studies of the complexities of gender roles (Crawford & Unger, 2000; Kilmartin, 2000).

In conclusion, even the limited hypothesis that some individuals whose orientation is predominantly homosexual can become predominantly heterosexual following reparative therapy is not supported by this study. It may be possible that some of the research participants might have a more fluid sexual orientation, such as bisexuality (Bohan, 1996).

Are Converts to Be Believed? Assessing Sexual Orientation "Conversions"

Kenneth M. Cohen, Ph.D., and Ritch C. Savin-Williams, Ph.D.

Counseling & Psychological Services, Gannett Health Center, Cornell University, Ithaca, New York 14853; e-mail: kmc17@cornell.edu; Department of Human Development, Cornell University, Ithaca, New York; e-mail: rcs15@cornell.edu.

The realization that subject selection criteria significantly influence results when investigating socially stigmatized populations is one of the earliest and most frequently documented lessons learned by sexologists. Examples of the biasing influence of subject selection are legendary, including the early discovery that recruitment venue could determine outcomes, such as age of coming out, number of gay friends, and self-esteem level (Harry, 1986) and, more recently, that selecting gay youth from support groups (Savin-Williams & Ream, in press) or with sex-atypical behavior (McDaniel, Purcell, & D'Augelli, 2001) escalates levels of reported suicidality. Reviewing problems inherent in sampling "homosexuality," Sandfort (1997) argued that "findings in a specific study depend heavily on the definition and operationalization of homosexuality adopted, and on the way the sample has been put together" (p. 261). Yet, sample selection remains one of the "unresolved issues in scientific sexology" reviewed by McConaghy (1999) several years ago in this *Journal*.

No where is this issue more pertinent than in assessing the highly volatile issue of sexual orientation change following reparative therapy. As Spitzer noted in his literature review, the question of the effectiveness and consequences of "reorientation" therapy has been a source of vitriolic controversy. Because of this current debate, we

believe that sexologists must be extra vigilant to conduct the most methodologically sophisticated, conservative investigations possible, which implies being particularly attentive to subject selection. We do not believe that it is justifiable for contemporary research scholars to presume their conclusions by selection criteria: Including subjects most likely to agree with the author's hypothesis while excluding those most likely to give disconfirming results. To permit conclusions about causality, which Spitzer denies making, but nevertheless embraces ("change in sexual orientation following some kind of therapy does occur in some gay men and lesbians"), the scientific method must be fully embraced, for only then will it be possible to determine whether events (sexual orientation change) are the consequence of chance, untested variables, or the study's independent variable (reparative therapy). Scientific research also depends on the willingness of the researcher to question the truthfulness of subject claims. Given the failure of previous scientific attempts to demonstrate change in sexual orientation (versus sexual identity or behavior), the findings by Spitzer deserve close methodological scrutiny.

In this commentary, we ignore Spitzer's apparent conceptual misunderstanding about the purpose of his study (a title and literature review that promises a study about changes in sexual orientation from homosexual to heterosexual, but a proposed hypothesis that merely suggests shifts in sexual attractions, an event not uncommon among many sexual-minority individuals—see Diamond, 2003a). We also ignore several methodological procedures that create doubt about whether Spitzer used proper scientific methods to ensure the validity of his data, including his definition of sexual orientation, the lack of reliability and validity data on instruments, and the failure to behaviorally anchor response items. Instead, our focus is on subject selection biases that raise serious questions about the veracity of subject claims of reorientation. It is our contention that Spitzer selected a unique group of subjects who were decidedly invested in demonstrating the possibility and benefits of reparative therapy. This one fatal flaw seriously diminishes the internal and external validity of his study and necessarily precludes the very conclusions Spitzer offered: "The mental health professions should stop moving in the direction of banning sexual reorientation therapy" and that "many patients . . . can make a rational choice to work toward developing their heterosexual potential and minimizing their unwanted homosexual attractions."

Who were Spitzer's subjects and how could this collective affect the internal and external validity of his findings? In terms of venue, volunteers were recruited through "repeated" notices to ex-gay religious ministries,