

CITATION

Schumm, W. R. (2015)
Navigating treacherous
waters—one researcher's
40 years of experience with
controversial scientific
research. *Comprehensive
Psychology*, 4, 24.

Navigating treacherous waters—one researcher's 40 years of experience with controversial scientific research^{1, 2}

Walter R. Schumm

Kansas State University

Abstract

Science often must deal with issues that are politically controversial. However, there are dangers in dealing with controversial research and serious risks to the process of doing science and to the credibility of science, particularly social science. Here, I discuss lessons learned from engaging in and criticizing controversial research for nearly four decades. Social science research as a process is being damaged by questionable research practices, several of which are discussed. Social science results are being misrepresented through a variety of weak or incorrect methodologies, each of which is discussed. Discourse about social science results often shifts from academic discussion into attempts to discredit those with whom one may disagree. Science and the public are not being well served by these problems, so new researchers and policymakers need to be aware of them. For teaching purposes, examples are also presented of controversial research in which new analyses offer different results than previously reported.

My journey dealing with controversial research may have begun with my older brother's dissertation (Schumm, 1966), about which he told me, when I was 15-yr. old, that probably one-third of what scientists had thought they knew about the research topic had been incorrect. That situation imbued me with a certain sense of skepticism about scientific research on one hand, but a sense of optimism on the other hand that one could – through better science – correct such things and make improvements. Only later did I find Cohen (1990) saying essentially the same thing with respect to social science – “One of the things I learned early on was that some things you learn aren't so” (p. 1304).

Even well-known historical events can be deconstructed statistically to show that things did not occur as we have been led to believe. Here are some examples of which readers may or may not be aware. The week before Pearl Harbor, the U.S. was trying to ambush the Japanese. In the RMS Titanic disaster, the lowest survival rates for men and the highest survival rates for women and children were among the middle class passengers, suggesting a new nonlinear theory of social class and compliance with social rules. The Challenger disaster could have been predicted in advance with simple statistics (Schumm, Webb, Castelo, Akagi, Jensen, Ditto, *et al.*, 2002). There are many other examples, in this author's own experience, in which research did not turn out as might have been expected. Grover, Russell, Schumm, and Paff-Bergen (1985) showed that the best predictor of later marital satisfaction was the length of time taken before the decision to marry, not the length of engagement. Gwanfogbe, Schumm, Smith, and Furrow (1997) reported that in some situations a wife might be happier in a bigamous marriage. Hendrix, Jurich, and Schumm (1995) showed that adverse effects on a veteran's family life after observing prisoner abuse lasted for decades. In Moxley, Eggeman, and Schumm

¹Address correspondence to Dr. Schumm, School of Family Studies and Human Services, Justin Hall, Kansas State University, 1700 Anderson Avenue, Manhattan, KS 66506-1403. This report is dedicated to the officers and crew of the U.S.S. Hannibal (AG-1), the U.S.S. Borie (DD-215), the U.S.S. Oklahoma (BB-37), the U.S.S. San Diego (CL-53), the U.S.S. Okanagan (APA 220), and the U.S.S. Salem (CA-139) (Schumm, 2002b). Only the Salem survives, as a museum ship in Massachusetts. Preparation of this manuscript was supported in part by a summer faculty grant from the Witherspoon Institute, Princeton, New Jersey; the Witherspoon Institute had no influence or control over the content of the manuscript.

²This paper was invited by the Editors as a discourse from a senior scientist familiar with long-running controversy. Hopefully it will provide a personal view of the problems with entrenched political bias in the social sciences, as demonstrated by Crawford, Duarte, Haidt, Jussim, Stern, and Tetlock (2015) It may be harder than we thought, but political diversity will (still) improve social psychological science. *Behavioral and Brain Sciences*, 38, e164.

(1986), marriage counseling often helped a couple decide what to do, but did not necessarily save marriages. Schumm's data (2010b) on parental sexual orientation did seem to significantly predict children's sexual orientation. Schumm, Jeong, and Silliman (1990) presented data indicating that Protestant fundamentalism was associated with marital distress. When there were differences in marital satisfaction between spouses, the wife was more often the less satisfied (Schumm, Jurich, Bollman, & Bugaighis, 1985) or wives were less satisfied overall with their marriages (Crawford, Schumm, & Schumm, 2015). Family violence is not confined to low-income families (Schumm, Martin, Jurich, & Bollman, 1982). Schumm, McCollum, Bugaighis, Jurich, Bollman, and Reitz (1988) reported that Hispanic families often valued family life more and were more satisfied with family life than Anglos, even though they were worse off financially. Legalization of same-sex marriage may create more rather than less inequality (Schumm, in press).

There are some who think that *most* early scientific findings are incorrect or inflated (Ioannidis, 2005, 2008, 2012). That may be true, but my focus here will be on findings associated with more controversial research, based on my experiences over the past 40 yr., primarily concerning situations where political forces had a vested interest in proving the null hypothesis or preserving traditional historical narratives.

First, I will discuss common problems to anticipate when doing controversial research; then I will discuss some common methodological errors that can complicate the interpretation and assessment of controversial research. Lastly, I will address generally baseless criticisms encountered when dealing with controversial research. The content areas will include topics of which I have detailed knowledge: homosexuality, policy with respect to homosexuals in the military, same-sex marriage and parenting, causes of Gulf War illnesses, religious behavior, etc. My goal is to present illustrative examples of issues associated with controversial research rather than presenting every possible example, which would require a very long book.

The reader should keep in mind this main point: *controversial research of any kind, on any topic, is likely to share these problems I point out from my own areas of research.*

Persistent Problems in Controversial Research

While all research is subject to limitations, my experience has been that controversial research appears to be vulnerable to more frequent and possibly more serious limitations. In some cases, I think the desire to score political points overrides normal scholarly caution. In other cases, I think there may be a tendency to "settle for" favorable results without much additional testing to attempt to disconfirm the desired results. In yet other cases, there are time pressures to get reports published and quality is sacrificed to increase the speed of pub-

lishing. Dr. Bruce Bell, a family researcher at the U.S. Army Research Institute, used to say you can have two of three – good, fast, or cheap. If you want "fast," then you have to spend more or accept lower quality research. There may be pressure from some grantors to produce findings to their liking. Regardless of the reasons, the next section will consider certain types of problems that seem to me to be more common to controversial research.

Confirmation Bias

Confirmation bias is when scholars "try to get the result they want" (Ioannidis, 2012, p. 650). Ioannidis (2012) noted that scholars can "continue adding and melding data, analyses, and subanalyses until something significant and publishable emerges" (p. 650). This may occur by examining and reworking data until a positive result is found in a sea of negatives or reworking data until a null result is found in a sea of positives. As Nickerson (1998) stated, "If one were to attempt to identify a single problematic aspect of human reasoning that deserves attention above all others, the confirmation bias would have to be among the candidates for consideration" (p. 175). Confirmation bias leads scholars to give more weight to evidence that supports their prior theory and less weight to evidence that disconfirms their prior theory. It appears to be a normal human tendency, once a stand is taken on an issue, to focus on "defending or justifying that position" (p. 177). Confirmation bias is related to *belief persistence*, in which "Once a belief or opinion has been formed, it can be very resistive to change, even in the face of fairly compelling evidence that it is wrong" (p. 187). Similarly, "many beliefs may be held with a strength or degree of certainty that exceeds what the evidence justifies" (p. 188).

As with beliefs, *theory persistence* is indicated in that "The history of science contains many examples of individual scientists tenaciously holding onto favored theories long after the evidence against them had become sufficiently strong to persuade others without the same vested interests to discard them" (p. 195). While one might hope that scientists might be less resistant to new ideas, Nickerson (1998) noted that scientists "often look much harder for evidence that is supportive of a hypothesis than for evidence that shows it to be false" (p. 207). Confirmation bias can become entangled with agendas regarding social justice (Allen, 2015, p. 165), where feelings about "the cause" may justify questionable research practices as long as the desired end is found and thereby justifies the means.

Refusal to Share Data

The Committee on Science, Engineering, and Public Policy of the National Academy of Sciences stated that "After publication, scientists expect that data and other research materials will be shared with qualified colleagues upon

request. Indeed, a number of federal agencies, journals, and professional societies have established policies requiring the sharing of research materials. Sometimes these materials are too voluminous, unwieldy, or costly to share freely and quickly. But in those fields in which sharing are possible, a scientist who is unwilling to share research materials with qualified colleagues runs the risk of not being trusted or respected. In a profession where so much depends on interpersonal interactions, the professional isolation that can follow a loss of trust can damage a scientist's work" (1995, p. 10). The American Psychological Association (APA, 2010) has stated that "Researchers must make their data available to the editor at any time during the review and publication process if questions arise with respect to the accuracy of the report. Refusal to do so can lead to rejection of the submitted manuscript without further consideration. In a similar vein, once an article is published researchers must make their data available to permit other qualified professionals to confirm their analyses and results (APA Ethics Code Standard 8.14a, Sharing Research Data for Verification). Authors are expected to retain raw data for a minimum of 5 yr. after publication of the research" (p. 12). In spite of such admonitions to make data available to other scholars, my experience and that of others (Wicherts, Borsboom, Kats, & Molenaar, 2006; Wicherts, Bakker, & Molenaar, 2011; Allen, 2015) has been mixed. Sometimes scholars, if asked, will share their data (Langbein & Yost, 2009; Regnerus, 2012a, b; Rosenfeld, 2014; Sullins, 2015 a, b, c) and some not (Brachman, Gold, Plotkin, Fekety, Werrin, & Ingraham, 1962; Golombok, Perry, Burston, Murray, Mooney-Somers, Stevens, *et al.*, 2003; Kaplan & Rosenmann, 2012). Ioannidis (2012, p. 650) has reported, as a major impediment to psychological research, the refusal of many scholars to make their data, analyses, or protocols available publicly.

At least in some cases, those who refuse to share data may be concerned that others would obtain "undesirable" results if allowed to investigate the raw data independently. Wicherts, *et al.* (2011) found that articles published by scholars who refused to share their data had more statistical problems or errors than those articles published by scholars who were willing to share their data. Sometimes one can re-engineer the original data using original reports (Brachman, Plotkin, Bumford, & Atchison, 1960; Plotkin, Brachman, Utell, Bumford, & Atchison, 1960; Brachman, *et al.*, 1962) as we did with early data on the anthrax vaccine (Schumm & Brenneman, 2004; Schumm, Brenneman, Arieli, Mayo-Theus, & Muhammad, 2004); sometimes, you can re-engineer enough of the data to improve your knowledge, but not perfectly (Schumm, 2014). If you cannot obtain the original data, you may have to try to replicate the study on your own, which at times may be impractical or too costly. At times, a data set may be available for a time publicly, but then may be retracted, as happened with the NESARC (National Epidemiological Survey of

Alcohol and Related Conditions) data, retracted by the U.S. government. Controversial data may be more difficult to obtain, overall, than less controversial data.

Resistance to Correction

As a new scholar (c. 1980), I assumed that most scientists would be eager to hear about mistakes they had made and eager to fix them. After all, my mentor, Dr. Charles Figley, had withdrawn an article accepted in the journal *Family Relations* when I reported to him that the main scale he had used as a one-dimensional measure was actually at least two-dimensional. Some graduate students feared I would be discharged from the Purdue graduate program, but nothing of the sort was even threatened. Later, we published three articles (Schumm, Figley, & Jurich, 1979; Schumm, Figley, & Fuhs, 1980, 1981) from that same data set, perhaps making up a bit for the withdrawn article. However, after I had published some general critiques of the literature (Schumm, Southerly, & Figley, 1980; Schumm, 1982, 1993) I began to realize that I had not made a lot of friends by noting that other scholars' work had serious flaws.

Not only may scholars themselves resist correction, but many journal editors will resist exposure of flaws in articles published under their watch. Some journals refuse to publish any critiques of their articles under any conditions. One journal refused to publish a critique on the intriguing "logic" that if the data in the original article were flawed, then so would be the critique, which led to my publishing that critique elsewhere (Schumm, 2003a). The editor did not retract the paper despite proven errors. Sometimes I have been criticized for only publishing critiques of the literature (Schumm, 1982; Schumm, 1993; Schumm, 2004g; Schumm, 2005a; Schumm, 2008; Schumm, 2010b, c, e, f; Schumm, 2011a, b; Schumm, 2012a, b, c; Schumm, 2013; Schumm, 2014), as if that activity were not "real" or valuable research. The watchdog role is very important, especially in areas of greater controversy; the field of economics even has its own online watchdog journal: *Econ Journal Watch*. When there is a lot of money or political power backing a given idea, then the watchdog role is even more important because of the pressures to conform to the "conventional answers" of the day, or the fear of not receiving adequate funding for research.

Bias Among Peer Reviewers

A reviewer of one of our articles (Schumm, Webb, Jurich, & Bollman, 2002) said that anything that implicated anthrax vaccine in long-term health problems of veterans would contradict government policy and create confusion in the field if published (p. 188). It is remarkable that in June 2002 the RAND Corporation had not released its Volume 3 of the Gulf War series, a volume on the role of immunizations in Gulf War illness. My research group was watching specifically to see how long it would take for that volume to be published (Schumm, *et al.*, 2002, p.

188), as at the time RAND had said it would be published within 1 yr. As of 2012, the vita of Dr. Beatrice Golomb *still* showed Volume 3 as “unreleased.” One gets the impression that if the U.S. Government does not want something published, it is not going to get published—ever. A vital and related point is whether peer review can be sound or effective if reviewers reject manuscripts solely because such manuscripts appear to contradict U.S. government policy, when the government will not allow research to be published no matter how much peer review might have supported publication. Certainly peer review is “blind,” but not necessarily to powerful policy makers or grant writers; would peer reviewers put their own research programs and reputations at risk to give unbiased reviews?

Another area where I have encountered reviewers who were illogical or biased concerns my attempts to statistically analyze religious behavior, including that related to Islam. I often received responses that reviewers did not like the idea of subjecting religion to statistical testing because, among other things, it meant that one might obtain results that differed from various religious tenets. My view is that *if* I could scientifically prove that a religion was incorrect in its history or doctrines, that would be a good thing, regardless of whatever religionists might think. Nevertheless, I have been able to publish a number of controversial studies (Faragallah & Schumm, 1996; Faragallah, Schumm, & Webb, 1997; Morgan-Miller, 2002; Schumm, 2002a; Schumm, 2003b, c, d; Schumm, 2004d, e, f; Schumm, 2005b; Schumm, 2006a, b; Schumm, Ferguson, Hashmat, & New, 2005; Schumm & Kohler, 2006). I have often encountered very harsh reviews, even for papers that merely pointed to the elephant in the living room; e.g., lesbian mothers having less stable relationships than heterosexual mothers (Schumm, 2010c). With reference to articles on same-sex parenting, Belcastro, Gramlich, Nicholson, Price, and Wilson (1993) concluded that “the system of manuscript review by peers, for minimum scientific standards of research, was compromised in several of these studies” (p. 117).

Literature Reviews May Ignore (Contrary) Findings

Literature reviews can reach sharply different conclusions in areas of controversy. For example, Ball (2003, p. 726) said he was not aware of any study done on the comparative stability of lesbian and heterosexual parents. Redding (2008) argued in his literature review that “lesbigay families are just as stable for childrearing as heterosexual families” (p. 164). Goldberg (2010, p. 26) argued that lesbian mothers might have *greater* stability than heterosexual parents. However, Kurdek (2005) reviewed the literature and stated that “The limited data available indicate that gay and lesbian couples may be less stable than married heterosexual couples” (p. 251). It is remarkable that good scholars can arrive at different conclusions in literature reviews when presumably they are all relying upon the same publications.

Biblarz and Stacey (2010) reviewed much of the literature on same-sex parenting, and among other things cited one article (MacCallum & Golombok, 2004) in concluding that lesbian mother families were less stable over time than heterosexual mother families (43% vs. 13% break-up rates over 6 yr., $p < .05$), although Tasker (2010) challenged the validity of their conclusion. There were other studies (Schumm, 2010c) that allowed comparisons of relationship stability as a function of parental sexual orientation, but these were not mentioned in Biblarz and Stacey’s review (Brewaeys, Ponjaert, van Hall, & Golombok, 1997; Chan, Brooks, Raboy, & Patterson, 1998; Fulcher, Chan, Raboy, & Patterson, 2002; Bos, Gartrell, Peyser, & van Balen, 2008; Gartrell & Bos, 2010). Fulcher, *et al.* (2002) and Chan, *et al.* (1998) found that 39% of lesbian mothers compared to 6% of heterosexual couples had broken up over 7 yr. ($p < .05$). Bos, *et al.* (2008) found that 48% of their lesbian couples had broken up over 10 yr., compared to 30% of a matched sample of heterosexual sisters ($p < .05$). Brewaeys, *et al.* (1997) reported a break-up rate over 5 yr. of 10% for lesbian mothers, compared to 4% for heterosexual mothers (ns). More recently, Rosenfeld (2014) found that of the married parents in his study, 25% of same-sex couples compared to 7.8% of heterosexual parents had broken up over 4 yr. (Table 1). Overall, break-up rates appeared to be about 15–20% for heterosexual parents over approximately 10 yr. compared to 40–45% for lesbian parents (Schumm, 2010c, pp. 504–505). Several other studies have reported stability rates for lesbian couples, even though they did not report rates for heterosexual couples. Stevens, Perry, Burston, Golombok, and Golding (2003) found 40% of one lesbian group of parents had separated by age 7 of their child, with 61% of a different group having separated within 4 yr. Tasker and Golombok (1997) found that as many as 75% of their lesbian parents had broken up by the time their child was 15-yr.-old. Kivalanka and Goldberg (2009) found that as many as 53% of lesbian parents had broken up by the time their child had become an adolescent. Golombok, Tasker, and Murray (1997) found that one-third of their lesbian mothers had broken up by the time their child was 6-yr.-old. There is a great deal of empirical data on the comparative stability of lesbian and heterosexual parents, but several major literature reviews (Ball, 2003; Redding, 2008; Biblarz & Stacey, 2010; Goldberg, 2010) did not explore the extent of these data or the direction of the effect.

In another recent example of an attempted comprehensive literature review, Fedewa, Black, and Ahn (2015) did a meta-analysis of *four* studies and concluded that parental and child sexual orientation were unrelated with $d = -0.06$, although they did report a significant result ($d = -0.53$, $p < .05$) regarding sexual (orientation) questioning by daughters of same-sex vs. heterosexual parents, on the basis of a single study (Bos & Sand-

TABLE 1

Rosenfeld's (2014) results examined in detail predicting relationship instability over 4 yr. from sexual orientation, parental status, and marital status including subgroups based on duration of romantic relationship

Parental Status	Group	Married		Unmarried	
		Same-Sex % (n)	Heterosexual % (n)	Same-sex % (n)	Heterosexual % (n)
Non-parents	All	11.9 (92)	3.9 (926)	29.6 (277)	41.7 (424)
	≤ 5 yr.	42.9 (14)	13.8 (80)	53.1 (98)	47.7 (262)
	≤ 10 yr.	25.8 (31)	10.9 (229)	42.9 (147)	44.2 (337)
	>10 yr.	4.9 (61)	1.6 (677)	14.6 (130)	32.2 (87)
	≤ 20 yr. and age ≤ 45	20.0 (30)	7.9 (315)	41.6 (113)	42.2 (255)
Parents	All	25.0 (4)	7.8 (488)	52.2 (23)	43.2 (118)
	≤ 5 yr.	100.0 (1)	27.6 (29)	64.3 (14)	49.4 (79)
	≤ 10 yr.	33.3 (3)	12.4 (129)	64.7 (17)	44.6 (92)
	>10 yr.	0.0 (1)	6.1 (359)	16.7 (6)	38.5 (26)
	≤ 20 yr. and age ≤ 45	50.0 (2)	8.9 (381)	44.4 (18)	43.8 (96)

fort, 2010). However, in my literature review of this topic (Schumm, 2013) I found nearly 10 times as many studies ($k=38$), and the meta-analysis did indicate a significant relationship.³ Furthermore, when Crowl, Ahn, and Baker (2008) did a similar meta-analysis 7 yr. earlier, they used more ($k=5$) studies on parental and child sexual orientation, yielding $d=0.20$, a small to moderate effect. It is not clear how a literature review by the same author(s) could find such stark differences using, presumably, the same studies, and yet overlook dozens of other relevant studies.

Belcastro, *et al.* (1993) noted in their analysis of the same-sex parenting literature that “most were biased towards proving homosexual parents were fit parents” (p. 117) and that “A disturbing revelation was that some of the published works had to disregard their own results in order to conclude that homosexual parents were fit parents” (p. 117). As early as 2005, I pointed out the large number of reviews of the literature that had concluded, against the evidence, that there were no differences among children raised by same-sex parents compared to heterosexual parents (Schumm, 2005a, 2008). One example of a such a review is Fitzgerald (1999), who concluded in her abstract that “The body of literature generally concludes that children with lesbian and gay parents are developing psychologically, intellectually, behaviorally, and emotionally in positive directions, and that the sexual orientation of parents is not an effective or important predictor of successful child

development” (p. 57). Yet within this article, as noted before (Schumm, 2004g, p. 422; 2005a, p. 437), she stated that “In summary, faced with these frequent methodological difficulties, the generalizability of these studies is limited and overall, they can best be described as descriptive and suggestive, rather than conclusive” (p. 69). While generalizability is critical to translating research into policy, Fitzgerald also concluded that “the legal community must not support policies of outright denial of rights to such things as adoption, foster parenting, reproductive technology, or retention of custody, simply on the basis of sexual orientation” (p. 70), which suggests not only generalization of the findings (in spite of the limitations), but that the generalization be used in a socially potent manner to change social policy and law.

As another example of a literature review with at least one important limitation, Fedewa, *et al.* (2015) cited Hawkins (2011) among the studies reviewed, but when I ran a meta-analysis across the four tests available (parent sexual orientation by two outcomes, parent-adolescent relationship quality and adolescent behavior problems) for Hawkins' research, I obtained an average Hedge's g of 0.323 with $z=3.06$ ($p=.002$). Yet I did not find an indication in Fedewa, *et al.* (2015) that Hawkins' results were statistically significant. This omission makes me question such literature reviews. Furthermore, in Hawkins' study the gay ($d=0.35$) and lesbian ($d=0.37$) parents scored higher on couple relationship quality than the heterosexual parents, which would be expected to lead to better parental outcomes than for the heterosexual parents, although the reverse actually occurred. Furthermore, Fedewa, *et al.* (2015) reported effect sizes for both lesbian mothers and gay fathers with respect to child outcomes; however, the number of effect sizes from the literature ranged between one and four for gay fathers but as high as 38 for lesbian mothers (pp. 20–24). While Fedewa, *et al.* stated that “The literature on gay fathers independent of

³When a formal meta-analysis was performed on data from several studies that had provided rates for children from both heterosexual and same-sex parent families (Gottman, 1989; Huggins, 1989; Javaid, 1993; Sirota, 1997; Kunin, 1998; Canning, 2005; Murray & McClintock, 2005; Rivers, Poteat, & Noret, 2008; Schumm, 2008; Regnerus, 2012a, b; Swank, Woodford, & Lim, 2013), an overall odds ratio of 3.12 (95% CI=2.53 to 3.83, $p<.001$) was obtained, suggesting that the odds that children from same-sex parent families would grow up to identify as LGB or to engage in same-sex sexual behavior were three times greater than for children of heterosexual parents.

lesbian mothers is limited" (p. 30), that did not moderate their conclusion that "children with lesbian or gay parents had higher levels of psychological well-being than children with heterosexual parents" (p. 28).

It is also common for literature reviews to reveal bias in the way that they inaccurately cite previous research findings. For example, Hooker (1957, 1958) found statistically significant differences between gay men and heterosexual men in Rorschach test results; however, many of the subsequent reviews of her research have concluded the opposite, that she had not found any significant results (Schumm, 2012c). Hooker did not misrepresent her research in these reviews; the misrepresentation was from the authors of the reviews. Research by Hosking, Mulholland, and Baird (2015) represents another example of this problem. They cite Sarantakos (1996a) as supporting a conclusion that "the positive health and social outcomes experienced by the children of gay and lesbian parents, which have been mapped out in quantitative and qualitative social science research (citing Sarantakos, among others) are regularly offered as proof that nonheterosexual parenting is not detrimental to the children" (p. 328). Sarantakos (1996a) found effect sizes as large as 3.75 between children of heterosexual married and same-sex parents, an effect size clearly substantially larger than zero or no harm. Obviously, although mistakes can occur, it is unethical to cite research as having shown no significant effects when there were significant effects.

To summarize my comments with respect to bias in literature reviews, it does not make sense to say on the one hand that research is of low quality and cannot be generalized or that we have very little data, and then, in interpretations, contrarily assume the research is methodologically sound enough to justify legal or social policy changes. It would be one thing to argue for legal or policy changes on the basis of their "rightness" regardless of the research (e.g., Constitutionality), but to use low-quality research to justify legal or policy changes is not a good idea.

The point is that in controversial areas of research literature reviews themselves may be biased against findings that might be contrary to the politically correct or theoretically accepted view – and literature reviews may even contradict themselves; e.g., Crowl, *et al.* (2008) compared to Fedewa, *et al.* (2015) in terms of the number of studies assessing the sexual orientation of children of gay or lesbian parents, with one reporting $d=0.20$ and another $d=-0.06$. Examples of excellent literature reviews about same-sex parenting in top tier journals that attempted to carefully evaluate each study under consideration include Allen (2015) and Marks (2012), who each reviewed over 50 studies for sample size, statistics used, any reports of statistical power or effect sizes, outcomes studied, any comparison groups used, whether the samples were random or not random, if any longitudinal data were available, and the gender of the parents.

Biased Citation and "Refusal to Cite"

Published literature is cited to support hypotheses or research questions. Citations of papers in articles may thus reflect authors' biases in formulating their research goals. Below is an interesting example showing that citation rates may reflect authors' biases rather than the quality of a cited article. It is also a way for groups of citing authors to "bury" contrary opinions or results. Some researchers argue that more highly cited articles are more credible or of higher quality. However, frequent citations may only reflect a general political or social bias and not methodological quality, as indicated in the following examples.

An interesting example is three articles published from the same data set (Mucklow & Phelan, 1979; Miller, Mucklow, Jacobsen, & Bigner, 1980; Miller, Jacobsen, & Bigner, 1981) by the same authors; two of these articles appeared in the same journal (Mucklow & Phelan, 1979; Miller, *et al.*, 1980). The two papers discussed results that appeared to be favorable to lesbian mothers (Mucklow & Phelan, 1979; Miller, *et al.*, 1981) and were cited (by Google Scholar as of December 18, 2015) 72 and 94 times, respectively—over eighteen times (166 to 9) more often than the one paper (Miller, *et al.*, 1980, cited 9 times) reporting results "not favorable" to lesbian mothers (Schumm, 2010d), despite the fact that the latter results were later replicated in part by Dundas and Kaufman (2000), so they were likely robust. Sometimes this non-citation bias or "refusal to cite" is even more direct: in their literature review, Fedewa, *et al.* (2015) could have cited any one of these three papers but cited only Mucklow and Phelan (1979), which happened to report more favorable results for lesbian families (it is not clear why they did not also cite Miller, *et al.*, 1981, except perhaps to avoid duplication).

Golombok's (2015) recent summary of research on same-sex families is described as follows: "*Modern Families* brings together research on parenting and child development in new family forms including lesbian mother families, gay father families.... The findings not only contest popular myths and assumptions about the social and psychological consequences for children of being raised in new family forms, but also challenge well-established theories of child development that are founded upon the supremacy of the traditional family." One might ask whether a text with such a comprehensive intent would cite any contrary evidence. With one exception (Allen, 2013), no research or literature reviews supporting a more conservative social perspective (e.g., papers by Schumm, Cameron, Marks, Sullins, or Regnerus) were cited.

Ideology may be Accepted over Scientific Evidence

An older, but clear example of ideological reasoning is found in Hooker (1957), who argued that it would only take one case among a sample of 30 homosexuals to prove that homosexuality was not "necessarily a symptom of pathology" (p. 30). Statistically, in the context of her experiment which compared 30 heterosexual men

with 30 homosexual men on mental health adjustment, that criterion would be satisfied if 1/30 homosexual men were mentally well-adjusted compared to 29/30 heterosexual men ($r=0.93$, $d>4.0$; Cohen, 1988, p. 22). In this case, there would have been an effect size difference larger than 4.0, eight times greater than the effect Cohen (1992) believed would be noticed by a careful observer; nevertheless, Hooker would have considered her hypothesis of no differences between the two groups of men to be supported. As a statistician, it is clear that if an effect size of 4.0 is considered not meaningful, then ideology will be allowed to trump empirical results.

In another example, Patterson (2005) stated that "Not a single study has found children of lesbian or gay parents to be disadvantaged in any significant respect relative to children of heterosexual parents" (p. 15). Yet, she discussed Sarantakos (1996a) on pages 6–7 and also Puryear's (1983) dissertation (p. 39). The former paper found numerous differences in educational outcomes, with substantial effect sizes (as large as 3.75) between outcomes for children of same-sex vs. heterosexual parents (Schumm, 2015c), while the latter dissertation reported significantly and substantially lower family togetherness in drawings by children of same-sex parents relative to children of heterosexual parents. Furthermore, Patterson (2005) cited two studies (Mucklow & Phelan, 1979; Miller, *et al.*, 1981) by the same authors using the same data that featured results favorable to lesbian families but omitted in her bibliography one study from the same authors (Miller, *et al.*, 1980) that had featured adverse results for lesbian families. Patterson (2005) also did not cite any of the additional research done by Sarantakos (1996b, 1998, 2000) on same-sex families, as described in more detail elsewhere (Schumm, 2015c). Other reviews of the literature have overlooked the breadth of Sarantakos's research on same-sex families (Marks, 2012; Herek, 2014; Allen, 2015).

De Facto Censorship

Some years ago, I was invited by a journal editor to participate in a debate about same-sex marriage. I agreed and provided a brief paper (Schumm, 2009) that was an extension of previous discussions about same-sex marriage and parenting (Schumm, 2004c, g). Several other authors did so as well. The pro-gay marriage side was represented, but the materials were reprints of previously published articles. A number of persons vehemently disagreed with the concept of having a debate about same-sex marriage because having a debate implied that no one side was absolutely right, and in their view one side *was* absolutely right. The end result was that the articles opposed to same-sex marriage were withdrawn/canceled and the entire project canceled. A similar situation occurred in which I was invited to a conference in Los Angeles to partake in a debate on same-sex marriage and parenting with the incentive that each contributor would be allowed to publish their

presentation in a university law review from that institution. It appeared that the progressive scholars realized that if this promise were kept, then there would be as many "anti-gay marriage" articles as pro-gay marriage articles, and so the entire idea was abandoned.

Back in 1996, when working at the Army Research Institute as a summer fellow, I had discovered that as many as 70% of female sergeants had experienced a divorce within a few months after their deployments to the Middle East for the first Persian Gulf War, a percentage found in both a sample of active component personnel and in a sample of reserve component personnel. However interesting such a finding may have been or however useful it may have been, publication of that result was forbidden by the government. Only years later was it partially published, which I dared to attempt thinking that the leaders who had resisted publication were no longer involved in military family research; i.e., in a position to hurt us for revealing the evidence we had found (Schumm, Nazarinia Roy, & Theodore, 2012). I have heard of other scholars who found things the government did not want "out" and they came to work only to find that their passwords had been changed and they were not given the passwords needed to continue to run statistical analyses or to prepare publications from their research.

Not only can the government be upset if you find a negative that they don't want, it can be a problem to not find a positive it does want. Along these lines, I was involved in a program evaluation involving nutrition in which there appeared to be no significant effects of the program, no matter what outcomes measures or subgroups of program clients were examined; but that answer was not acceptable to the officials in Washington, DC, who were overseeing the project, and, without my participation, the research team continued to crunch numbers until they found something that could be presented to prove the success of that very expensive taxpayer-supported program. In other words there was a censorship of unwanted null results, because the *null* results would show that millions of federal dollars were being wasted or mismanaged.

Sometimes one can fight back against government assumptions that implicitly censor activity with social science research. The U.S. military appears to have assumed in practice that staff rides (learning tours of old battlefields) are reserved for senior officials rather than for lower ranking enlisted or junior officers because either they will not be useful for junior personnel or are too expensive to be provided for all military personnel. Schumm, Turek, and McCarthy (2003) conducted and evaluated a staff ride for an entire Army Reserve unit to the Civil War battlefield at Lexington, Missouri, with a focus on providing enlisted personnel and junior officers an experience normally reserved for non-commissioned or commissioned officers at senior military schools. Publishing the evaluation of that staff ride "for

all" was one way to challenge the assumption that staff rides would not work for less experienced personnel or would be too expensive to justify as part of a Reserve or National Guard unit's weekend training. Although staff rides for all ranks of military personnel have not become standard practice, anyone who wishes to look up if a staff ride "for all" has ever been tried can now do so—evidence that was not available before.

Inequitable Criticism of Methods to "Discredit" Contrary Findings

As compared to a "refusal to cite," an author may cite minimally for the purpose of unfairly or unequally criticizing work that does not agree with their assumptions. Herek (2014) cited Allen, Regnerus, and Sarantakos but primarily to highlight the alleged limitations and legal irrelevance of their results. Herek (2006) has emphasized "the importance of examining the entire body of research rather than drawing conclusions from one or a few studies" (p. 610), but chose to focus on only *one* article by Sarantakos in his critique rather than examining Sarantakos's "entire body of research," which is extensive. Likewise, Manning, Fetto, and Lamidi (2014) cited Allen, *et al.* (2013) and Regnerus (2012a, b), but not Sarantakos (1996a), primarily in an attempt to discredit their results.

Different standards can be used for evaluating research. At a trial in Florida in 2008, the expert witnesses for the other side decried my reporting of results where the observed significance level was $p=.10$, although this was meant to parallel what the researchers had done themselves (Tasker & Golombok, 1997, p. 47). Later, a close examination of how my critics had reported results found that they had often done the same thing themselves, reporting results for $p=.10$ (Schumm, Pratt, Hartenstein, Jenkins, & Johnson, 2013). As noted elsewhere (Schumm, 2012b, 2014), Regnerus (2012a, b) received intense criticism for his research on same-sex families because he included in his analyses families in which children had only lived with a same-sex parent for a few years rather than their entire lives. However, other scholars (Golombok, *et al.*, 2003) had done exactly the same thing, yet never received this heated criticism, as discussed in more detail elsewhere (Schumm, 2014).

A similar situation occurred with respect to the U.S. government's citation of research on Gulf War illnesses related to exposures from the first Persian Gulf War (e.g., vaccines, pyridostigmine bromide tablets, insecticides, insect repellants) – such citations were very infrequent. For example, The Institute of Medicine of the National Academy of Sciences, with support from Contract V101(93)P-2155 from the Department of Veterans Affairs, produced a text on the effects on health of serving in the first Persian Gulf War (Institute of Medicine, 2006). Although the Ohio Desert Storm Study had produced numerous publications on factors related to Gulf War illnesses, including vaccines (Schumm, Reppert, Jurich, Bollman,

Webb, Castelo, *et al.*, 2002; Schumm, Webb, Jurich, & Bollman, 2002; Schumm, 2004b; Schumm & Brenneman, 2004; Schumm, Brenneman, Arieli, Mayo-Theus, & Muhammad, 2004; Schumm, 2005c; Schumm, Jurich, Bollman, Webb, & Castelo, 2005; Schumm, Jurich, Webb, Bollman, Reppert, & Castelo, 2005a; Schumm, 2006c; Schumm & Nass, 2006), nerve agent tablets (Schumm, Reppert, Jurich, Bollman, Castelo, Sanders, *et al.*, 2001; Schumm, Reppert, Jurich, Bollman, Webb, Castelo, *et al.*, 2002), and exposure to nerve agents or other factors (Schumm, Webb, Bollman, Jurich, Reppert, Castelo, *et al.*, 2004; Schumm, Jurich, Webb, Bollman, Reppert, & Castelo, 2005b; Schumm, Jurich, Webb, Bollman, Reppert, Sanders, & Castelo, 2007) none were cited in the IOM (2006) review of the literature on Gulf War health, though Golomb (2008) cited some of them in a later report on Gulf War illness and toxins. Furthermore, our reports presaged the conclusion of the Research Advisory Committee on Gulf War Veterans' Illnesses (2014) that "chemical exposures, not psychological stressors or psychiatric disorders, are the cause of Gulf War illness and other health and functional disorders in Gulf War veterans" (p. 24). Later in the same report, it was stated that "The evidence is particularly compelling for pesticides and pyridostigmine bromide" as well as exposure to "nerve gas agents" (p. 38) as factors implicated in the development of Gulf War illnesses, all factors for which our research found early evidence. In other words, although our research was among some of the first to identify some of the factors now believed to be associated with Gulf War illness, one would be hard pressed to find official government documents with citations to that effect.

Furthermore, the RAND corporation produced a number of reports on possible causes of Gulf War illnesses (RAND, 2005⁴). When we contacted RAND in June 2002 (Schumm, Webb, Jurich, *et al.*, 2002, p. 188, footnote 2), we were told that the report on immunizations (vaccines, volume 3, being prepared by Dr. Beatrice Golomb) would be published by October 15, 2002. In 2005, RAND estimated that the immunization report would be published that year, stating "RAND additionally investigated the efficacy of the anthrax and botulinum toxoid vaccines and reviewed the history of anthrax vaccine production. The results of that review will appear in a forthcoming RAND report, which is expected to be published in 2005" (RAND, 2005, p. 2). However, that RAND report, Volume 3 of the set of reports on Gulf War illnesses, has not yet been published as of 2015. The Research Advisory Committee on Gulf War Veterans' Illnesses (2014) discussed comments by Dr. Golomb, who was described in the report as follows: "Dr. Golomb commented on the vaccine sentence. She thought that the sentence should read that current results have been conflicting and have shown weak as-

⁴See www.rand.org/content/dam/rand/pubs/research_briefs/2005/RAND_RB7544.pdf

sociations because these studies have not adjusted for other exposures" (p. 11). Thus, although the role of vaccines in Gulf War illnesses has remained at least an open question, the RAND report on immunizations has never been published despite the fact that scholars have been waiting for it for over 13 years.

Misleading Statements

Statistical statements can lead the unsuspecting to draw incorrect conclusions. For example, Herek (2014) argued that "the available empirical data are consistent with the conclusion that the vast majority of those children [of same-sex parents] grow up to be heterosexual" (p. 593). Likewise, Patterson (2013, also see 2009a) stated that "Overall, the clearest conclusion from these and related studies is that the great majority of children from lesbian or gay parents grow up to identify as heterosexual" (p. 31). Such statements sound as if they are claiming there are no differences in children's sexual orientation as a function of parental sexual orientation, but technically they are not; rather, they are saying that at least 50% and perhaps more of the children of same-sex parents identify as heterosexual eventually at some point in their adulthood. Such statements obscure a reality in which children from same-sex parents are much more likely to identify as non-heterosexual or to experiment with same-sex romantic relationships at some point in their social development, as has been detailed elsewhere (Schumm, 2013). Allen (2015) has noted the same phenomenon, stating that "the conclusion that 'the large majority of sons of gay fathers are heterosexual' is hardly noteworthy" (p. 172) given that no one would expect that paternal sexual orientation would *determine* a child's sexual orientation.

Another common practice is to discuss a paper with which you disagree and speculate about its possible flaws without hard evidence. For example, Herek (2014) argued that stigma and prior divorce might account for Sarantakos's (1996a) findings about same-sex parenting. It is true, they "might." However, Herek had no evidence of this, and his argument failed: in at least one area, the children of same-sex parents did *better* than those of heterosexual parents, which is difficult to explain if teacher and community bias was entirely to blame for the *less* favorable results. Furthermore, Hawkins (2011) found that same-sex parents rated their children higher on conduct problems than did heterosexual parents, even though the same-sex parents rated their own couple relationships as more satisfactory than did heterosexual parents, an example that does not fit Herek's apparent assumptions.

Perhaps another issue is the use of numerous control variables. Adding new variables to any regression model, if those variables are not important, can bias the results of tests of significance for other variables that are important. The coefficients or effect sizes of those variables may not be biased, but the statistical tests can become biased. It is

essential that any new variables added have a strong theoretical basis for their inclusion, especially if the objective of the researchers is to "prove" the null hypothesis regarding variables already in the model. Virtually any independent variable can be reduced to statistical non-significance if one includes enough unimportant new variables in the overall model. Thus, it is easy to say something like "after we added several new control variables, variable X no longer significantly predicted Y." That sounds profound, like a new discovery, but it is merely what can happen when you add any new variables to any larger model. The burden should be on the scholar who adds the new variables to justify their selection and their placement in the model (e.g., perhaps they would better serve as mediating or moderating variables than as control variables). If the goal is to "prove" the null hypothesis, then the addition of suppressor, distorter, or mediating variables should be considered, as well as the addition of control variables (Schumm & Crawford, in press).

Violations of Human Rights or Scientific Fraud

Although this may be less common than some of the other concerns, one may encounter a violation of the rights of human subjects, especially in older studies. I will discuss two sets of studies with respect to these issues.

First, in May 1957, a test of a human anthrax vaccine began at the Arms Mill in Manchester, New Hampshire. By August, the treatment group had received three active inoculations and the placebo group its control inoculations. As early as the 27th of August, workers began to fall ill of anthrax and three had died by September 8th. A fourth worker died on November 4th. None of these workers were diagnosed to the public with anthrax and some never received the proper treatment for their anthrax infections and died as a consequence, often without knowing they had been exposed to anthrax infection. Yet, as noted previously (Schumm, 2005d, p. 344), the senior author (Brachman, *et al.*, 1962) stated that "When possible cases of cutaneous or inhalation anthrax were reported among the employees, I was immediately notified and I flew to the mill in order to confirm the diagnosis" (Brachman, 2005). Despite its speed, that protocol did not do enough in time to keep several workers from dying of anthrax infections.

Between the start of the vaccinations and the start of the epidemic only 3–4 mo. occurred at most. Yet, it was stated that the epidemic "presented an unusual opportunity to study both the epidemiology of this disease and the effectiveness of an anthrax vaccine which had been given to some of the workers several months before the epidemic" (Brachman, Plotkin, Bumford, & Atchison, 1960, p. 6). Unless the mill workers had received all three injections, a full test of the vaccine would not have been possible. Yet, the time between receiving the last injection and the start of the epidemic was likely less than a month or two, not several months. My guess is that the

government or someone affiliated with the vaccine test deliberately infected the mill workers in order to help test the vaccine. However, the bottom line is that the workers were allowed to die without informed consent as to the nature of the trials and without adequate treatment, which I believe were severe (and lethal) violations of their human rights (Schumm, 2005d). Hopefully, today, institutional review boards would never allow such practices to occur in medical or social science research.

Ioannidis (2012) has stated that “questionable research practices are probably very common” and that “Occasionally, results may even be entirely fabricated” (p. 650). John, Loewenstein, and Prelec (2012) have discussed the relatively high frequency of questionable research practices (QRPs) admitted by social scientists. Stroebe, Postmes, and Spears (2012) have enumerated a number of exposed cases of scientific fraud and indicate that top tier journals are more likely to have accepted articles later proven fraudulent than are lower tier journals, perhaps the opposite of what would be expected if the screening procedures of top tier journals were actually more effective than those of lower tier journals.

A possible example of scientific fraud would be the early studies by Hooker (1957, 1958). At the very least, she did find statistically significant differences between homosexual and heterosexual men between her two groups (although each group also included bisexual men) and the two groups were significantly different *a priori* on key demographic variables, which could have biased the judges who rated the men (Schumm, 2012c). Her research has been incorrectly cited by many scholars as having proven that there were no differences between the two groups of men. Furthermore, Cameron and Cameron (2012) challenged my critique as too soft, claiming that Hooker's research was fraudulent, not merely misrepresented by later reviewers. I do not think that Cameron and Cameron (2012) were able to prove fraud, but they raised a number of important points. To date, no one has challenged their assessment by submitting a rebuttal.

Fraud in controversial areas is possible. An article published in the prestigious journal *Science* (LaCour & Green, 2014) was recently exposed as having been based, apparently, on fabricated data. The study had received funding from the Williams Institute (p. 1369), some of whose scholars had demanded retraction of Regnerus (2012a, b) because of a disagreement over measurement technique of family types. Yet, fabrication of data and reporting of detailed statistical analyses and charts, as done in LaCour and Green (2014), is surely far more worthy of retraction (McNutt, 2015a, b) than were any of Regnerus's possible errors.

Common Methodological Problems in Controversial Research

Above, I have discussed some of the process issues that may bias social science. Next, I will address at least ten

of the more common methodological problems I have encountered over the years in dealing with research on controversial issues. It can be argued that scientific peer review should have detected many of these problems and required their correction prior to publication, but my focus here is not the issue of ineffective peer review.

Disparities Between Theory and Analysis

It is particularly revealing when authors report nonlinear patterns (e.g., quadratic) but analyze their data with linear statistical models. It is relatively easy to evaluate nonlinear (i.e., quadratic, cubic, etc.) trends in data, so lack of capability cannot be blamed. For example, Price (2002) dealt with the controversial issue of the relationship between degree of Islamization in nations and their respect for human rights. He stated in his article that the results appeared to be curvilinear, but he analyzed his data using linear correlations only (Schumm, 2003a). More recently, I noted that an article (Stanger-Hall & Hall, 2011) dealing with the controversial area of sex education that had reported nonlinear patterns but the analysis was only in terms of linear statistics. In some cases, scholars have used difference scores when interaction effects might have been more appropriate for testing (Schumm & Kirn, 1982). In controversial research it may be more common to find a mismatch between expressed nonlinear theory and the linear statistics used to test that nonlinear theory.

Falsehoods Uncritically Accepted

Recently, I noticed that the figure “six to fourteen million” children of same-sex parents in the USA being mentioned frequently in the literature in both scholarly and legal articles (Schumm & Crawford, in press). In fact, some courts appear to have accepted this as a fact. Recent estimates are on the order of thirty to seventy times lower (i.e., 200,000 vs. 6 to 14 million). How did this error occur? Our investigation (Schumm & Crawford, in press) found that the fourteen million number was from a USA Today article in 1984 (Peterson) that was picked up in four chapters by Bozett (1987a, b; 1993) and Green and Bozett (1991), as well as in an article in *Child Development* by Patterson (1992). It was interesting to me that the citations in all of those five sources were incorrect (citing page 3 or page 30 rather than the actual page number, 3-D). The newspaper article gave no source or basis for its estimate. However, once it was picked up, even though with an incorrect citation, in scholarly books and journal articles, it became an unquestioned fact, cited at least 65 times, even as recently as 2013 by a social scientist (Raley). One scholarly example includes Dundas and Kaufman (2000), who indicated that “gay and lesbian parents in the United States [were] raising between 6 and 14 million children” (p. 66). The figure was cited in numerous law reviews (e.g., Strasser, 2011) and in some same-sex marriage court cases as a

fact (Schumm and Crawford, in press). Moreover, some cited the figure as a low estimate (Ahmann, 1999) and others as pertaining only to adopted or foster children (Mabry, 2005), the estimated total number of all children, therefore, being much greater. A similar incorrect estimate of 8–10 million children of gay and lesbian parents began with minutes from an American Bar Association meeting (Bureau of National Affairs, 1987) that was picked up by the Editors of the *Harvard Law Review* (1989, p. 1629; 1990, p. 119).

Another example of a falsehood accepted widely is the idea that same-sex attracted youth or the children of same-sex parents are bullied more than heterosexual children or the children of heterosexuals. However, Rivers, *et al.* (2008) compared children of same-sex and opposite-sex parents and found that the *latter* were victimized more often ($d=0.28$). Sullins (2015c), in a much larger sample of families, also found that the children of heterosexuals were more often bullied than were the children of same-sex parents. Rivers and Noret (2008) compared same-sex and opposite-sex attracted students in Britain and found that while same-sex attracted students were slightly more likely to be victimized ($d=0.11$), they were more likely to perpetrate bullying as well ($d=0.22$). Space precludes a more comprehensive analysis of bullying, but at least some studies have found that bullying patterns do not fit conventional expectations. Sometimes research does not take perpetration into account, but Robinson, Espelage, and Rivers (2013), though they found higher rates of victimization among younger LGB youth, found that bullying decreased among both heterosexual and LGB youth, especially girls, over time and that LGB status predicted emotional distress above and beyond any effects of bullying or prior emotional distress – a more complex pattern than might have been anticipated.

Poorly Described, Non-blinded, or Biased Samples

Sometimes something as simple as how many people participated in a study can be uncertain. During our examination of research on smallpox vaccine, we determined that different articles in top-tier medical journals reported varying sample sizes for the same study (Schumm, Nazarinia, & Bosch, 2009). In earlier research on anthrax, Brachman, *et al.* (1962) may have conducted only a single-blinded study, as the lead researchers did vaccinations of the workers and attempted to respond to sickened workers directly. Furthermore, in early reports Brachman and his colleagues (Brachman, *et al.*, 1960; Norman, Ray, Brachman, Plotkin, & Pagano, 1960; Plotkin, *et al.*, 1960) stated that there were 150 participants in each arm of the study, but later they changed the numbers to 149 and 164 (Brachman, *et al.*, 1962).

In studies of homosexuality, Hooker (1957, 1958) personally invited men to participate in her studies and may have informed some of them of the purposes of the research, meaning that her study was not double-

blinded. In addition, both of her groups included 10% bisexuals, so her comparisons were not between strictly gay men vs. strictly heterosexual men. Golombok, *et al.* (2003, p. 22) could not obtain enough cases (only 18) of lesbian families from a large random sample (nearly 14,000) from Avon in southwestern Britain, so they supplemented their data with convenience sampling; furthermore, their sample of lesbian families included between three and fifteen (of 28) cases in which the focal child had spent more time outside of a lesbian family than in one, meaning that the sample was not of children that had been with a lesbian mother or family since birth—a problem Regnerus encountered and for which he (but not Golombok, *et al.*) received much criticism (Schumm, 2012b, 2013, 2014).

Farr, Forssell, and Patterson (2010) obtained samples of same-sex and heterosexual parents, but the response rate for the former was 75% compared to 41% for the latter, raising the spectre of selection bias. Furthermore, average household incomes were over \$150,000, suggesting that the samples were not representative of average U.S. parents. Finally, it was stated that “All parents noted that some individual... provided outside care for their child on a regular basis” (p. 168). Thus, in one study alone, we find that there was probably selection bias in the sampling, bias in terms of socioeconomic status, and some question as to what extent the child was being raised by its parents vs. other caretakers. Furthermore, the caretakers were asked to rate the child, which could have led to bias since many of the caretakers were in the financial employ of the parents. Sullins (2015a) also found sample bias problems with respect to research on same-sex parenting.

These are only a few examples of how we may not really know who the participants in a study were or how the family's characteristics changed over time. In some cases, sample bias can lead to relatively weak independent variables, as will be discussed more later. It might seem simple to compare same-sex parents with heterosexual parents, but often the samples are not pure and may be biased in ways that could distort the results. Some researchers may try to absolve themselves of such issues by saying it is difficult to find pure samples, but comparing samples that are not pure family types is an important issue, as Regnerus discovered from his many critics (Schumm, 2012b, 2013, 2014).

Furthermore, if samples are different in terms of socioeconomic status, parental education, number of children in the families, age of parents, duration of the parental relationship, region of the nation, religiosity, hours employed per week, or other variables, then those differences should be controlled statistically before drawing conclusions about parental sexual orientation or any other single variable (Schumm, 2010f). Biased samples are problematic regardless of whether the results seem to favor or not favor same-sex parenting or any other

outcome (Allen, 2015). As Wilkinson and the Task Force on Statistical Inference (1999) noted, non-random samples are fraught with risk of bias and that “there are no valid excuses, financial or otherwise, for avoiding an opportunity to double-blind” (p. 596). Allen (2015) has observed that “A proper probability sample is a *necessary condition* for making a general claim about an unknown population, based on a sample” (p. 159).

Omission of Important Dependent Variables

Langbein and Yost (2009) tested five possible outcomes of legalization of same-sex marriage; finding no relationship between some sort of same-sex legal unions in a state and those selected outcomes, they deemed that there were no “negative externalities” associated with same-sex marriage. However, proving that five variables are not significantly related to a predictor variable does not mean that there are no other variables yet to be found significant; that is an improper generalization. When I analyzed fertility rates as a function of years since same-sex marriage, I did find a significant direct and a significant indirect effect of same-sex marriage on fertility (Schumm, 2015a), for example.

Sometimes, significant outcomes may only be detected over time in longitudinal studies: the missing outcome variable is the post-treatment effect. When I was studying divorce as a function of overseas deployments, it was widely regarded that there was only an artificial relationship due to a backlog of divorces caused by the inability of service members to get divorced while deployed. Yet our research, particularly our longitudinal research, did find significant relationships between deployments and divorce (Schumm, Bell, & Gade, 2000; Schumm, *et al.*, 2012). Similarly, few same-sex parenting studies have assessed change over time in child outcomes across different types of families. In one of the few truly longitudinal studies, Lavner, Waterman, and Peplau (2012) reported that outcomes were similar after two years for children of same-sex and heterosexual parents, a fact that obscured the longitudinal reality that problems had increased slightly for children of same-sex parents but had declined moderately for children of heterosexual parents over the 2-yr. period. In other words, their study revealed that a one-time status report on outcomes for children might not provide as much information about changes in children’s development as a report that took account of changes over time between outcome assessments. Farr (2014) appeared to have found a similar pattern over time. More long-term longitudinal studies are needed if we are to better understand how child outcomes change over time across different types of families.

Missing Data

The APA (2010) notes that “missing data can have a detrimental effect on the legitimacy of the inferences drawn

by statistical tests. For this reason, it is critical that the frequency or percentages of missing data be reported along with any empirical evidence and/or theoretical arguments for the causes of data that are missing” (p. 33). Yet, there have been a number of controversial studies in which the extent of missing data seemed to me to be extreme, nor was it explained, as required by the APA. Sometimes the missing data were not reported directly but appeared as much lower degrees of freedom than might have been expected. For example, Fulcher, *et al.* (2002) began with 80 participants but some of their *t* tests had only 50 degrees of freedom, an indication of extensive missing data. Erich, Kanenberg, Case, Allen, and Bogdanos (2009) started with 259 families, but missing data reduced their participant count to as few as 115 parents (56% missing data) and 78 children (70% missing data). Wainright, Russell, and Patterson (2004) had as much as 29.5% missing data, while their later study (Wainright & Patterson (2008) had as much as 46.6% missing data. In the case of Wainright, *et al.* (2004; Wainright & Patterson, 2006, 2008) the problem of missing data was compounded by the fact that up to 61% (Sullins, 2015d) or more of their “same-sex” mother families were heterosexual families, and it is not certain how missing data were distributed among the presumed and actual lesbian parent families.

Sometimes problems related to missing data occur before data collection, when response rates are low (Bos, 2010) or biased towards one group or another. A retention rate over time was reported to be 93%, but what was not specified was how many of the same-sex parents went through a gender change or left their same-sex partner for an opposite-sex partner. Gartrell and Bos (2010) did not mention how many same-sex partners had made such changes by year 17 of their study, though earlier at the 5-yr. mark two women had left their partner for a man and one woman had undergone a sex change and become a man (Gartrell, Banks, Reed, Hamilton, Rodas, & Deck, 2000, p. 543). To the best of my knowledge, Gartrell and Bos and their colleagues have never discussed the extent to which their lesbian mothers became involved with men or changed their gender – or how those changes may have affected their children.

Using Too Many Independent Variables

Cohen (1990) stated that “One thing I learned over a long period of time that is so is the validity of the general principle that less is more, except of course for sample size.... I have encountered too many studies with prodigious numbers of dependent variables, or with what seemed to me far too many independent variables, or (heaven help us) both” (p. 1304). Rosenfeld (2010) may serve as such an example, due to the use of dozens of independent variables. Likewise, Langbein and Yost (2009) also used over 65 independent variables

when they had between 141 and 153 cases, creating a ratio of cases to independent variables far below the recommended level of 10 (Schumm, *et al.*, 1980, p. 253), even below the level of five to nine deemed as acceptable by Vittinghoff and McCulloch (2007).

Adding independent variables to a regression analysis may reduce the significance of the variables first entered into the model. One corollary of using many independent variables with listwise deletion is that the overall missing data can increase substantially, effectively reducing the sample size and statistical power. More discussion on this issue can be found elsewhere (Schumm & Crawford, *in press*).

Omission of Important Mediating Variables

Sometimes important “surprises” can be hidden in research models that include mediating variables. When I was predicting support for same-sex marriage and parenting using the NFSS data set ($N=2,988$), I used a number of independent variables, including gender, education, sexual orientation, having had a parent who was perceived as having had a same-sex romantic affair, political orientation, and others. However, belief in sex without commitment was a relevant mediating variable that was remarkably the strongest predictor of support for same-sex marriage, while being predicted by many of the independent variables (Table 2).⁵ Notably, the correlates of belief in sex without commitment do not appear to be related to a high quality of life (Table 3). Regnerus's (2014) much larger study ($N=15,738$) found that those who supported same-sex marriage were up to several times more likely to endorse, among other issues, that viewing pornography is OK, that premarital cohabitation is good, that no-strings-attached sex is OK, that marital infidelity is sometimes OK, and that it is OK for three or more adults to live in a sexual relationship with each other. Unless this sort of mediating variable is used properly, one could draw inaccurate conclusions about why various groups support same-sex marriage at various levels (Tables 2 and 3).

Failure to Understand Clustering of Social Phenomena

One of the challenges of understanding modern human sexuality is that quite a few variables tend to cluster together, as part of the Second Demographic Transition (SDT; Lesthaeghe, 2010, 2014). Higher rates of premarital sex, lower fertility rates, later age at marriage, more people never getting married, increase of same-sex marriage, and higher rates of cohabitation are all asso-

ciated with each other. Trying to study any one of those variables in isolation is risky because the outcome may depend on how you use the other variables in a model. Those variables might act as distorters, suppressors, or extraneous variables depending on the situation and the specific theoretical model being tested. Inclusion or exclusion of any of those variables might dramatically change the outcome of the specific model being tested.

Langbein and Yost (2009) used several aspects of the Second Demographic Transition as both independent and dependent variables, which could have strongly influenced the outcomes they observed. In particular, one reason that Langbein and Yost (2009, p. 301) obtained $R^2 \geq 0.92$ was likely due to their predicting of related SDT variables (e.g., divorce rates) from one or more other related SDT variables (e.g., marriage rates). However, clustering does not mean that two SDT phenomena will never remain significantly correlated after controlling for other factors; Negy, Pearte, and Lacey (2013) found among a sample of several hundred young adults that attitudes toward same-sex marriage and toward polygamy were positively correlated ($r_p = 0.28$, $p < .05$, $d = 0.58$) even after statistically controlling for several variables, including autonomy, political conservatism, religiosity, conventionalism, traditionalism, gender, and social desirability (p. 65). Their results are consistent with a theory that approval of same-sex marriage could lead to more favorable attitudes toward polygamous marriage, although their results were correlational rather than causal.

Statistical Ignorance: Attempts to “Prove” the Null Hypothesis

Cohen (1990) noted that “The null hypothesis.... is *always* false in the real world” (p. 1308). When someone tries to prove that which “is always false,” then it is possible that political considerations are taking the limelight. Rather than proving the null, the goal should be to identify the effect sizes, however small, of the apparent associations among the variables under study (Schumm, 2010f). In other words, truth should matter more than publishability (Nosek, Spies, & Motyl, 2012), and the consideration of effect sizes is one of the better ways to get at factual truth, although trivial but significant results or non-significant but meaningful effects may tend to get published if their results are deemed politically desirable. Herek (2006) has stated that “The null hypothesis..... cannot be proved” but he goes on to argue that “A more realistic standard is the one generally adopted in behavioral and social research, namely, that repeated failures to disprove the null hypothesis are accepted provisionally as a basis for concluding that the groups, in fact, do not differ” (p. 610). Later, Herek (2006) argued that “empirical research to date has consistently failed to find linkages between children's well-being and the sexual orientation of their parents. If gay,

⁵Regnerus (2012c) found similar results but focused on the pornography part of casual sex. Schneider (2013) criticized Regnerus's findings as from a “pretty nutty professor” (p. 8), and argued that a generalized social tolerance was underlying acceptance of both pornography and same-sex marriage. I used a variety of control variables to serve as proxies for social tolerance, but future research might try to measure and control for social tolerance specifically.

TABLE 2
 Ordinary least squares (ols) regression models predicting support for same-sex marriage and parenting from the New Family Structures Study (NFSS) using weighted data

	Standardized Regression Coefficients for Different Models and Mediating or Dependent Variables				
	A	B	C	D	E ⁱ
Independent variables ⁶					
Same-sex parental romance ever in family of origin ^a	0.07‡	-0.06†	0.03	0.01	0.02
Age of respondent	-0.03	-0.11‡	-0.03	0.02	-0.01
Gender of respondent ^b	0.14‡	-0.03	-0.15	0.02	0.14‡
Educational attainment ^c	-0.07‡	0.21‡	0.02	0.04*	0.05†
Ever bullied as a child ^d	0.11‡	-0.17‡	0.02	0.05†	0.03*
Respondent sexual orientation ^e	—	-0.14‡	0.22‡	0.07‡	0.12‡
Quality of family life as a child scale ^f	—	—	-0.09‡	0.02	0.01
Sex without commitment scale ^g	—	—	—	0.38‡	0.38‡
Political orientation ^h	—	—	—	—	0.29‡
Adjusted <i>R</i> ²	0.045	0.109	0.074	0.161	0.374
<i>F</i>	28.4	59.1	32.4	41.3	180.8
<i>df</i>	5, 2887	6, 2828	7, 2745	8, 2710	9, 2701
<i>p</i>	<.001	<.001	<.001	<.001	<.001

Note Model A: Predicted variable is sexual orientation; Model B: Predicted variable is quality of family life during childhood scale; Model C: Predicted variable is sex without commitment scale; Model D: Predicted variable is political orientation; Model E: Predicted variable is support for same-sex marriage and parenting scale. ^aCoded: 0=never, 1=yes, ever. ^bCoded: male=1, female=2. ^cHigher scores represent higher educational attainment. ^dCoded: 0=never bullied, 1=bullied. ^eCoded: 0=heterosexual, 1=nonheterosexual. ^fCoded: 1=strong disagreement, 5=strong agreement that childhood family life was positive. ^gCoded so that a higher score represents a more favorable attitude toward sex without commitment. ^hCoded: 1=conservative, 5=liberal. ⁱCoded so that a higher score represents a more favorable attitude toward same-sex marriage and same-sex parenting. When mean substitution was used instead of listwise deletion of cases with missing data, the results were nearly identical. Adjusted *R*² never differed by more than .005 and standardized regression coefficients never differed by more than .02. However, the degrees of freedom for Model E were 9, 2978 with similar adjustments for the other models. **p* < .05. †*p* < .01. ‡*p* < .001.

lesbian, or bisexual parents were inherently less capable than otherwise comparable heterosexual parents, their children would evidence problems regardless of the type of sample. This pattern clearly has not been observed. Given the consistent failures in this research literature to disprove the null hypothesis, the burden of empirical proof is on those who argue that the children of sexual minority parents fare worse than the children of heterosexual parents" (p. 614). Recently, Maxwell, Lau, and Howard (2015) concluded that "Enormous sample sizes, much larger than those typical in psychology, are generally required for demonstrating that an effect is so small that it can essentially be regarded as null" and that "the continuation of underpowered stud-

ies in many areas of psychology.... undermines scientific psychology" (p. 496). There is a danger of using the excuse of difficulty (large sample sizes) to justify disregard of the risks of underpowered studies or the need to demonstrate equivalence ($d < 0.10$) before accepting a null hypothesis. The danger would be of accepting the null hypothesis before actually showing equivalence.

For example, Herek (2006) has stated that "empirical data addressing this question [if there is a tendency for gay or lesbian parents to raise children who grow up to be gay, lesbian, or bisexual] are limited." He seems to accept the null hypothesis by stating that "To the extent that data are available, however, they show that the vast majority of children raised by lesbian and gay parents eventually grow up to be heterosexual" (p. 613). On the one hand, what might be true of the vast majority might not be true for a statistically significant minority (hence, rejecting the null hypothesis). On the other hand, his argument obscures the nature of the outcome variable, which might be, e.g., experimentation with same-sex romantic sexuality rather than adoption of a same-sex sexual orientation identity. However, Schumm (2010b, 2013) found as many as 38 studies that had addressed this issue (is that limited data?), and many of them found a significant association between parental sexual orien-

⁶Scale items and Cronbach's *α*s from the New Family Structures Study (NFSS). Quality of life as a child ($\alpha=0.89$): My family relationships were safe, secure, and a source of comfort (Q28a); We had a loving atmosphere in our family (Q28b); All things considered, my childhood years were happy (Q28c); My family relationships were confusing, inconsistent, and unpredictable (recoded) (Q28g). Sex without commitment ($\alpha=.74$): It is a good idea for couples considering marriage to live together in order to decide whether or not they get along well enough to be married to one another (Q109c); It is OK for two people to get together for sex and not necessarily expect anything further (Q109d); Viewing pornographic material is OK (Q109i). Support for same-sex marriage and parenting ($\alpha=.74$): It should be legal for gays and lesbians to marry in America (Q109e); Gay and lesbian couples do just as good a job raising children as heterosexual couples (Q109m).

TABLE 3
Correlations between sex without commitment scale and other variables using both unweighted and weighted data from the New Family Structures Study (NFSS)

Variable	Zero-order Correlations	
	Unweighted	Weighted
Religious variables		
Importance of religious faith (Q70)	-0.51‡	-0.50‡
Religious attendance (Q71)	-0.53‡	-0.50‡
Organized religion is a negative force (Q74)	0.32‡	0.30‡
Family of origin background		
Loving atmosphere in family of origin (Q28b)	-0.12‡	-0.10‡
Problematic/Risky/Criminal behaviors		
View pornography often (Q82b)	0.36‡	0.36‡
Drink alcohol often (Q82c)	0.27‡	0.26‡
Often drink alcohol in order to get drunk (Q82d)	0.26‡	0.27‡
Use marijuana often (Q82e)	0.20‡	0.22‡
Use other illegal drugs often (Q82f)	0.04*	0.08‡
Smoke tobacco often (Q82g)	0.16‡	0.12‡
Gamble for money often (Q82h)	0.09‡	0.12‡
Ever had legal problems from drinking (Q83)	0.10‡	0.14‡
Ever had problems with close relationships due to drinking (Q84)	0.08‡	0.09‡
Ever used illegal drugs (Q85)	0.19‡	0.19‡
Ever arrested (Q86)	0.15‡	0.17‡
Ever convicted of a crime (Q88)	0.07†	0.10‡
Ever spent time in jail (Q89)	0.09‡	0.09‡
Mental health		
Felt depressed in the past week (Q76e)	0.09‡	0.08‡
Felt sad in the past week (Q76i)	0.08‡	0.10‡
I am an impulsive person (Q77l)	0.14‡	0.12‡
I get angry easily (Q77c)	0.08‡	0.09‡
I get stressed out easily (Q77f)	0.05†	0.09‡
Do things without thinking about the consequences (Q77j)	0.07‡	0.11‡
Like to try new things even if have to break rules (Q77k)	0.21‡	0.19‡
Have had suicidal thoughts in past year (Q81)	0.09‡	0.04+
Health and happiness		
Life happiness (Q79)	-0.12‡	-0.10‡
Ever had a sexually transmitted infection (Q124)	0.11‡	0.08‡
Number of abortions obtained (Q133)	0.06†	0.08‡
Current physical health (Q54)	-0.09‡	-0.04+
Relationship values		
Marriage is an outdated institution (Q109a)	0.38‡	0.35‡
Children do better if they are raised with a mother and a father (Q109b)	-0.31‡	-0.25‡
Relationship issues		
Worried that my partner does not really love me (Q75d)	0.14‡	0.11‡
Have thought about leaving current partner (Q97)	0.15‡	0.12‡
Ever cohabited outside of marriage (Q99)	0.27‡	0.22‡
Current relationship happiness (Q108)	-0.09‡	-0.12‡
Ever cheated sexually on your partner (Q127)	0.22‡	0.21‡

Note Correlations can be converted to effect sizes (*d*); $r \geq .10, d \geq 0.20$; $r \geq .196, d \geq 0.40$; $r \geq .243, d \geq 0.50$; $r \geq .287, d \geq 0.60$; $r \geq .371, d \geq 0.80$; $r \geq .447, d \geq 1.00$ (Cohen, 1988, p. 22). + $p < .10$. * $p < .05$. † $p < .01$. ‡ $p < .001$.

tation and children either adopting a same-sex sexual identity or experimenting with same-sex sexual activity or romantic relationships.

As Herek implied, if there are no patterns of differences in child outcomes as a function of parental sexual orientation (given comparable parent characteristics), it should not be possible to find such patterns involving effect sizes larger than 0.20 in the literature. However, there are many such examples (Schumm, 2015c). For example, Golombok, *et al.* (1997) compared heterosexual two-parent and single parent heterosexual families to lesbian mother families. The heterosexual families differed significantly from the lesbian families in terms of mother's age ($p < .05$), social class ($p < .001$), and family size ($p < .0001$). The lesbian mothers reported lower levels of depression than did the two-parent heterosexual mothers ($d = 0.31$) and higher levels of mother's warmth to child ($d = 1.04$, $p < .05$). The children of the lesbian mothers reported greater peer acceptance ($d = 0.19$) than did the children of the two-parent heterosexual mothers (and greater than that of the children of single parent heterosexual mothers as well). What is remarkable is that in this study the two-parent heterosexual mothers and their families were disadvantaged in terms of age (younger), family size (more children), fewer socioeconomic resources, children with lower levels of peer acceptance, higher levels of maternal depression, and higher levels of maternal stress ($d = 0.37$), but their children reported *higher* levels of *cognitive competence* ($d = 0.94$, $p < .001$) and *physical competence* ($d = 0.55$, $p < .01$), as reported previously (Schumm, 2011b, p. 92). In a later study, Golombok, *et al.* (2003) also found that children from two-parent heterosexual families reported greater cognitive competence ($d = 0.14$) and physical competence ($d = 0.38$) than children from two-parent lesbian families, even though the latter families had greater socioeconomic status, greater maternal acceptance, lower stress, fewer children, and less frequent corporal punishment (Schumm, 2011b, p. 93). In other words, there are at least two studies in which very disadvantaged heterosexual parent families were compared to very advantaged lesbian families, and yet the children in the former reported substantially higher levels of physical and cognitive competence than those in the latter. What would the results have been had the pre-existing differences between the families been controlled statistically? No doubt, the children of the heterosexual families would have fared even better. Possibly, such results tell us that same-sex parents are intrinsically less effective as parents, but if they have a variety of important advantages then their children will turn out only slightly worse than the children of heterosexual parents. We can not know for sure, because there were no complete statistical controls for all of the pre-existing differences between the two groups of parents. Some scholars (Adams & Light, 2015) have argued that consensus has been

achieved about same-sex parenting, but it should raise concern if "consensus" was achieved with improper sampling and statistical methods.

Likewise, Herek (2014) rejected the results of Sarantakos's (1996a) research even though that research found effect sizes as large as 3.75 between child outcomes for heterosexual and same-sex parents (Schumm, 2015c). There is an important difference in our approaches to assessing research about null hypotheses. The Golombok, *et al.* (1997, 2003) studies found at least some significant differences favoring the children of heterosexual parents, even though those parents were substantially disadvantaged without statistical controls being used to control for all of those disadvantages. Herek (2014) dismissed what Sarantakos (1996a) found because of the unproven speculative *possibility* that some of the differences might have been a result of teacher bias, although there was no statistical evidence of such bias. Statistically significant evidence against the null hypothesis in at least some studies does not fit the narrative that no studies have ever found any such patterns. One could point to meta-analyses (e.g., Fedewa, *et al.*, 2015) to argue that the overall pattern supported the null hypothesis, but most meta-analyses have not factored in the pre-existing differences between the different groups of parents; as I described above, this uncorrected sampling bias can be extremely misleading.

Inconsistent Results Within or Between Studies

Wainright, *et al.* (2004) reported means and standard deviations for parental warmth and care from adults and peers (p. 1892). However, when they reported results for the same two variables later, using exactly the same sample (Wainright & Patterson, 2008, p. 122), the means were not the same.

In Erich, *et al.* (2009), results differed from Tables 2 to 3 with respect to age of the sample of children at adoption and number of previous placements (Schumm, 2010f, p. 963), an example of inconsistency of reporting within a single study. As noted earlier, some studies have featured inconsistent sample descriptions across different published articles (Schumm, *et al.*, 2009).

Several Low Quality Studies Equal One High Quality Study

Having numerous weaknesses or flaws in research is bad enough, but the bizarre idea expressed in the title of this section has been circulated before U.S. courts, most recently concerning same-sex marriage and parenting. One example of this is from Herek (2006), who said that, rather than using random samples and equivalence testing to evaluate null hypotheses, "[a] more realistic standard is the one generally adopted in behavioral and social research, namely, that repeated failures to disprove the null hypothesis [with convenience samples] are accepted provisionally as a basis for conclud-

ing that the groups, in fact, do not differ" (p. 610). "Low quality" refers to studies based on small samples and nonrandom data collection, often not double-blinded. Small samples can mean that even if medium to large effect sizes exist, they will not be statistically significant. Nonrandom sampling prohibits results from being generalized to the larger U.S. population and should never be used to establish public policy until replicated sufficiently on properly selected samples. If the study is not double-blinded, then bias can result from either the researcher(s) hinting at or the participants recognizing the types of responses desired. Such studies often do not even attempt to control for such biases or for pressures to respond in a social desirable manner.

For example, a large, random study with many controls (Regnerus, 2012a, b, d) was ridiculed (see Redding, 2013; Schumm, 2013; Turner, 2015; Yancey, 2015 for more details), while a small ($N=32$), nonrandom study without even basic demographic information except the age range of respondents (Leddy, Gartrell, & Bos, 2012) has, to the best of my knowledge, never been criticized. Redding (2013) suggested that what mattered here was not the methodology, but whether the study's conclusions agreed or disagreed with certain preconceived notions. As I pointed out, although the New Family Structures Study (NFSS) has many faults, it has no more faults than many other articles on same-sex parenting published in a variety of journals, including "top tier" journals (Schumm, 2012b). In reference to the same-sex parenting literature, Allen (2015) has concluded that "A series of weak research designs and exploratory studies do not amount to a growing body of advanced research" (p. 173). In other words, cumulative scientific knowledge – sound enough to be presented to judicial authorities – should be built upon sound scientific research involving sound theory; random, unbiased data; and sound statistical testing rather than merely a series of extremely limited and weak studies with many of the previously mentioned limitations, often in multiple combinations.

Problems with Statistical Interpretation

Failure to Understand or Report Effect Sizes

Cohen (1990) has stated that "I have learned and taught that the primary product of a research inquiry is one or more measures of effect size, not p values" (p. 1310). Since 1994, the APA (1994, p. 18) has been recommending that scholars report effect sizes as well as significance levels. In 2001, the APA highlighted the importance of reporting effect sizes, noting that "it is almost always necessary to include some index of effect size or strength of relationship" (p. 25). More recently, the APA (2010) reiterated that statement, "For the reader to appreciate the magnitude or importance of a study's findings, it is almost always necessary to include some

measure of effect size in the Results section" (p. 34), citing Cohen's d value as one option. The APA also stated that "When applying inferential statistics, take seriously the statistical power considerations associated with the tests of hypotheses. Such considerations relate to the likelihood of correctly rejecting the tested hypotheses, given a particular alpha level, effect size, and sample size. In that regard, routinely provide evidence that the study has sufficient power to detect effects of substantive interest. Be similarly careful in discussing the role played by sample size in cases in which not rejecting the null hypothesis is desirable (i.e., when one wishes to argue that there are no differences), when testing various assumptions underlying the statistical model adopted..." (p. 30). Wilkinson and the Task Force on Statistical Inference (for the APA) specifically stated that "Always provide some effect-size estimate when reporting a p value" (1999, p. 599). Warner (2013, p. 107) has cited as "small" effect sizes where Cohen's $d \leq .20$, with "medium" effect sizes between 0.20 and 0.79. Amato (2012) offered a slightly different interpretation of effect sizes with those between 0.20 and 0.39 deemed "moderate" and those at or above 0.40 deemed "strong" (above .60 were "very strong"). It is clear that the APA's recommendations on reporting effect sizes have been often ignored (Schumm, 2010f). Ioannidis (2012) has likewise cited "underpowered studies" (p. 650) as a major impediment to self-correction in social science.

The most honest "broker" of a study's outcome is the effect size of the result, *not* the significance level of the findings (Schumm, 2010f). Statistical significance can be manipulated by using a small sample to favor an acceptance of, or failure to reject, the null hypothesis, or a larger sample to favor a rejection of the null hypothesis. As Erich, *et al.* (2009) observed, failure to reject a null hypothesis could "be a function of small sample sizes and high standard deviations rather than there being no actual significant differences" (p. 401). Allen (2015) has stated that "The very small sample sizes found in many of these studies creates a bias toward accepting a null hypothesis of 'no effect' in outcomes between same-sex and heterosexual households" (p. 164). Likewise, Rosnow and Rosenthal (1996) stated that "Just because a p value is reported as 'statistically significant' does not mean that the effect was large, nor does a p value reported as 'nonsignificant' imply a trivial result" (p. 331). Unfortunately, "Many researchers also continue to obsess on p values to the exclusion of effect sizes and statistical power" (Rosnow & Rosenthal, 1996, p. 331).

Cohen (1992) indicated that "My intent was that medium ES represent an effect likely to be visible to the naked eye of a careful observer" (p. 156). Furthermore, Cohen (1988) observed that "Many effects sought in personality, social, and clinical-psychological research are likely to be small effects as here defined, both because of the attenuation in validity of the measures employed

and the subtlety of the issues frequently involved" (p. 13). Therefore, one should be cautious about rejecting a null hypothesis when effect sizes are ≥ 0.20 , even if the results are not significant statistically, unless the samples are very large. Moreover, according to Cohen (1992), effect sizes ≥ 0.50 would be substantial enough to be observable without the use of statistics while "small" effect sizes might still have considerable practical importance.

Wainright, *et al.* (2004) is often cited as a random sample study that did not find any significant differences between children of lesbian mothers and heterosexual parents (Patterson, 2009a). Rosenfeld (2010, p. 756) cites Wainright, *et al.* (2004), following Meezan and Rauch (2005), as one of the four highest quality studies ever conducted with respect to same-sex parenting, as well as citing Wainright and Patterson (2006, 2008) as exemplars of "nationally representative probability samples" (p. 756). Only 17 of the 44 children of lesbian mothers were actually children of lesbian mothers (the other 27 had heterosexual parents, Sullins, 2015d), and much of the data were missing (p. 1892, in terms of self-esteem, anxiety, and depression); data were only reported for 27 of 44 children of same-sex parents (39% missing) and for 37 of 44 children of heterosexual parents (16% missing). Still, there were results that were *unfavorable* with respect to the children of lesbian mothers. In terms of depressive symptoms, the effect sizes were 0.13 and 0.23 for sons and daughters, respectively. In terms of anxiety, the effect sizes were 0.79 ($p < .05$, $95\%CI = 0.06, 1.53$) and 0.33, for sons and daughters, respectively. In terms of parental warmth, the effect sizes were 0.16 and 0.36. Although most of the results were not significant statistically, Wainright, *et al.* (2004) did not report effect sizes, which often did not favor outcomes for the children of lesbian parents in spite of the weak independent variable (mixed group of parents vs. a group of heterosexual parents) and a small sample size decreased further from 88 to 64 by missing data. The effect sizes were of small/medium to large magnitude for four of the six comparisons.

In a larger sense, it is odd that the APA would cite studies like Wainright, *et al.* (2004) or reviews by Herek (2014) in defense of same-sex parenting when such studies or reviews did not follow the APA's own recommendations for reporting effect sizes and other methodological requirements. Herek (2014) spent several pages discounting Sarantakos (1996a) and never discussed the very strong effect ($d > 1.00$) sizes involved in that study, despite calls by the APA to do so when evaluating research. Recently, Cheng and Powell (2015) reanalyzed data from the New Family Structures Study (NFSS; Regnerus, 2012a, b) and reported that they only found four significant results, compared to the twenty or more reported by Regnerus (2012a, b). However, they did not report effect sizes for any of their results. Because

they reduced the number of same-sex parent families considerably, it is actually possible that the effect sizes were unchanged, but due to the smaller sample statistical significance was lost. Had they shown that the effect sizes were also reduced, that would have implied far more strongly that the previous NFSS results were a result of poor measurement and methodology.

Failure to Use Theory or Measures Regarding Social Desirability Response Bias

Herek (2014) acknowledged that social pressures may cause human research participants to "intentionally give researchers inaccurate self-reports" (p. 597). In particular, parental self-reports are subject to social desirability response bias. This problem may be amplified when parents are aware that their answers might be used to advocate for their same-sex parental rights, which can occur when a study is not double-blinded. Many authors of reports on same-sex parenting have expressed an awareness of this problem, but the majority of such reports did not follow up on this theoretical expectation by measuring social desirability, as it might apply to the self, ratings of one's children, or ratings of one's relationships with others (Appendix).

For example, Gartrell, Hamilton, Banks, Mosbacher, Reed, Sparks, *et al.* (1996) indicated that "To the extent that these subjects [i.e., lesbian mothers] might wish to present themselves and their families in the best possible light, the study findings may be shaped by self-justification and self-presentation bias" (p. 279). Later, Erich, *et al.* (2009) stated that "Responses of this sort are subject to the effects of social desirability and impression management" (p. 403). Surprisingly, while Erich, *et al.* measured social desirability, they did not control for it in their statistical analyses. Thus, in some cases social desirability has been measured, but was not used or was measured incorrectly (e.g., measuring individual social desirability rather than relationship social desirability). Despite reluctance on the part of many to control for social desirability response bias, bias related to social desirability responding has been discussed for some time (Phillips & Clancy, 1972; Schumm, Bollman, & Jurich, 1981, 1982; Schumm, Hess, Bollman, & Jurich, 1981; Nederhof, 1985; Kozma & Stones, 1987). It is also possible for researcher bias to play a role. As Erich, Leung, and Kindle (2005) noted, "Social justice agendas may have distorted interpretations of research findings" (p. 46); likewise, Stacey and Biblarz (2001) acknowledged that "ideological pressures constrain intellectual development in this field" (p. 160) and that the personal values of scientists in this area "play a greater part than usual in how they design, conduct, and interpret their studies" (p. 161). It is remarkable that scholars may recognize the importance of social desirability in research in general and in same-sex parenting in particular, but few have measured or controlled for it. Lick, Patterson, and Schmidt (2013) are a recent excep-

tion, and found that controlling for social desirability did change some of their results (p. 243).

Model Selection Problems

When a researcher presents a complex model in which, e.g., one outcome is predicted from perhaps 80 independent variables, it might seem that the very best model had been selected. But the truth is that with 80 independent variables there are thousands of possible ways to have selected various combinations of those independent variables. Each of those combinations may yield a different statistical result. The researcher is free to keep testing different models until, possibly, the one outcome desired is achieved, perhaps to show that a particular variable is or is not significant statistically. As Studenmund (2010) as noted, "the weakness is that researchers can estimate many different specifications until they find one that 'proves' their point, even if many other results disprove it" (p. 156) and "One of the weaknesses of econometrics is that a researcher can potentially manipulate a data set to produce almost any results by specifying different regressions until estimates with the desired properties are obtained. Thus, the integrity of all empirical work is potentially open to question" (pp. 170–171). Sensitivity analysis is one approach to overcoming this limitation.

Sometimes the truth is "on the bubble." When we were studying anthrax vaccine and its association with Gulf War illnesses, in our models sometimes it would be a significant predictor, and at other times not quite ($p < .10$). Other predictors were more robust with respect to the different models we tested. Since the researcher can test as many models as possible until the "right" answer is obtained, the odds are high that the researcher will capitalize on chance and the results may not replicate in future studies (Maxwell, Lau, & Howard, 2015). It also appears common in controversial research for researchers to measure a variety of control variables but then not use them as statistical controls, as has been discussed at length elsewhere (Schumm, 2005a; Schumm, 2008; Schumm, 2010f; Schumm, 2011a, b; Schumm, 2012a). The APA (2001, p. 5) recommends the appropriate use of statistical controls.

As one example, Erich, Leung, and Kindle (2005) predicted family functioning from parental sexual orientation and a few statistical controls, except for parental education, which was higher for the same-sex parents by an effect size of approximately 0.53 (Erich, Leung, Kindle, & Carter, 2005; Schumm, 2010f, p. 959), which corresponds to $r = -0.26$. Rosenfeld (2010) stated that "the second-most-important factor in childhood progress through school appears to be parental educational attainment" (p. 762), which suggests that parental education might indeed have something to do with family functioning. Herek (2014) stated that "parental socioeconomic status" (p. 606) can affect children's outcomes.

Erich, Leung, and Kindle, (2005, p. 55) found that heterosexual parental orientation predicted family functioning positively ($b = 0.17, p < .10$), and might have done so significantly had educational differences between the same-sex and heterosexual parents been statistically controlled for. To show how this might be: if one sets the association between heterosexual parenting and family functioning to 0.17 and the association between education and heterosexual parenting to -0.26 , then if the association between education and family functioning is between 0.20 (then $r_p = 0.25, p < .05$) or .50 ($r_p = 0.36, p < .01$), with $N = 72$, the association between heterosexual parenting and family functioning would be statistically significant as a partial correlation. One has to wonder why a number of other variables were controlled for statistically but not education, for which there was a strong pre-existing difference between the two groups being compared. Using plausible assumptions about the association between parental education and family functioning, it appears that had education been controlled for, heterosexual parenting would have been positively and significantly associated with better family functioning in Erich, Leung, and Kindle's (2005) data.

Another issue with model selection concerns the use of mediating variables. Mediating variables can illuminate how processes may be occurring. Splitting the sample on the mediating variable is sometimes seen as a way to "control" the variable, but often its role will seem to disappear. For example, suppose that same-sex parents are less stable than heterosexual parents. It is possible that sexual orientation might predict parental stability, which would predict other child outcomes, such as children's educational outcomes. Potter (2012) found that children of same-sex parents fared worse on educational outcomes than children from heterosexual parents, until one controlled for parental stability. Such a result seems to imply that parental sexual orientation has no effect on child outcomes. However, it is entirely possible that parental instability mediates the association between parental sexual orientation and child outcomes. Potter's data indicated that most same-sex parents did not have stable relationships (Schumm, 2012b), although Cheng and Powell (2015) remain uncertain about the legitimacy of the coding for same-sex parents in Potter's data. The data should have been tested for the significance of the mediating or indirect effect of parental sexual orientation on child outcomes, operating through parental instability. Similarly, one might use parental instability as a control (exogenous) variable rather than as a mediating or intervening variable to make it appear that, taking stability into account statistically, there are "no differences in outcomes" [which is how Herek (2014, p. 602) interpreted the results of studies by Potter (2012) and Rosenfeld (2010)].

If the independent variable (e.g., same-sex parenting) is strongly related to instability and instability is

strongly related to adverse child outcomes, at the very least there might be substantive and significant indirect or mediating effects of the independent variable on the outcomes. Herek (2014) has recognized that instability is more likely among same-sex parents, at least in their past (p. 606). Thus he concluded that “variables related to family stability should be accounted for when comparing groups” (p. 606), which leaves open the question of how to do so. Using stability as a mediating variable rather than a control variable will illuminate how parental variables affect child outcomes. Herek (2014, pp. 610–616) criticized the Regnerus (2012a, b) study for confounding parental sexual orientation and family stability (among other issues) but overlooked the fact that other researchers (e.g., Golombok, *et al.*, 2003) have results with similar limitations (Schumm, 2012b). Both of these studies are important, if citations are any indication, with 122 for Regnerus (2012a) and 301 for Golombok, *et al.* (2003) as of December 18, 2015, in Google Scholar.

Model specification or selection is extremely important to interpretation, especially in terms of the use or non-use of relevant mediating variables. Any particular selection of variables may lead to one result, while a different selection of variables could lead to an entirely different result. Cheng and Powell (2015), e.g., used a different set of control variables than Regnerus (2012a, b) and found different results. Regnerus (2012a) controlled for respondent's age, gender, race/ethnicity (two levels), mother's education, household income while growing up, experience being bullied as a youth, and the gay-friendliness of state legislation. Cheng and Powell (2015) added controls for father's age, region, residential area, and added further options for race/ethnicity; they also redefined same-sex parent status and eliminated cases in which respondents provided inconsistent or illogical responses to a number of questions. With their changes, Cheng and Powell (2015) found that of 24 significantly different outcomes between same-sex and other families, only 4 remained significant. They conclude that their results support “the longstanding body of scholarship that confirms minimal differences in the consequences of living with same-sex or opposite-sex parents” (p. 625).

Inaccurate Statistical Procedures and Reporting

In dealing with controversial research, it is possible to find instances of non-significant results being reported as significant, reports of non-significant results that actually were significant statistically, and instances of reports where significant results were contained within the data but not reported as such.

Brachman, *et al.* (1960) claimed that in May 1957, 300 mill workers at the Arms Mill in Manchester, New Hampshire voluntarily participated in a test of human anthrax vaccine, with 150 receiving the vaccine and 150 a placebo. They claimed that the difference between no

infections in the treated group and 4 infections in the placebo group was significant statistically, by a chi-squared test ($p = .044$). However, the chi-squared test is an approximation. The Fisher's Exact Test (FET) is more accurate, and, in this case, the results with the FET were *not* significant. Furthermore, it is not clear how a later analysis (Brachman, *et al.*, 1962) of the same test involved only 149 vaccinated workers and a placebo group of 164 workers. There also were many design flaws and analytical methods used in the anthrax vaccine studies, as detailed elsewhere (Schumm & Brennehan, 2004; Schumm, *et al.*, 2004; Schumm, 2005c; Schumm & Nass, 2006).

It is possible to find research where results reported as non-significant were actually significant but the statistical test was applied incorrectly. For example, Rivers (2000) reported non-significant results for single and multiple suicide attempts since a function of school absenteeism, indicating that students who were more often absent from school were *not* more likely to attempt suicide. However, if one combines the data for single and multiple attempts, since those outcomes are mutually exclusive, then one finds that 43% (36/83) of those who had been absent from school, compared to 21% (7/33) of those not absent, had attempted suicide (two-sided Fisher's Exact Test, $p < .05$; odds ratio = 2.85, 95%CI = 1.11, 7.29, $p < .05$), a result that was statistically significant, the opposite of what Rivers (2000) reported.

Reviewing Tasker and Golombok (1995, 1997) one can easily come away with the sense that having a lesbian mother has no effect on the child's own sexual orientation in terms of attraction, identity, or behavior. However, more careful statistical analyses indicated that if an adolescent reported having same-sex sexual attractions, they were significantly more likely to report actual experience with same-sex sexual behavior if their mother was lesbian than if she was a heterosexual single parent (Schumm, 2004g). Furthermore, children of lesbian mothers were significantly more likely to have considered the possibility of becoming involved in a same-sex sexual relationship than were children of heterosexual mothers (Schumm, 2004g). Also, it was apparent that several of the children of lesbian mothers had considered engaging in same-sex sexual relationships even though they had *never* experienced same-sex sexual attractions. In other words, same-sex parenting was associated with children's same-sex sexual behavior even when same-sex sexual attractions were *absent*. The point here is that parental modeling and expectations may affect children, apparently independently of any biological connection related to same-sex sexual attraction, although it is possible that some children may indicate a same-sex sexual orientation as part of social desirability or an attempt to please parents to whom they are strongly attached. Whether such possibilities are true or not is not the point here, but that if the fine points of results are overlooked or covered up the

chance of understanding some of the deeper processes of parenting may be greatly reduced.

Using Weak Independent Variables

Cohen (1990) disagreed with the practice of collapsing variables into yes/no categories, as a practice of the “willful discarding of information” (p. 1306). Use of weak independent variables decreases the chance of rejecting the null, or increases the chances of “supporting the null hypothesis,” as some researchers incorrectly interpret it. This situation will be considered below in a number of important studies.

Landbein and Yost (2009) used whether or not a state had some type of legal union for same-sex couples as their independent variable. At the time, only one state had approved same-sex marriage *per se*. Testing for the effect of same-sex marriage (with only one state out of 50 allowing it) confounds “same-sex marriage” with the state’s many other characteristics. For example, if same-sex marriage predicted lower marriage rates, how would one know that the cause was “same-sex marriage” vs. a host of other possible factors? As an analogy, scientists would give little credit to a study of Hispanics when the study sampled 49 Anglos and one Hispanic participant. In addition, treating legalization of same-sex marriage as a yes/no variable assumes that any effect of same-sex marriage does not change over time or with time. A better, stronger measure of possible effects of legalizing same-sex marriage would be to consider how many years same-sex marriage has been legal in a particular state. As one example, when I considered the number of years that same-sex marriage had been in effect in a state as an independent variable, it predicted lower rates of fertility (Schumm, 2015a).

Wainright, *et al.* (2004) and Wainright and Patterson (2006, 2008) reported results from the ADD HEALTH study, in which they compared 44 adolescents from heterosexual families with 44 adolescents allegedly from same-sex families. However, Patterson (2009b) admitted that at least 26 of the 44 same-sex families may have been heterosexual families; Sullins (2015d) reviewed the same 44 cases and found that 27 were miscoded heterosexual families. That means that 61% of the “same-sex” families were miscoded. This weakens any effect of “same-sex parenting” that might be involved in the comparison of the two groups used. Despite the weakness of the independent variable, an effect size of 0.79 (anxiety) was found in favor of the sons of heterosexual parents by Wainright, *et al.* (2004); effect sizes as large as 0.27 in favor of the children of heterosexual parents were found by Wainright and Patterson (2006). These severe limitations of the Wainright and Patterson studies, acknowledged by Patterson (2009b), did not stop Herek (2014) from stating that the three studies used “forty-four adolescents parented by female couples who reported they were married or in a marriage-like

relationship” (p. 603), a statement that is not correct (Sullins, 2015d).⁷

Hooker (1957, 1958, 1978, 1993) is renowned for her early studies on gay men. While many believe that she compared assessments from 30 gay men with 30 heterosexuals, in fact, she compared two groups which each included three (10%) bisexual men. As noted elsewhere (Schumm, 2012c), while she did find differences between the two groups of men, those differences might have been larger statistically if she had not included so many bisexuals in each of her two groups. Furthermore, demographic variables significantly distinguished the two groups of men. Comparisons of the two groups were not in terms of demographic variables, which thus should have been controlled statistically. The judges who were evaluating the men might have been biased by the demographic characteristics of the men (Schumm, 2012c).

Golombok, *et al.* (2003) included at least 3 and possibly as many as 15 (of 28) lesbian families in which the child had spent more time outside a lesbian family than in one (Schumm, 2014). At least one child had spent 9 mo. or less in a lesbian family out of nearly 10 yr. of its life, and yet the child’s outcomes were attributed to being in a lesbian family rather than to whatever its previous family structures may have been. Other children, more appropriately to the study, had been raised from birth in a lesbian family. Treating these very different family histories (i.e., born into a lesbian family vs. entering into a lesbian family in the most recent year of 10 yr.) as if they were the same muddies the interpretation of any reported effects of being raised in a lesbian-parent family. Number of years spent in a lesbian family might be a stronger independent variable than current familial status alone. Despite these issues, Rosenfeld (2010), following Meezan and Rauch (2005), cited Golombok, *et al.* (2003) as one of the four “highest-quality studies in this field” (p. 756). If so, this situation must be improved.

In Regnerus’s (2012a, b, d) NFSS study, Cheng and Powell (2015) have detailed numerous concerns, including missing data and a variety of measurement concerns, including out-of-normal-range answers to survey questions. However, the most serious criticism was that the independent variable called “type of family,” including allegedly same-sex families, was not measured consistently, thoroughly, or accurately. Others have made similar criticisms (Gates, *et al.*, 2012; Sherkat, 2012; Anderson, 2013; Ball, 2013; Becker & Todd, 2013;

⁷Sullins (2015d), after restricting Wainright’s sample to same-sex families and comparing unmarried and married same-sex couples, found that for some outcomes the children of married same-sex couples fared less well than children of unmarried same-sex couples or of heterosexual couples. Such an outcome raises the possibility that the duration of same-sex family life might play a role in outcomes for children rather than mere status of having been in a same-sex family for an unspecified period of time or the marital status of the parents (which might be correlated with years duration of the parental relationship). This will be revisited in the next paragraphs.

Perrin, Cohen, & Caren, 2013; Siegel, Perrin, Dobbins, Lavin, Mattson, Pascoe, *et al.*, 2013; Infanti, 2014; Reiss, 2014; Kaplan, 2015), although some have defended Regnerus's research (Destro, 2012; Monte, 2013; Redding, 2013; Wood, 2013; Yancey, 2015).

Detailed Examples of Poor Measurement and Statistics Combined with Hasty Applications to Social Policy

The following two detailed treatments of studies show in detail how poor measurement or statistical analysis can cause real confusion about what research results mean and their potential meaning for policy or legal changes. The important question to keep in mind is: if controversial issues are so poorly researched, is it reasonable to create public policy based on the studies; and if we do so, is it at all reasonable to shut down debate and research on these topics because it runs against those new policies ("public opinion" or "majority scientific opinion")? The eventual cost of such actions will be the impoverishment, even discrediting, of science (Duarte, *et al.*, 2014) and lack of empirically-based correction in public and private organizations.

Example 1: Gays in the Military

Whether homosexuals should serve in the military has a contentious history. President Clinton changed the policy to "Don't Ask, Don't Tell" (DADT), but under President Obama federal policy was changed to allow open service by all homosexuals. At present, the policy is being further changed to allow transgendered persons to serve openly. One major issue was whether the policy change would affect military readiness, probably through reduced social cohesion within units that had gay or lesbian members; another major issue was the effect on retention and recruiting. Pro-gay advocates, of course, have had an answer to these questions: "twenty-four nations now allow gays and lesbians to serve in their armed forces; none has seen any impairment to cohesion, recruitment, or fighting capability" (Frank, 2009, p. 160). Scholars used to be skeptical of absolute statements using words such as "none" or "any." The following example suggests that skepticism may well be in order, even today. Another issue was how well military standards regarding proper sexual conduct could be enforced equally for heterosexual, gay, and lesbian members (Schumm, 2004a). What the U.S. Army appears to have done, according to reports this author has received from an inside source, is that shower areas have been converted into individual stalls, individuals have been forbidden to appear naked in front of others (i.e., soldiers sleep in their uniforms in bed or in sleeping bags and must enter and exit individual shower stalls wearing robes or towels), and romantic involvements between soldiers (same-sex or heterosexual) have been discouraged (one heterosexual soldier was discharged from basic training for

merely *flirting* with a female soldier in formation). Such changes to shower facilities and regulations are not without financial and retention costs but appear to have been instituted to reduce any real or imagined effect of gay or lesbian soldiers serving openly in close quarters.

In my opinion, the debate on this issue has featured more heat than light, since empirical research with U.S. service members has been relatively scarce with respect to changing DADT. Thus, a recent report by Kaplan and Rosenmann (2012) assumed greater importance because it was an empirical investigation of social cohesion in military units and within military units of the Israeli armed forces, which some have claimed were able to allow gay and lesbian soldiers to serve openly without any difficulties (Kaplan, 2003). Kaplan and Rosenmann (2012) reported in an analysis of data from over 400 Israeli soldiers that perceived unit social cohesion was unrelated to perceptions of having or having had a gay or lesbian soldier in the unit (no time frame specified). As such, on the surface their results provided support for those who argued that ending the "gay ban" would not adversely influence military readiness or retention. But would a closer look at the research, with measurement and interpretation issues in mind, cast a different light on the interpretation?

The specific question used to assess whether a person had a gay or lesbian unit member was "Do you know, or have known in the past, of a homosexual or lesbian soldier in your unit?" (p. 427). This independent variable was weak for two reasons. First, the responses were merely "yes," "no," and "possibly." The high percentage (26%, p. 432) of "possibly" answers indicates uncertainty among many current unit members about their unit's history with gay personnel. There is a great deal of ambiguity about what the answer "possibly" might have meant to the respondents. It could mean that they simply did not know. It might mean that they had suspicions but no definite evidence. It might mean that they believed there were gays or lesbians in their unit, but they did not think they could prove it, if required to do so. It might mean that there were bisexuals in their unit but they were not sure if bisexuality counted as gay or lesbian as it was not part of the question. There might also have been some ambiguity about whether "soldier" was intended to also include noncommissioned officers, warrant officers, or commissioned officers. To this author, the answers "yes" or "no" appear to have much less ambiguity and might merit comparison directly, without introducing the ambiguity of the "possibly" answer.⁸

Second, there was no consideration of the rank of the respondent, the duration of time in the unit, activi-

⁸One indication of the ambiguous nature of the "possibly" response is that of the 26 possible contrasts, 16 times the mean score on unit social cohesion for the "possibly" response was closer to the mean score for the "no" response while 10 times it was closer to the mean score for the "yes" response, so there was no clear trend in the meaning of "possibly."

TABLE 4
 Effect Sizes and Significance Levels For Kaplan and Rosenmann (2012) For Unit Social Cohesion as a Function of Knowledge of Gay or Lesbian Peers in own Unit (Yes vs. No) For Noncombat and Combat Units

Type of Unit Unit Social Cohesion Items	Noncombat Units		Combat Units	
	ES	p	ES	p
Brotherhood	0.25	.21	0.10	.59
Warmth and physical closeness	-0.03	.86	0.34	.06
Wish to disclose personal issues	0.17	.22	0.23	.40
Chemistry and shared language	0.41	.04	-0.04	.85
Wish for validation	-0.01	.96	0.09	.63
Sense of social belonging	0.33	.10	0.09	.61
Intimacy	0.24	.23	0.41	.03
Enjoying doing things together	0.15	.44	0.20	.29
Love	0.21	.30	0.22	.22
Desire to be with them	0.17	.38	0.31	.10
Competitiveness	-0.29	.15	0.02	.90
Envy	-0.14	.46	0.17	.36
Admiration	0.27	.22	0.54	.004
Average effect size				
Positive emotions	0.24	0.27		
Negative emotions	-0.22	0.10		
One sample chi-squared test for positive emotions	81% ($p = .035$)		91% ($p = .007$)	

Note $df = 198$ for noncombat units and 204 for combat units. Cohen's d is reported for effect size (ES). Hedge's g corrects downward for bias in d , but the correction factor was trivial (0.995) with $df = 150$.

ties in the unit, most recent time with the unit, or any other meaningful factors. While Kaplan and Rosenmann (2012) reported effect sizes associated with their MANOVAs, they did not report effect sizes for each of their univariate analyses. These data were re-analyzed (Table 4; Schumm, 2004a). For service members from *noncombat units* there were four negative effect sizes (only one of whose magnitude was >0.20), two for positive emotions and two for negative emotions (Table 4). Of the remaining nine positive emotions, nine effect sizes were positive (greater social cohesion in units without a gay member); and of those nine effect sizes, six were greater than 0.20. One of the results appeared to be statistically significant at $p < .05$ and another at $p < .10$. The 9/11 (81.8%) split for positive emotions was significant ($p = .035$) by a one-sample chi-squared test. For service members from *combat units* there was one negative effect size, essentially zero (Table 4). Of the remaining 12 effect sizes, all were positive and seven were greater than or equal to 0.20. The percentage of positive outcomes for the positive emotions (10/11, 90.9%) was significant ($p = .007$) by a one-sample chi-squared test. Two of the results appeared to be significant at $p < .05$ and two at $p < .10$. Of the 22 tests for positive emotions, 19 resulted in positive effect sizes, a result significantly different from a 50/50 split ($p = .001$).

Thus, my findings with Kaplan and Rosenmann's data (Schumm, 2004a) were different than theirs and others (e.g., Knapp, 2008), with far more indications that

social cohesion was adversely impacted by the recalled presence of gay or lesbian service members, especially for combat units. I found three statistically significant results and three statistical trends among the 26 comparisons (one significant result and two or three trends might have been expected by chance alone). Of the 22 comparisons of positive emotions associated with unit social cohesion, 19 yielded positive effect sizes; and of those, 12 involved "medium" effect sizes by Warner's (2013) standard (e.g., >0.20) while one more effect size was exactly 0.20. By Amato's (2012) standards, three of the results would have been strong and ten moderate [59% (13/22) of those tested]. Four of the other effect sizes were between 0.15 and 0.19. The average effect size for positive emotions in noncombat units was 0.24, while it was 0.27 for combat units. Thus, the overall trend was for *reduced* perceived unit social cohesion, *lower* positive emotions, and *greater* negative emotions in units with known gay or lesbian peer service members compared to units without known gay or lesbian peer service members. This is not quite the same as "there is no possibility for making an argument based on evidence that lifting the ban would harm the military" (Belkin, 2001, p. 104). While obviously Belkin made that statement before Kaplan and Rosenmann (2012) published their study, no one had studied unit social cohesion this way—in other words, there was "no evidence" because no one had yet studied the issue in as much detail as did Kaplan and Rosenmann (2012).

Recently, Belkin, Ender, Frank, Furia, Lucas, Packard, *et al.* (2013) have argued, with limited evidence, that the repeal of DADT did not adversely affect unit cohesion in the U.S. military, and the Kaplan and Rosenmann (2012) report appears to support their argument. Indeed, Kaplan and Rosenmann (2012) concluded that “knowledge of gay peers did not yield decreased social cohesion” (p. 419), a finding they interpreted as calling “into question the assumption that openly gay service reduces social cohesion” (p. 431). While they admit that “hypotheses can never be conclusively rejected,” they stated that “we believe that in light of the theoretical and methodological considerations presented above, our findings inform the ongoing debate around DADT and its repeal and call into question the concern for unit social cohesion once gay soldiers are allowed to serve openly” (p. 433). Yet reanalysis of their data indicate otherwise, with most effects being associated adversely with awareness of gay men in the unit presently or in the past.

In a more recent report (Kaplan & Rosenmann, 2014), the authors compared social cohesion with male unit peers and within a romantic relationship with a girlfriend. Notably, the girlfriend was rated higher on love ($d=1.42, p<.05$), warmth and physical closeness ($d=1.84, p<.05$), seeking validation ($d=0.90, p<.05$), disclosing personal issues ($d=1.45, p<.05$), doing things together ($d=0.85, p<.05$), desire to be together ($d=1.42, p<.05$), “chemistry” and shared language ($d=0.69, p<.05$), and intimacy ($d=1.98, p<.05$), while unit peers were rated similarly on comradeship ($d=0.02, ns$) and higher on competitiveness ($d=0.44, p<.05$). Again, they did not report effect sizes or significance levels from comparing the means scores, although my calculations of the effect sizes found many effect sizes to be in the “large” or “very large” range. However, the different patterns in ratings between non-romantic and romantic friends reinforce the idea that structuring the military so that romantic peers (same-sex or heterosexual) would serve together in the same units would be confusing at best and against good order and discipline at worst, because of inherently conflicting loyalties across the same dimensions of social cohesion where romantic friends, for the most part, would receive far higher social cohesion ratings than other military peers.

Example 2: Gay Marriage and Implications

To make good decisions, one needs good information. U.S. courts have been flooded in recent decades with cases involving marriage and parenting rights of gays, lesbians, and bisexuals. Social science has played a role in many of these decisions (Schumm & Crawford, *in press*). In many situations, the courts have been presented with inferior and/or incorrect social science. However, detecting the problems with some of that social science has not always been easy. Here I present just one example of how

low-quality research has been used in a recent trial concerning gay marriage, namely *Obergefell v. Hodges*, which eventually reached the U.S. Supreme Court and led to the Court imposing a requirement on all states to provide same-sex marriage licenses. The research of Rosenfeld (2014) played an important role in the original trial in the state of Michigan because it implied that stability rates for same-sex and opposite-sex couples were similar if marital status was taken into account (thus, providing an argument from social science that providing legal marriage for same-sex couples would stabilize the relationships of same-sex couples and benefit their children through the greater stability of their parents).

One critical problem with Rosenfeld (2014) is that the response rate was only 13% (p. 909). Of course, Rosenfeld is not alone; Allen (2015, p. 160) has noted that Bos (2010) reported research that featured a response rate of only 3.6%. Herek (1998) criticized research in which samples had only a 20% response rate and had been published in lower tier journals, but I have yet to see anyone likewise critique Rosenfeld (2014) for a much lower response rate. It may be simply a matter of Rosenfeld (2014) having found what was expected; hence, his immunity from criticism (Redding, 2013).

Umberson, Thomeer, Kroeger, Lodge, and Xu (2015) described Rosenfeld's (2014) research as having proven that “same-sex and different-sex couples have similar break-up rates once marital status is taken into account” (p. 98). At the same time, Umberson, *et al.* (p. 101–102) recommend taking parental status and relationship duration into account when comparing same-sex and different-sex couples, particularly in terms of stability. Rosenfeld (2014) stated that “After controlling for marriage and marriage-like commitments, the break-up rate for same-sex couples was comparable to (and not statistically distinguishable from) the break-up rate for heterosexual couples” (p. 905). However, Rosenfeld used a weak independent variable and a questionable dependent variable.

In terms of his dependent variable, he included 96 couples (of 3,009 couples) as “stable” when at least one partner had died over the 4-yr. period. This attrition was not mentioned by Rosenfeld (2014) even though the American Psychological Association recommended discussion of such issues as far back as 1999 (Wilkinson and the Task Force on Statistical Inference, p. 596). The APA (2010) also indicates that such troublesome observations should not be omitted “to present a more convincing story” (p. 12). In terms of independent variables, Rosenfeld (2014) included mixed-orientation marriages among the heterosexual marriages, although mixed-orientation marriages are likely to have lower stability rates and lower satisfaction rates (Tornello & Patterson, 2012). Removing the dead couples and the mixed-orientation couples from the analyses, the results obtained are presented in Table 1.

Among all unmarried couples, both parents and non-parents, the instability rates were very similar, with lower instability rates in some rows for same-sex couples. Similar does not mean “low”: among couples of either sexual orientation, with romantic relationships of 10 yr. or less, break-up rates were high, between 40% and 65%. Among all married couples, the patterns were very different. Instability rates were generally lower among married couples compared to unmarried couples. First, however, the instability rates for married same-sex couples were approximately three times higher in most categories than the rates for married heterosexual couples. Second, the rate differential between married and unmarried status was much less for same-sex couples; in other words, it seems that marriage tended to stabilize heterosexual couples far more than same-sex couples. Third, there were virtually no cases for married same-sex parents ($n=4$) while there were 488 cases for married heterosexual parents. Although the breakup rate for married same-sex parents was 25% over 4 yr., to achieve a statistically significant difference the breakup rate for heterosexual married parents would have had to be less than one percent over 4 yr., a rate about half of the typical breakup rate per year for married heterosexual couples. This lack of data means that it is not valid to draw conclusions about the effect of marriage on instability for same-sex parents or for the effect of parental status on married same-sex couples. Given that Rosenfeld deliberately oversampled same-sex couples, the lack of data for married same-sex parents is a huge concern.

Regardless of multivariate results, if one looks at non-parents for whom there are sufficient data the break-up rates for unmarried couples may be similar, but the break-up rates for married same-sex and married heterosexual couples are quite different, which indicates that “controlling for marriage” is a questionable procedure. It is difficult to draw conclusions about married same-sex parents vs. married heterosexual parents because there are virtually no cases for the former group. Again, in this context, “controlling for marriage” means little because for one subgroup there are virtually no marriages of parents to be controlled for. Thus, the anticipated benefits of legal marriage for the children of same-sex parents are not at all clear.

In both of these examples to compare outcomes for two groups, one should clearly define the two groups and be sure that measurement leads to pure groups rather than overlapping or ambiguous groups (Cheng & Powell, 2015). Using answers of “yes” or “no” to vague questions does not create a strong independent variable. When independent variables are weak or sample sizes are very small, non-significant findings may well be the result of poor methodology more than anything else.

Psychological and Political Reaction in Social Science

In principle, scientists are supposed to be unbiased, like referees in sports, making empirical “calls” based on the situation, not on which team they hope wins the game. Abbott (2012), along such lines, argued that “Readers assume researchers and authors are unbiased and objective and that their statistics have not been manipulated to support a personal theory; nevertheless, some researchers seek to advocate or support a social and political agenda” (p. 36). Scientists are subject to ordinary human tendencies, such as the confirmation bias. But if an honest discussion of controversial issues is to occur, we must value everyone’s right to free expression, especially in academic venues (Williams, 2011).

There is an important distinction between disagreeing with what someone has said and attempting to attack him as a person, to discredit him or deny his personal or academic credibility. Sometimes the attempt to discredit makes no pretensions to formal critique; e.g., as Stacey and Biblarz (2001) observed, value differences can lead to contentious debates in the field about same-sex marriage or parenting. Superficial attempts to discredit those with whom one may disagree are common but very harmful to debate. At times there is unusual behavior from those who should know better. For example, once while I was debating Stacey and Biblarz at a national conference, a professor yelled at me in front of a large audience that I “did not know anything about qualitative or quantitative research.” However, a year or so later, when an ACLU speaker at another national conference said I had betrayed the conservative cause and supported gay adoption rights, she and other scholars of the same political views warmed back up to me after discussion.

There are a variety of ways that scholars (and even more often, lawyers) take more subtle but equally *ad hominem* “cheap shots” at those with whom they disagree in areas of controversial research. There are multiple techniques designed to either stop a conversation in the immediate sense or, much more seriously, serve to shut down all public debate on specific topics. This is unfortunate, as Abbott (2012) observes, because skepticism is a part of the essence of science; irrefutable dogma should be the domain of religion, not science.

“Quality” of Journals

Publishing in a so-called “top tier” journal means less than one might think. My own analysis of citations from my “top tier” and “lower tier” journal publications suggests that “top tier” articles can go uncited for decades—“lower tier” articles can have many more citations than “top tier” articles (Schumm, 2010a, d). Some social science organizations (e.g., the APA, the American Sociological Association, the National Council on Family Relations) have taken political stances with respect to controver-

sial issues. The danger is that because these organizations also sponsor most U.S. professional scholarly journals in their topic areas, there is a risk of biasing which articles are accepted or rejected on the basis of political conformity to the organization's stated political positions rather than scholarly merit. Since some of the journals of nearly every social science organization are considered more "top tier" than others, the bias likely influences both higher and lower tier journals. "Top tier" is a subjective label now mainly based on the impact factor, which is typically interpreted in an inaccurate manner, particularly if it is misattributed as a measure of the quality of every article published in a journal. Impact factors can be and are padded by numerous editorial practices, such as asking authors to cite the journal more often, by including editorial pieces or short comments that cite previous articles in the same journal (or other journals owned by the same publisher), or by narrowing the scope of articles accepted to increase the chances of citing previous articles on the same topics. Journal impact is greatly influenced by the size and reach of the marketing arm of the publisher, which obviously has nothing to do with article quality. The most important aspect of a research article is its own quality, not where it happened to be published.

"Legitimate" Scholars

It requires no effort or justification to declare that someone who disagrees with you is "not a legitimate scholar." If the scholar under discussion has a Ph.D. and has published widely, and in my case has taught numerous statistics and research methodology courses at graduate and undergraduate levels, then I would argue that it is nonsense to declare such a person as "not legitimate." Such declarations are especially easy on the Internet, where criteria for scholarly success are seldom discussed by those who are eager to discredit their political opponents. The Ph.D. is a doctor of philosophy degree, which to some extent implies that the scholar is able to comment legitimately on a wide variety of research, even that outside his or her field of specific expertise. In other words, a poor statistical analysis is "poor," whether it is published in a medical journal or in a sociological journal, and a sociologist should be able to critique the medical journal's articles and a medical researcher the sociological journal's articles.

During the Nazi era, a pamphlet entitled "100 Authors Against Einstein" (Israel, Buckhaber, & Weinmann, 1931) was published to discredit Einstein, who replied "Why 100 authors? If I were wrong, then one would have been enough."⁹ In other words, scientific consensus can be wrong and the goal should be to determine which research is correct, not how many votes there are in favor of one side or the other. Along similar lines, it is

⁹Clark, D., January 7, 2013, "A Hundred Authors Against Einstein." <http://www.weeklysciencequiz.blogspot.com/2013/01/a-hundred-authors-against-einstein.html>

of concern that scholars who have published unpopular findings are being accused of having published "hate-speech" (Turner, 2015, p. 109) or those who have criticized inferior research studies are being levied with "charges of hate" (Yancey, 2015, p. 26) for not accepting such research at its face value. Notably, one activist organization went directly to the publisher of a journal of which I am the editor and challenged the publication of a comment by Cameron and Cameron (2012) that was critical of one of my editorials (Schumm, 2012c). The attack was not based on the quality of the comment but was a personal attack on Paul Cameron, who had co-authored the comment. Yancey (2015) reported of Regnerus that "his detractors condemned his research in public forums and attacked his character. His university conducted an inquiry into the research and an audit was made about the peer review process by which his article was accepted. Even a petition was forwarded to have his article removed from the journal that published it. His detractors argued that his research was abnormally bad and motivated by his Catholic faith" (p. 26).

"Legitimate" Studies

All studies have some methodological limitations. Just because a published study has some limitations does not make it illegitimate, i.e., meaningless. Sometimes there are severe limitations of usefulness for policy or legal purposes, typically because the sampling or procedures prevent results from being generalized. Logically, the usefulness of such results is *limited*; it is not accurate to say that the entire study and its results are "illegitimate." If the data were faked or the research was conducted in an unethical manner in violation of human rights, those issues would make a published article illegitimate. Unfortunately, it is all too easy to label a study "not legitimate" without having to provide any evidence. For example, the scholarly research of Regnerus (2012a, b) was not only challenged for a variety of problems (e.g., Cheng & Powell, 2015), but "his work was characterized as a form of hate speech" (Turner, 2015, p. 109) and attacked in a variety of ways, as discussed previously (Yancey, 2015).

Politically, it is convenient to present an alleged "consensus" and dismiss any contrary findings as inherently illegitimate. But that action overlooks the fact that science often advances the most when presented with contradictory findings that must be resolved. If scientists routinely dismissed research findings just because they differed from previous research, science would make no progress.

Statements of Limitations Illogically "Disarm" Criticism

Wilkinson and the Task Force on Statistical Inference (1999) pointed out that some authors seem to think if they "confess" the limitations of their study, then they are "absolved" of the consequences of those limitations.

Wilkinson, *et al.* specifically noted that “Confession should not have the goal of disarming criticism” (1999, p. 602). Acknowledging limitations is “for the purpose of qualifying results and avoiding pitfalls in future research” (p. 602). In other words, if one’s study is of low quality—small sample, not double-blinded, used questionable measures, etc.—then it should not be expected that such a study will contribute much to the overall research literature, much less policy and law.

The flip side of this logic about “confession” is also sometimes (perhaps unwittingly) used. Even if low-quality studies are replicated, confessing the limitations do not make the research more applicable to public policy—or, more directly, replication cannot resolve consistent methodological problems. In controversial research with small and low-quality samples, it is not uncommon for an author to admit limitations, but nevertheless interpret the results as having important policy implications. According to Wilkinson and the Task Force on Statistical Inference of the APA (1999), applying research with many serious limitations to develop policy and law is inappropriate. To give a single example, Herek (2014) has argued that certain research with major limitations should be considered relevant for public policy decision-making. At the same time he dismissed studies with similar limitations, with which he disagreed (Sarantakos, 1996a; Regnerus, 2012a, b; Allen, 2013). Unfortunately, these limitations are common in research on both sides of the issue (Schumm, 2012b) for various reasons; until we have better studies, weak research should not be a basis for public policy.

At times, such behavior by researchers has enormous consequences to the public. Langbein and Yost (2009) admitted that their study had limitations, including the small number of states that had approved gay marriage, and that “it may be too early to tell exactly what the effects of laws regulating same-sex marriage are at this point” (p. 306). They admitted that “We cannot say that we have disproved the existence of a link between laws permitting gay marriage and a negative impact on ‘family values’ indicators, but we can say that no such link is demonstrated in the data that we analyzed here” (p. 306). Nevertheless, they proceed to claim that “Permitting gay marriage does no harm, and making it legal may even be beneficial....” Thus, despite the fact that only one state (Massachusetts) had approved of gay marriage in their data and that effects of legal changes might take many years to accrue to a statistically significant level, they did not hesitate to draw the policy conclusion that gay marriage “does no harm.” Their research was very influential in trials involving the legality of same-sex marriage, in spite of major empirical flaws (Allen & Price, 2015)—for example, having a ratio of cases to variables of less than 3 to 1 in a multivariate analysis (Schumm & Crawford, in press). My point is that while it is good for a researcher to admit limitations

of a study, the implications of such limitations must be taken seriously in all interpretative commentary.

“This Scholar Has Been Discredited”

Ad hominem critique should always be suspect in science. If the person making this sort of comment is not a scholar, of course the comment should be given no weight. However, even if the critic is a scholar it may deserve no weight. I will provide two examples of how this might be done to damage of a scholar’s reputation, but has implications for the public.

It is possible that this type of attack is used most by lawyers in court when dealing with expert witnesses on research topics. At one trial, I was criticized for being ignorant of statistical principles. First, I had cited research where results with $p < .10$ had been reported (because the authors had reported this alpha level), a situation that is increasingly common in research (Schumm, *et al.*, 2013). I was able to show that critics had also published research where $p < .10$ while other scholars had used $\alpha = .10$ as their criterion for statistical significance, even in “top tier” journals (Schumm, *et al.*, 2013). Second, I had submitted a working paper to the court in which I had provided statistical results using more than one statistic (e.g., chi-square, *t* test, or correlation). This is a no-win situation if your opponent seeks to discredit you. If you report the best statistic (A), then they will say you did not use the other one (B). If you report B, then they will say you did not use the best one (A). And if you report both A and B, then they will say you must not know the difference. I reported both A and B (even C, sometimes) so I would have both at hand, because the court had said we could not introduce new results into the discussion at trial. Accordingly, I did not want to get caught having reported A and then be asked about B, which if I said what it was, I would have been liable for violating the court’s orders to have submitted all testimony in writing before the trial. That is, my attempt to meet the court’s mandated requirements led to my being vulnerable to accusations of not knowing the difference between statistics A and B (or C). Such Catch-22 situations are common in courts and are much used by lawyers.

On the other hand, and more crucially, it is easy for scholars to claim that a study has been discredited by citing some other scholar who has said so. For example, Umberson, *et al.* (2015) stated that “the findings from this study have been largely discredited” (p. 99) in reference to Regnerus’s (2012a) New Family Structures Study (NFSS). Yet Umberson, *et al.* (2015) omitted the fact that virtually every limitation of the Regnerus study could be found in many accepted studies on same-sex relationships (Schumm, 2012b)¹⁰; logically, if the problems with the Regnerus study made the results completely useless,

¹⁰I was challenged by a blogger, Scott Rose, with respect to one of my statements (Schumm, 2012b) but refuted that challenge later (Schumm, 2014).

then one could easily argue the same for many other highly regarded studies. At any rate, the term “discredited” is tossed around far too much, often with far less genuine evidence that one might suppose.

High (or Low) Quality = Cited Often (or Infrequently)

In one analysis, I found that the *lower* the methodological quality of a study, the more likely it was to have been cited in the scholarly literature in one topic area (Schumm, 2008). For example, I found two journal articles (Mucklow & Phelan, 1979; Miller, *et al.*, 1980), both published in the same journal by the same group of scholars at about the same time and using the same sample. The study of lower methodological quality (Mucklow & Phelan, 1979, cited 72 times according to Google Scholar as of December 18, 2015) has been cited far more often than the other (Miller, *et al.*, 1980, cited 9 times as of December 18, 2015). The only real difference in this natural experiment appears to have been that one study presented (currently) politically correct results, while the other did not (Schumm, 2010d). The point is that merely being cited often does not prove that the study was of high quality in an absolute sense nor that it was of higher quality compared to other studies. It may merely be a paper that says what scholars want to hear and are glad to cite, accordingly, to support their own opinions (Redding, 2013).

Sometimes apparent low citation rates are associated with selecting one article out of dozens that a scholar has published while ignoring the rest, as appears to have happened with Sofirios Sarantakos on the issue of same-sex parenting (Marks, 2012; Herek, 2014, pp. 607–610; Allen, 2015; Schumm, 2015c). It also can occur when the scholar is not from the U.S. or has published in journals not widely read in the United States. The bottom line is that the quality of a study should be based on its own merits; at best, citations are a weak indicator of quality. Citations may largely reflect results that are socially desired (Redding, 2013).

For years, the *journal* impact factor was used as a measure of the *article* quality—an odd and statistically impossible presumption. Now a growing body of evidence has turned researchers’ favor to the “article impact,” although institutions seem to be slow on the uptake. There is not yet much research on the effects of the publisher size and marketing activities, the effects of Internet search engines, etc. on citations, but I expect that within a decade, we will be seeing that even article impact metrics are less accurate than evaluating studies on their own merits.

“No Study Has Found any Evidence” Against a Favored Viewpoint

If this is true, it is possibly because no one has tried to find any evidence regarding the hypothesis in question.

Perhaps evidence has been found, but the study has been overlooked. Marks (2012) noticed that Patterson (2005) claimed that no study had ever found evidence not in favor of same-sex parenting, when there had been at least one such study (Sarantakos, 1996a) and probably many more (Schumm, 2015c). When someone says that no study has ever found any contrary evidence, I always question how comprehensive their literature review was, because results from a large number of studies are seldom unanimous when effect sizes are the basis for comparison. This issue is most likely an example of confirmation bias, as discussed previously. This problem could be exacerbated by the “file drawer” effect (researchers not bothering to publish negative findings or findings inconsistent with their own political or religious values).

Another example of this problem may be found in a research article by Hosking, Mulholland, and Baird (2015), in which they analyzed public voices of 7 sons and 8 daughters, ages 3 to 29, of same-sex parents. Their review of the literature concluded that “The research consensus was that while the children of gay and lesbian parents have more open ideas about sexual identity and gender fluidity, they identify as heterosexual in rates comparable to the rest of the population” (p. 340). Schumm (2010b, 2013) found that daughters of lesbian mothers were most likely to identify or experiment with nonheterosexual romantic relationships. Despite their conclusion about sociological consensus, what did Hosking, *et al.*’s (2015) own internal research evidence suggest? Of the eight daughters, one was likely too young (age 9) to identify with a sexual orientation and two other daughters had gay fathers, leaving five daughters with only lesbian mothers. One daughter identified as a lesbian (p. 339) while another hinted at some sexual orientation fluidity (being straight but not narrow, p. 340). Thus, internally, their research suggested a 20-40% rate of nonheterosexuality, compared to about 2-5% for many studies on personal sexual orientation, a pattern inconsistent with their own evaluation of the research literature.

APA Standards for Social Science Research Must be Adhered to

Although I have been critiquing social science research for over 35 yr., this does not mean that I am opposed to people doing whatever research they want to do. Rather, I am concerned that low-quality social science research – and low-quality medical research – is being used to create government policy and federal law. I did not invent such standards for research. The APA (1994, 2001, 2010) has long provided information on standards for research. It concerns me when researchers routinely violate APA standards for research but also expect that the APA will accept their research as useful for estab-

lishing policy and law or as a basis for creating APA positions on political and legal controversies.

Concern with standards for research need not be the end of social science research. Limitations of time and funding can restrict what can be done. For example, one's sample size might be limited by funding restraints. However, accurate reporting of results, reporting of effect sizes, use of reliable and valid measures, discussion of limitations, blinding of studies, statistical controls for known concerns such as social desirability response bias, acknowledgement of other viewpoints, and preparation of comprehensive literature reviews can be done even if funding is relatively low.

Patterson (2009a) cited numerous studies that did not comply with APA standards for research in order to support her call for changes to laws about same-sex marriage and parenting. A low-quality study can have many valid purposes – creating ideas for further, better research; raising questions about previous research; being a useful example of how low quality can bias research outcomes; putting some “flesh” on otherwise dry research concepts; proving that some otherwise rare cases exist, etc. The purpose for which it must not be used is the development of new public policies, regulations, or laws.

Implications

For Students

I think it is important for both undergraduate and graduate students to learn that research studies can actually have genuinely serious limitations, which in some cases may largely negate their scientific value. This realization should not discourage students from reading or using research but should encourage them to treat research with a certain amount of skepticism, especially when powerful political, governmental, or financial interests are backing a particular side of an issue. When a professor exposes the weaknesses of published research in the classroom, that should be *appreciated*.

I often point out to students doing research that they need to construct their research with strong methodology that is capable of disproving even the things in which they believe most strongly. For example, if one were studying abortion attitudes among students in Roman Catholic high schools, one should set up the research so that results could be obtained that disagreed with the expectations or doctrines of the Roman Catholic church hierarchy. Research should not be designed to “prove” one side or the other, but to test the hypothesis. This means that if you are hoping to not reject the null hypothesis, you should use as large a sample as possible, to reduce the chances of finding a small to medium size effect that would not be statistically significant with a small sample size. If you want to see the null hypothesis rejected, you should focus on effect size, not

statistical significance, because of the chance of finding a significant but trivial result.

For Professors

As a professor, I have found many students to be appreciative of learning how to evaluate published research in detail. However, some students who have strong political or social opinions may not appreciate hearing about limitations of research. Once a student dropped my class because I had the class walk through several medical articles to see for themselves that the research was corrupt – sample sizes and dates did not match up, and eventually the authors admitted to over a dozen mistakes in what they had reported (Schumm, *et al.*, 2009). Apparently, the student's relatives included a physician and she had difficulty accepting the possibility that medical research could be corrupt. Another time, I showed how several scholars (Balsam, Beauchaine, Mickey, & Rothblum, 2005; Balsam, Rothblum, & Beauchaine, 2005) had found certain results they did not like and had stated that they did not want the public to become aware of those results (Balsam, Rothblum, & Beauchaine, 2005, p. 484). I thought this was an interesting twist – results are published in a top tier journal but the authors do not want the public to find out about the research. Why, after all, would one publish if one does not want the public to know? One graduate student deemed that to be offensive and dropped my class.

At one point when there were three professors (myself among them) criticizing U.S. government research, we compared notes and realized that two of us had seen our homes or property near our home vandalized and set on fire. At another time, when I was publishing quite a bit of controversial research and providing some of that research to a state representative, legal actions of a severe nature were taken against some of my adult children by the state along the same lines as my research. Coincidence? Who knows? I doubt there will ever be any proof of any malicious actions by state or federal governments. But such “coincidences” can make one wonder what it takes to make oneself “an enemy of the state” through politically incorrect research. The easy way out would be to only critique research that was mostly irrelevant or that had no possibility of offending anyone. Although I agree it is risky to point out the limitations of research backed by powerful interests, it is even more vital to use examples that are meaningful in application. I agree with Nickerson (1998, p. 205), who seemed to believe that teaching students to overcome any of their own tendencies toward confirmation bias was an important educational objective, although it may not be clear what approaches might work the best for such an objective. At the same time, engaging in controversial research can be exciting, as one may discover important results that lead to positive social change.

For Attorneys

Unfortunately, attorneys are usually required as part of their job to take one side or the other in a dispute. I believe that this tends to cause them to see the world in black and white rather than a more realistic and complex range of grays, which social science research often finds the world to be. Nickerson (1998) specifically noted that science and law differ in important ways, one being that lawyers deliberately make a case for their side and are by no means obligated to look fairly at both sides of an issue, stating that attorneys on neither side are “committed to an unbiased weighing of all the evidence at hand, but each is motivated to confirm a particular position” (p. 175). Allen (2015, p. 155) has reported how research in controversial areas of social science has been sparked by legal cases. Patterson (2009a) stated that “it should be recognized that the reality is more complex than is usually acknowledged in legal and policy debates” (p. 733). I think there is a risk that attorneys on both sides of an issue may latch onto results they find useful for their side of the dispute without as much careful evaluation as might be warranted.

Our discussion (Schumm & Crawford, in press) of how an unsupported newspaper article (Peterson, 1984) was eventually cited by Patterson (1992), and, likewise, unsupported comments at an American Bar Association meeting (Bureau of National Affairs, 1987) were cited in the *Harvard Law Review* (Editors, 1989), with both being cited as fact in over seventy later scholarly articles and law reviews, shows examples of scholars and lawyers using information that supports their causes, regardless of its factual or non-factual nature. These are also cautionary examples for researchers tempted to be lazy in literature reviews, or who exclusively use particular article aggregators on the Internet.

Attorneys and experts can have conflicts over such matters. One time an attorney asked me to find several “bad things” about lesbian and gay persons to help their case. Unfortunately, those “bad things” were not correct, empirically. Consequently, I lost that consulting job. My sense is that many attorneys need to learn more about statistics and research methodology if they want to coordinate more effectively with social scientists on legal issues. Social scientists need to be advised of the key legal questions and where social science might or might not contribute to the discussion.

Social scientists need to maintain their independence, even if that does not please attorneys. At one trial I was told to testify however I wanted, even if it helped the other side, because the attorneys were interested in the facts. In contrast, in some cases attorneys have badgered social scientists to give only “yes” or “no” answers to questions. Yet, in the real and complicated world of social science, the truth cannot be expressed in simplistic answers; thus, social scientists have to remember on the witness stand that they swore

to tell the truth, not to give “yes” or “no” answers. Lawyers may also try to get social scientists to comment on matters outside their expertise, such as where certain groups of people might “spend eternity spiritually.” I would argue that the best answer for a social scientist is that such an issue is outside their expertise or several echelons above their pay grade!

Simplistic answers can be wrong. Once, I was asked in a deposition whether I thought homosexuality was sinful. I replied that homosexuality can be defined in terms of attraction, identity, and behavior, as well as other matters, and that functional relationships can be defined in terms of several issues. Thus, using an overly simplistic binary coding (low/high) on the 14 possible issues, I would have some 66,000 or more possible combinations of issues. I left it up to the lawyer to pick one of the 66,000 and allow me to discuss it at my leisure. The lawyer wanted me to discuss main effects, but the problem was that if there are interactions, then main effects do not convey the facts. The lawyer ultimately decided that it was not worth paying me over a hundred dollars an hour to follow rabbit trails. My experience has also been that many lawyers do not understand motivations of scientists who do research, and getting into debates with them when the terms are not understood is probably futile. Moreover, a lack of understanding of statistics by lawyers or judges can make accurate presentation of the evidence challenging at best, and can hinder effective use of expert witnesses in both their initial testimony and any rebuttals to their testimony. I do not wish to imply that all attorneys are guilty as charged here, but in my limited experience I have encountered far more problems than I would have anticipated.

Social science has an important role in guiding public policy and the law, but it should be accurate and well-replicated science, and all “sides” should present their perspectives and deal with criticisms. If only one side is allowed to present their perspectives or to deal with criticisms, I do not think that the results will be good for public policy or the law.

One issue that seems to be overlooked far too often is, *what criteria* would matter for decisions based on social science? For example: in the case of adoptions by LG couples, is the goal to show that at least 1% of same-sex parents are fit? That is not really a social science criterion, because (1) empirical results are not that precise and (2) logic alone would suggest that a few parents from any social group are likely to be excellent parents. In a contrasting example, perhaps the goal is to show that the children of one group of parents are five times more likely to have emotional disturbances than the children of another group of parents? For example, suppose I could provide evidence that the children of same-sex parents were twice as likely to be emotionally disturbed as the children of heterosexual parents. Would that be “strong” enough to convince supporters

of gay rights to not recommend that same-sex parents be allowed to adopt? What about eight times as likely? Would that be “strong” enough? My sense is that no criterion would be “strong” enough, because the issue is the rights of parents, and the arguments disregard the rights of children altogether.

Given the above example of ambiguities in how social science research is used (or not) to create public policy, a researcher might well ask, why bother with research if its results do not matter in the discussion anyway? My view is that one would have to start with agreed-upon criteria for what would constitute “proof” before each side presented their evidence. Otherwise, each side will continue to change the criteria to keep their side of the case alive. An interesting comparison here might be how convicted felons forfeit, in some states, voting rights, gun ownership rights, or the privilege of adopting a child, among other things. Surely, there are some convicted felons who were wrongly convicted and were innocent. Surely, there are some convicted felons who have become rehabilitated into outstanding citizens and represent no danger to anyone. But does that mean that *all* convicted felons should be granted such rights or privileges? Obviously, the states do not agree on such issues.

For Courts

To make legal decisions from scholarly data, one should have agreed-upon criteria; otherwise, decisions will be more random than rational. Without criteria, how will a court know how to weigh different research results? One important criterion would be effect size. As noted previously, Cohen (1992) indicated that “My intent was that medium ES represent an effect likely to be visible to the naked eye of a careful observer” (p. 156). Furthermore, Cohen (1988) observed that “Many effects sought in personality, social, and clinical-psychological research are likely to be small effects as here defined, both because of the attenuation in validity of the measures employed and the subtlety of the issues frequently involved” (p. 13). Cohen’s idea and Amato’s (2012) comment in which he argued for accepting effect sizes of 0.20 to 0.39 as “moderate” indicate that courts should accept effect sizes of 0.20 or greater as of potential value, even if the results (because of small sample sizes) were not statistically significant.

Another important criterion should be whether or not the sample(s) used were random in nature as opposed to nonrandom or convenience samples. Only results from random samples from known populations should be generalized, for purposes of policy or law, to whole populations. For example, suppose one collects data on parents whose annual household incomes are an average of \$200,000 per year using a convenience sample. Such results would probably not generalize to parents with much lower income, even if the sam-

ple had been random. Yet the nonrandom nature of the sample further disallows generalization of *any* results to the entire population of parents or their children. Convenience samples may show us that a certain type of family exists, but they do not tell us what percentage of these families may be highly functional nor whether generally these families are more or less functional than other families in the general population.

A third criterion should be the use of strong (i.e., reliable, valid) independent variables for predicting child outcomes, as discussed previously. A much longer list of criteria could be developed, of course. But my point here is that without criteria of which everyone is aware, the discussion to make little sense. For example, suppose one researcher discusses an effect size result of 0.30. By Cohen’s and Amato’s criteria, that should be an important result. However, another researcher might dismiss such a result, especially if it was not statistically significant. The court must decide what effect size will be deemed useful or not useful or researchers will quibble forever over the meaning of research results.

There can be “non-criteria” as well. For example, I showed (Schumm, 2013) that dozens of scholars are on record, from 1975 to 2014, reporting that there is no relationship whatsoever between parental and child sexual orientation. So many have argued that one might legitimately believe that there was a scientific “consensus” in the field of social science about that matter. Courts often may accept scientific consensus as a benchmark for making decisions related to science. However, of some 38 studies reviewed in Schumm (2013), the vast majority found a positive association between parental and child sexual orientation, a result with which a few progressive scholars agreed. In other words, sometimes a minority scholarly opinion is correct even in the face of apparent scholarly “consensus” in the other direction (Adams & Light, 2015).

Another example would be research based on maternal reports of child development; if parental social desirability bias is not taken into account, the results may mean very little because the outcome variable may be measuring parental social desirability more than anything else. Using a measure of individual social desirability (Crowne & Marlowe, 1964; Zhou, Eisenberg, Wang, & Reiser, 2004, p. 357; Lick, Tornello, Riskind, Schmidt, & Patterson, 2012; Lick, *et al.*, 2013) or relationship social desirability (Edmonds, 1967; Schumm, Bollman, *et al.*, 1981, 1982; Schumm, Hess, *et al.*, 1981; Schumm, Akagi, & Bosch, 2008) may under-control for *parental* social desirability (the Appendix gives examples of items that could be used to measure each of these three concepts), even if other forms of social desirability significantly predict a child’s social functioning or other measures of development (e.g., Zhou, *et al.*, p. 360). To date, I am not aware of any study that has ever measured and controlled for parental social desir-

ability when comparing maternal reports of child development or adult child reports of parent's functioning as a function of the parent's sexual orientation. Indeed, Lo, Vroman, and Durbin (2015) have recently indicated that "There is an extensive literature on response bias and social desirability (e.g., Edwards, 1957), but to our knowledge no studies have explored parental social desirability and its impact on parent ratings of child behavior" (p. 287). I have little doubt that a meta-analysis of maternal reports of child development might show few or minor differences as a function of maternal sexual orientation (e.g., Fedewa, *et al.*, 2015), but without controls for parental social desirability response bias those findings might only reflect bias rather than an accurate assessment of a child's developmental status. My point is that even if I bring to a court a letter or list of several hundred scholars who claim such and such to be true, that really might mean very little from an even-handed scientific perspective. Empirical facts should be established by data, not opinion, even given large numbers of opinions, even from good scholars.

Some court decisions call for Solomonic wisdom. The story goes that two women approached King Solomon of Israel and both claimed maternity of the same infant. Solomon developed a test to deal with the conflict, saying the baby should be cut in half so that each mother would get an equal share. The real mother protested and told him to give the baby to the other woman, while the other woman thought an "equal split" was a great idea. If a court is presented with a "no difference" hypothesis, that two variables are not at all related, in any study ever, then a Solomonic test can be devised. For example, a court might ask the other side to present three different outcomes (with effect sizes of 0.20 or greater) for which there were at least three scientific studies (preferably from random samples) each that disagreed with the "no difference" hypothesis, taking into account the relative strength of the predictor variable(s) in each study. If the "no difference" side gets "edgy" about this Solomonic test, then that is a clue that they really do not believe their own hypothesis. If the "difference" side feels comfortable or excited about that test, that shows they probably have the facts to back up their claims of difference. If the situation is reversed (if the goal is to disprove a difference), then one side might claim there had never been any studies that did not find at least some differences. Then the 3 outcomes/3 studies "test" could be geared to finding examples of results with effect sizes smaller than 0.20, regardless of their statistical significance.

Conclusions

There are positive ways to approach controversial research with creative ideas and methods that allow considerable progress.

I agree with Elovitz (1995), who argued that social science has been misused in the briefings before various

courts dealing with same-sex parenting issues. While I think it can be shown that both sides of many issues have misrepresented the research literature, the specific preponderant theme in my own experience with controversy has been the claim that same-sex parenting does not have any influence on children, a "no-differences" hypothesis that I believe can be refuted (Schumm, 2004g; Schumm, 2008; Schumm, 2010b; Schumm, 2011a, b; Schumm, 2013; Schumm, 2015b, c) despite the claim of sociological "consensus" (Herek, 2006; Patterson, 2009a; Manning, *et al.*, 2014; Adams & Light, 2015). First, I think that social science research as a process is being damaged by unscholarly practices such as refusal to share data, biased reviewers, weak or incomplete literature reviews, resistance by journals or authors to corrections to their results, biased citations based on political agreeableness more than scientific merit, censorship of ideas or research contrary to more powerful interests, literature reviews that overlook contrary results, and even human rights violations, among others (John, *et al.*, 2012).

Second, I think the meaning of social science data is being misrepresented due to a variety of weak or incorrect methodologies, including acceptance of incorrect facts, inaccurate use of statistics, use of too many or weak independent variables, omission of important dependent variables, lack of respect for effect sizes, attempts to prove the null hypothesis, questionable model selection, inconsistent results presented within or across published articles, use of nonrandom, biased samples, misalignment of theory and analysis, misuse of mediating variables, and high levels of missing data.

Third, discourse about social science results often shifts from academic discussion into attempts to discredit those with whom one may disagree, via a number of "cheap shot" criticisms which do not relate directly to the validity of research results but far more to academic status. Science and the public are not being well served by these problems, so policymakers need to be aware of them.

Those who interface with social science need to be fully aware of its limitations and how researchers' biases can selectively influence interpretations of results or bias results themselves. Controversial research presents even greater problems than usual. Therefore, greater caution should be used with respect to the acceptance of research in such areas and the need for constructive civil discourse may need to be reinforced more than usual.

References

- Abbott, D. A. (2012) Do lesbian couples make better parents than heterosexual couples? *International Journal of Humanities and Social Science*, 2(13), 30-46.
- Adams, J., & Light, R. (2015) Scientific consensus, the law, and same sex parenting outcomes. *Social Science Research*, 53, 300-310.
- Ahmann, E. (1999) Working with families having parents who are gay or lesbian. *Pediatric Nursing*, 25, 531-535.

- Allen, D. W. (2013) High school graduation rates among children of same-sex households. *Review of Economics of the Household*, 11, 635-658.
- Allen, D. W. (2015) More heat than light: a critical assessment of the same-sex parenting literature, 1995-2013. *Marriage & Family Review*, 51, 154-182.
- Allen, D. W., Pakaluk, C., & Price, J. (2013) Nontraditional families and childhood progress through school: a comment on Rosenfeld. *Demography*, 50, 955-961.
- Allen, D. W., & Price, J. (2015) Same-sex marriage and negative externalities: a critique, replication, and correction of Langbein and Yost. *Econ Journal Watch*, 12, 142-160.
- Amato, P. R. (2012) The well-being of children with gay and lesbian parents. *Social Science Research*, 41, 771-774.
- American Psychological Association (APA). (1994) *Publication manual of the American Psychological Association*. (4th ed.) Washington, DC: Author.
- American Psychological Association (APA). (2001) *Publication manual of the American Psychological Association*. (5th ed.) Washington, DC: Author.
- American Psychological Association (APA). (2010) *Publication manual of the American Psychological Association*. (6th ed.) Washington, DC: Author.
- Anderson, E. (2013) The need to review peer review: the Regnerus scandal as a call to action. *Journal of Gay & Lesbian Mental Health*, 17, 337-351.
- Ball, C. A. (2003) Lesbian and gay families: gender nonconformity and the implications of difference. *Capital University Law Review*, 31, 691-749.
- Ball, C. A. (2013) Social science studies and the children of lesbians and gay men: the rational basis perspective. *William & Mary Bill of Rights Journal*, 21, 691-764.
- Balsam, K. F., Beauchaine, T. P., Mickey, R. M., & Rothblum, E. D. (2005) Mental health of lesbian, gay, bisexual, and heterosexual siblings: effects of gender, sexual orientation, and family. *Journal of Abnormal Psychology*, 114, 471-476.
- Balsam, K. F., Rothblum, E. D., & Beauchaine, T. P. (2005) Victimization over the life span: a comparison of lesbian, gay, bisexual, and heterosexual siblings. *Journal of Consulting and Clinical Psychology*, 73, 477-487.
- Becker, A. B., & Todd, M. E. (2013) A new American family? Public opinion toward family status and perceptions of the challenges faced by children of same-sex parents. *Journal of GLBT Family Studies*, 9, 425-448.
- Belcastro, P. A., Gramlich, T., Nicholson, T., Price, J., & Wilson, R. (1993) A review of data based studies addressing the effects of homosexual parenting on children's sexual and social functioning. *Journal of Divorce & Remarriage*, 20, 105-122.
- Belkin, A. (2001) The Pentagon's gay ban is not based on military necessity. *Journal of Homosexuality*, 41, 103-119.
- Belkin, A., Ender, M. G., Frank, N., Furia, S. R., Lucas, G., Packard, G., Samuels, S. M., Schultz, T., & Segal, D. R. (2013) Readiness and DADT repeal: has the new policy of open service undermined the military? *Armed Forces & Society*, 39, 587-601.
- Biblarz, T., & Stacey, J. (2010) How does the gender of parents matter? *Journal of Marriage and Family*, 72, 3-22.
- Bos, H. M. W. (2010) Planned gay father families in kinship arrangements. *Australia and New Zealand Journal of Family Therapy*, 31, 356-371.
- Bos, H. M. W., Gartrell, N. K., Peyser, H., & van Balen, F. (2008) The USA National Longitudinal Lesbian Family Study (NLLFS): homophobia, psychological adjustment, and protective factors. *Journal of Lesbian Studies*, 12, 455-471.
- Bos, H. M. W., & Sandfort, T. G. M. (2010) Children's gender identity in lesbian and heterosexual two-parent families. *Sex Roles*, 62, 114-126.
- Bozett, F. W. (1987a) Children of gay fathers. In F. W. Bozett (Ed.), *Gay and lesbian parents*. New York: Praeger. Pp. 39-57.
- Bozett, F. W. (1987b) Gay fathers. In F. W. Bozett (Ed.), *Gay and lesbian parents*. New York: Praeger. Pp. 3-22.
- Bozett, F. W. (1993) Gay fathers: a review of the literature. In L. D. Garnets & D. C. Kimmel (Eds.), *Psychological perspectives on lesbian and gay male experiences*. New York: Columbia Univer. Press. Pp. 437-457.
- Brachman, P. S. (2005) *Letter to Food and Drug Administration, Division of Dockets Management*, 5630 Fishers Lane, Room 1061, Rockville, Maryland 20852. Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, March 2.
- Brachman, P. S., Gold, H., Plotkin, S. A., Fekety, F. R., Werrin, M., & Ingraham, N. R. (1962) Field evaluation of a human anthrax vaccine. *American Journal of Public Health*, 52, 632-645.
- Brachman, P. S., Plotkin, S. A., Bumford, F. H., & Atchison, M. M. (1960) An epidemic of inhalation anthrax, the first in the twentieth century. II. Epidemiology. *American Journal of Hygiene*, 72, 6-23.
- Brewaews, A., Ponjaert, I., van Hall, E. V., & Golombok, S. (1997) Donor insemination: child development and family functioning in lesbian mother families. *Human Reproduction*, 12, 1349-1359.
- Bureau of National Affairs, Inc. (1987, August 25) Reports, proposals, and rulings: ABA annual meeting provides forum for family law experts - homosexuality and family law. *The Family Law Reporter*, 13(41), 1512-1513.
- Cameron, P., & Cameron, K. (2012) Re-examining Evelyn Hooker: setting the record straight with comments on Schumm's (2012) reanalysis. *Marriage and Family Review*, 48, 491-523.
- Canning, T. T. (2005) Gay and heterosexual fathers: a comparative analysis of child behavior and well-being. *Dissertation Abstracts International, Section B: Science and Engineering*, 66(07), 3995.
- Chan, R. W., Brooks, R. C., Raboy, B., & Patterson, C. J. (1998) Division of labor among lesbian and heterosexual parents: associations with children's adjustment. *Journal of Family Psychology*, 12, 402-419.
- Cheng, S., & Powell, B. (2015) Measurement, methods, and divergent patterns: reassessing the effects of same-sex parents. *Social Science Research*, 52, 615-626.
- Clark, D. (2013, January 7) A hundred authors against Einstein. *Weekly Science Quiz*. Retrieved from www.weeklysciencequiz.blogspot.com/2013/01/a-hundred-authors-against-einstein.html.
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*. (2nd ed.) Hillsdale, NJ: Erlbaum.
- Cohen, J. (1990) Things I have learned (so far). *American Psychologist*, 45, 1304-1312.
- Cohen, J. (1992) A power primer. *Psychological Bulletin*, 112, 155-159.
- Committee on Science, Engineering, and Public Policy. (1995) *On being a scientist: responsible conduct in research*. Washington, DC: National Academy Press.
- Crawford, D., Schumm, W. R., & Schumm, M. R. G. (2015) Does the "his and hers" marital satisfaction pattern still persist? Poster session presented at the annual conference of the National Council on Family Relations, Vancouver, British Columbia, November 14, 2015.

- Crawford, J. T., Durate, J. L., Haidt, J., Jussim, L., Stern, C., & Tetlock, P. E. (2015) It may be harder than we thought, but political diversity will (still) improve social psychological science. *Behavioral and Brain Sciences*, 38, e164 (online, 14 pages).
- Crowl, A., Ahn, S., & Baker, J. (2008) A meta-analysis of developmental outcomes for children of same-sex and heterosexual parents. *Journal of GLBT Family Studies*, 4, 385-407.
- Crowne, D. P., & Marlowe, D. (1964) *The approval motive*. New York: Wiley.
- Destro, R. A. (2012) "You have the right to remain silent": does the U.S. Constitution require public affirmation of same-sex marriage? *BYU Journal of Public Law*, 27, 397-440.
- Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. E. (2015) Political diversity will improve social psychological science. *Behavioral and Brain Sciences*, 38, e130, (online, 13 pages).
- Dundas, S., & Kaufman, M. (2000) The Toronto lesbian family study. *Journal of Homosexuality*, 40(2), 65-79.
- Editors of the *Harvard Law Review*. (1989) Developments in the law: sexual orientation and the law. *Harvard Law Review*, 102, 1508-1671.
- Editors of the *Harvard Law Review*. (1990) *Sexual orientation and the law*. Cambridge, MA: Harvard Univer. Press.
- Edmonds, V. H. (1967) Marital conventionalization: definition and measurement. *Journal of Marriage and the Family*, 29, 681-688.
- Edwards, A. L. (1957) *The social desirability variable in personality assessment and research*. New York: Dryden.
- Elovitz, M. E. (1995) Adoption by lesbian and gay people: the use and mis-use of social science research. In M. E. Elovitz & C. Schneider (Eds.), *Legal issues facing the nontraditional family*, 1995. New York: Practising Law Institute. Pp. 171-191.
- Erich, S., Kanenberg, H., Case, K., Allen, T., & Bogdanos, T. (2009) An empirical analysis of factors affecting adolescent attachment in adoptive families with homosexual and straight parents. *Children and Youth Services Review*, 31, 398-404.
- Erich, S., Leung, P., & Kindle, P. (2005) A comparative analysis of adoptive family functioning with gay, lesbian, and heterosexual parents and their children. *Journal of GLBT Family Studies*, 1, 43-60.
- Erich, S., Leung, P., Kindle, P., & Carter, S. (2005) Gay and lesbian adoptive families: an exploratory study of family functioning, adoptive child's behavior, and familial support networks. *Journal of Family Social Work*, 9, 17-32.
- Faragallah, M. H., & Schumm, W. R. (1996) Willingness to have one's children serve in the military: an indicator of acculturation among Arab immigrants to the United States: a brief report. *Journal of Political and Military Sociology*, 24, 105-115.
- Faragallah, M. H., Schumm, W. R., & Webb, F. J. (1997) Acculturation of Arab-American immigrants: an exploratory study. *Journal of Comparative Family Studies*, 28, 182-203.
- Farr, R. H. (2014) Longitudinal outcomes for children adopted by lesbian, gay, and heterosexual parents. Paper presented at the annual conference of the National Council on Family Relations, Baltimore, MD, November 19-22.
- Farr, R. H., Forssell, S. L., & Patterson, C. J. (2010) Parenting and child development in adoptive families: does parental sexual orientation matter? *Applied Developmental Science*, 14, 164-178.
- Fedewa, A. L., Black, W. W., & Ahn, S. (2015) Children and adolescents with same-gender parents: a meta-analytic approach in assessing outcomes. *Journal of GLBT Family Studies*, 11, 1-34.
- Fitzgerald, B. (1999) Children of lesbian and gay parents: a review of the literature. *Marriage & Family Review*, 29, 57-75.
- Frank, N. (2009) *Unfriendly fire: how the gay ban undermines the military and weakens America*. New York: St. Martin's Press.
- Fulcher, M., Chan, R. W., Raboy, B., & Patterson, C. J. (2002) Contact with grandparents among children conceived via donor insemination by lesbian and heterosexual mothers. *Parenting: Science and Practice*, 2, 61-76.
- Gartrell, N., Banks, A., Reed, N., Hamilton, J., Rodas, C., & Deck, A. (2000) The National Lesbian Family Study: 3. Interviews with mothers of five-year-olds. *American Journal of Orthopsychiatry*, 70, 542-548.
- Gartrell, N., & Bos, H. (2010) US National Longitudinal Lesbian Family Study: psychological characteristics of 17-year-old adolescents. *Pediatrics*, 126, 1-9.
- Gartrell, N., Hamilton, J., Banks, A., Mosbacher, D., Reed, N., Sparks, C. H., & Bishop, H. (1996) The National Family Lesbian Study: 1. Interviews with prospective mothers. *American Journal of Orthopsychiatry*, 66, 272-281.
- Gates, G. J., et al. (2012) Letter to the editors and advisory editors of *Social Science Research*. *Social Science Research*, 41, 1350-1351.
- Goldberg, A. S. (2010) *Lesbian and gay parents and their children: research on the family life cycle*. Washington, DC: American Psychological Association.
- Golomb, B. A. (2008) Acetylcholinesterase inhibitors and Gulf War illnesses. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 4295-4300.
- Golombok, S. (2015) *Modern families: parents and children in new family forms*. Cambridge, UK: Cambridge Univer. Press.
- Golombok, S., Perry, B., Burston, A., Murray, C., Mooney-Somers, J., Stevens, M., & Golding, J. (2003) Children with lesbian parents: a community study. *Developmental Psychology*, 39, 20-33.
- Golombok, S., Tasker, F., & Murray, C. (1997) Children raised in fatherless families from infancy: family relationships and the socioemotional development of children of lesbian and single heterosexual mothers. *Journal of Child Psychology and Psychiatry*, 38, 783-791.
- Gottman, J. S. (1989) Children of gay and lesbian parents. *Marriage and Family Review*, 14, 177-196.
- Green, G. D., & Bozett, F. W. (1991) Lesbian mothers and gay fathers. In J. C. Gonsiorek & J. D. Weinrich (Eds.), *Homosexuality: research implications for public policy*. Thousand Oaks, CA: Sage. Pp. 197-214.
- Grover, K. J., Russell, C. S., Schumm, W. R., & Paff-Bergen, L. A. (1985) Mate selection processes and marital satisfaction. *Family Relations*, 34, 383-386.
- Gwanfogbe, P. N., Schumm, W. R., Smith, M., & Furrow, J. L. (1997) Polygyny and marital/life satisfaction: an exploratory study from rural Cameroon. *Journal of Comparative Family Studies*, 28, 55-71.
- Hawkins, S. A. (2011) Family relationships and adolescent behavior: a look at families headed by heterosexual, lesbian, and gay parents. Unpublished doctoral dissertation, The Claremont Graduate Univer., Claremont, CA.
- Hendrix, C. C., Jurich, A. P., & Schumm, W. R. (1995) Long term impact of Vietnam War service on family environment and satisfaction. *Families in Society*, 76, 498-506.
- Herek, G. M. (1998) Bad science in the service of stigma: a critique of the Cameron group's survey studies. In G. M. Herek (ed.), *Stigma and sexual orientation: understanding prejudice against lesbians, gay men, and bisexuals*. Thousand Oaks, CA: Sage. Pp. 233-255.
- Herek, G. M. (2006) Legal recognition of same-sex relationships in the United States: a social science perspective. *American Psychologist*, 61, 607-621.
- Herek, G. M. (2014) Evaluating the methodology of social science research on sexual orientation and parenting: a tale of three

- studies. *University of California-Davis Law Review*, 48, 583-622.
- Hooker, E. (1957) The adjustment of the male overt homosexual. *Journal of Projective Techniques*, 21, 18-31.
- Hooker, E. (1958) Male homosexuality in the Rorschach. *Journal of Projective Techniques*, 22, 33-54.
- Hooker, E. (1978) Epilogue. *Journal of Social Issues*, 34, 131-135.
- Hooker, E. (1993) Reflections of a 40-year exploration: a scientific view on homosexuality. *American Psychologist*, 48, 450-453.
- Hosking, G., Mulholland, M., & Baird, B. (2015) "We are doing just fine": the children of Australian gay and lesbian parents speak out. *Journal of GLBT Family Studies*, 11, 327-350.
- Huggins, S. L. (1989) A comparative study of self-esteem of adolescent children of divorced lesbian mothers and divorced heterosexual mothers. In F. Bozett (Ed.), *Homosexuality and the family*. New York: Harrington Park. Pp. 123-135.
- Infanti, A. C. (2014) The house of Windsor: accentuating the heteronormativity in the tax incentives for procreation. *Washington Law Review*, 89, 1185-1233.
- Institute of Medicine. (2006) *Gulf War and health: health effects of serving in the Gulf War*. Vol. 4. Washington, DC: National Academies Press.
- Ioannidis, J. P. A. (2005) Why most published research findings are false. *PLOS Medicine*, 2(8), e124.
- Ioannidis, J. P. A. (2008) Why most discovered true associations are inflated. *Epidemiology*, 19, 640-648.
- Ioannidis, J. P. A. (2012) Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7, 645-654.
- Israel, H., Buckhaber, E., & Weinmann, R. (1931) *A hundred authors against Einstein*. Leipzig, Germany: R. Voigtlander Verlag.
- Javaid, G. A. (1993) The children of homosexual and heterosexual single mothers. *Child Psychiatry & Human Development*, 23, 235-248.
- John, L. K., Loewenstein, G., & Prelec, D. (2012) Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, 524-532.
- Kaplan, D. A. (2003) *Brothers and others in arms: the making of love and war in Israeli combat units*. New York: Harrington Park Press.
- Kaplan, D., & Rosenmann, A. (2012) Unit social cohesion in the Israeli military as a case study of "don't ask, don't tell." *Political Psychology*, 33, 419-436.
- Kaplan, D., & Rosenmann, A. (2014) Toward an empirical model of male homosocial relatedness: an investigation of friendship in uniform and beyond. *Psychology of Men & Masculinity*, 15, 12-21.
- Kaplan, R. A. (2015) "It's all about Edie, stupid": lessons from litigating *United States v. Windsor*. *Columbia Journal of Gender and Law*, 29, 85-103.
- Knapp, D. E. (2008) Ready or not? Homosexuality, unit cohesion, and military readiness. *Employee Responsibilities and Rights Journal*, 20, 227-247.
- Kozma, A., & Stones, M. J. (1987) Social desirability in measures of subjective well-being: a systematic evaluation. *Journal of Gerontology*, 42, 56-59.
- Kunin, J. D. (1998) Predictors of psychosocial and behavioral adjustment of children: a study comparing children raised by lesbian parents to children raised by heterosexual parents. *Dissertation Abstracts International, Section B: Science and Engineering*, 59(06), 3094.
- Kurdek, L. A. (2005) What do we know about gay and lesbian couples? *Current Directions in Psychological Science*, 14, 251-254.
- Kuvalanka, K. A., & Goldberg, A. E. (2009) "Second generation" voices: queer youth with lesbian/bisexual mothers. *Journal of Youth and Adolescence*, 38, 904-919.
- LaCour, M. J., & Green, D. P. (2014) When contact changes minds: an experiment on transmission of support for gay equality. *Science*, 346(6215), 1366-1369.
- Langbein, L., & Yost, M. A., Jr. (2009) Same-sex marriage and negative externalities. *Social Science Quarterly*, 90, 292-308.
- Lavner, J. A., Waterman, J., & Peplau, L. A. (2012) Can gay and lesbian parents promote healthy development in high-risk children adopted from foster care? *American Journal of Orthopsychiatry*, 82, 465-472.
- Leddy, A. A., Gartrell, N., & Bos, H. (2012) Growing up in a lesbian family: the life experiences of the adult daughters and sons of lesbian mothers. *Journal of GLBT Family Studies*, 8, 243-257.
- Lesthaeghe, R. (2010) The unfolding story of the Second Demographic Transition. *Population and Development Review*, 36, 211-251.
- Lesthaeghe, R. (2014) The second demographic transition: a concise overview of its development. *PNAS*, 111(51), 18112-18115.
- Lick, D. J., Patterson, C. J., & Schmidt, K. M. (2013) Recalled social experiences and current psychological adjustment among adults reared by gay and lesbian parents. *Journal of GLBT Family Studies*, 9, 230-253.
- Lick, D. J., Tornello, S. L., Riskind, R. G., Schmidt, K. M., & Patterson, C. J. (2012) Social climate for sexual minorities predicts well-being among heterosexual offspring of lesbian and gay parents. *Sexuality Research and Social Policy*, 9, 99-112.
- Llopis-Jensen, C. (2015, December 6) Schumm's father an expert in trial on adoption ban: Florida 2008 case sought to deter same-sex couples. *The Topeka Capital-Journal*, pp. A1, A8.
- Lo, S. L., Vroman, L. N., & Durbin, C. E. (2015) Ecological validity of laboratory assessments of child temperament: evidence from parent perspectives. *Psychological Assessment*, 27, 280-290.
- Mabry, C. R. (2005) Opening another exit from child welfare for special needs children: why some gay men and lesbians should have the privilege to adopt children in Florida. *St. Thomas Law Review*, 18, 269-324.
- MacCallum, F., & Golombok, S. (2004) Children raised in fatherless families from infancy: a follow-up of children of lesbian and single heterosexual mothers at early adolescence. *Journal of Child Psychology and Psychiatry*, 45, 1407-1419.
- Manning, W. D., Fetto, M. N., & Lamidi, E. (2014) Child well-being in same-sex families: review of research prepared for American Sociological Association Amicus Brief. *Population Research and Policy Review*, 33, 485-502.
- Marks, L. (2012) Same-sex parenting and children's outcomes: a closer examination of the American Psychological Association's brief on lesbian and gay parenting. *Social Science Research*, 41, 735-751.
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015) Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist*, 70, 487-498.
- McNutt, M. (2015a) Editorial expression of concern. *Science*, 348(6239), 1100.
- McNutt, M. (2015b) Editorial retraction. *Science*, 348(6239), 1100.
- Meezan, W., & Rauch, J. (2005) Gay marriage, same-sex parenting, and America's children. *The Future of Children*, 15, 97-115.
- Miller, J. A., Jacobsen, R. B., & Bigner, J. J. (1981) The child's home environment for lesbian vs. heterosexual mothers: a neglected area of research. *Journal of Homosexuality*, 7, 49-56.
- Miller, J. A., Mucklow, B. M., Jacobsen, R. B., & Bigner, J. J. (1980) Comparison of family relationships: homosexual versus heterosexual women. *Psychological Reports*, 46, 1127-1132.

- Monte, W. (2013) Where the Holy See and science agree: children do best in a stable natural family. *Ave Maria International Law Journal*, 2, 217-250.
- Morgan-Miller, N. W. (2002) An exploratory study of different types of violence presented in early Christian and Islamic documents. *Psychological Reports*, 91, 520-524.
- Moxley, V., Eggeman, K., & Schumm, W. R. (1986) An evaluation of the "Recovery of Hope" program. *Journal of Divorce*, 10, 241-261.
- Mucklow, B. M., & Phelan, G. K. (1979) Lesbian and traditional mothers' responses to adult response to child behavior and self-concept. *Psychological Reports*, 44, 880-882.
- Murray, P. D., & McClintock, K. (2005) Children of the closet: a measurement of the anxiety and self-esteem of children raised by a non-disclosed homosexual or bisexual parent. *Journal of Homosexuality*, 49, 77-95.
- Nederhof, A. J. (1985) Methods of coping with social desirability: a review. *European Journal of Social Psychology*, 15, 263-280.
- Negy, C., Pearte, C., & Lacefield, K. (2013) Young adults' attitudes toward polygamous marriage as a function of gender, attitudes toward same-sex marriage, and other personality constructs. *Marriage & Family Review*, 49, 51-82.
- Nickerson, R. S. (1998) Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175-220.
- Norman, P. S., Ray, J. G., Jr., Brachman, P. S., Plotkin, S. A., & Pagan, J. S. (1960) Serologic testing for anthrax antibodies in workers in a goat hair processing mill. *American Journal of Hygiene*, 72, 32-37.
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012) Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7, 615-631.
- Patterson, C. J. (1992) Children of lesbian and gay parents. *Child Development*, 63, 1025-1042.
- Patterson, C. J. (2005) Lesbian and gay parents and their children: summary of research findings. In American Psychological Association (Ed.), *Lesbian & gay parenting*. Washington, DC: American Psychological Association. Pp. 5-22.
- Patterson, C. J. (2009a) Children of lesbian and gay parents: psychology, law, and policy. *American Psychologist*, 64, 727-736.
- Patterson, C. J. (2009b) Lesbian and gay parents and their children: a social science perspective. In D. A. Hope (Ed.), *Contemporary perspectives on lesbian, gay, and bisexual identities: Nebraska Symposium on Motivation*. New York: Springer Science and Business Media. Pp. 141-182.
- Patterson, C. J. (2013) Children of lesbian and gay parents: psychology, law, and policy. *Psychology of Sexual Orientation and Gender Diversity*, 1, 27-34.
- Perrin, A. J., Cohen, P. N., & Caren, N. (2013) Responding to the Regnerus study: are children of parents who had same-sex relationships disadvantaged? A scientific evaluation of the no-differences hypothesis. *Journal of Gay & Lesbian Mental Health*, 17, 327-336.
- Peterson, N. (1984, April 30) Coming to terms with gay parents. *USA Today*, p. 3D.
- Phillips, D. L., & Clancy, K. J. (1972) Some effects of "social desirability" in survey studies. *American Journal of Sociology*, 77, 921-940.
- Plotkin, S. A., Brachman, P. S., Utell, M., Bumford, F. H., & Atchison, M. M. (1960) An epidemic of inhalation anthrax, the first in the twentieth century: I. Clinical features. *American Journal of Medicine*, 29, 992-1001.
- Potter, D. (2012) Same-sex parent families and children's academic achievement. *Journal of Marriage and Family*, 74, 556-571.
- Price, D. (2002) Islam and human rights: a case of deceptive first appearances. *Journal for the Scientific Study of Religion*, 41, 213-225.
- Puryear, D. (1983) A comparison of children of lesbian and single parent heterosexual mothers on three measures of socialization. Unpublished doctoral dissertation, The California School of Professional Psychology, Berkeley, CA. *Dissertation Abstracts International*, 40(07), 3418B-3419B.
- Raley, J. A. (2013) Adolescents with same sex parents: does it make a difference? *Adolescent Psychiatry*, 3, 329-334.
- RAND Corporation. (2005) *Examining possible causes of Gulf War Illness: RAND policy investigations and reviews of the scientific literature*. Santa Monica, CA: RAND National Defense Research Institute and RAND Health.
- Redding, R. E. (2008) It's really about sex: same-sex marriage, lesbian parenting, and the psychology of disgust. *Duke Journal of Gender Law & Policy*, 15, 127-193.
- Redding, R. E. (2013) Politicized science. *Society*, 50, 439-446.
- Regnerus, M. (2012a) How different are the adult children of parents who have same-sex relationships? Findings from the New Family Structures Study. *Social Science Research*, 41, 752-770.
- Regnerus, M. (2012b) Parental same-sex relationships, family instability, and subsequent life outcomes for adult children: answering critics of the New Family Structures Study with additional analyses. *Social Science Research*, 41, 1367-1377.
- Regnerus, M. (2012c, December 20) Porn use and supporting same-sex marriage. *Public Discourse*. Retrieved from <http://www.thepublicdiscourse.com/2012/12/7048/>.
- Regnerus, M. (2012d) Response to Paul Amato, David Eggebeen, and Cynthia Osborne. *Social Science Research*, 41, 786-787.
- Regnerus, M. (2014, August 11) Tracking Christian sexual morality in a same-sex marriage future. *Public Discourse*. Retrieved from www.thepublicdiscourse.com/2014/08/13667/.
- Reiss, I. L. (2014) Exploring the relation of values, power, and advocacy in American sexual science. *International Journal of Sexual Health*, 26, 1-12.
- Research Advisory Committee on Gulf War Veterans' Illnesses. (2014) *RAC-GWVI committee meeting minutes* (March 19). Washington, DC: Department of Veterans' Affairs.
- Rivers, I. (2000) Social exclusion, absenteeism and sexual minority youth. *Support for Learning*, 15, 13-18.
- Rivers, I., & Noret, N. (2008) Well-being among same-sex- and opposite-sex attracted youth at school. *School Psychology Review*, 37, 174-187.
- Rivers, I., Poteat, V. P., & Noret, N. (2008) Victimization, social support, and psychosocial functioning among children of same-sex and opposite-sex couples in the United Kingdom. *Developmental Psychology*, 44, 127-134.
- Robinson, J. P., Espelage, D. L., & Rivers, I. (2013) Developmental trends in peer victimization and emotional distress in LGB and heterosexual youth. *Pediatrics*, 131, 423-430.
- Rosenfeld, M. J. (2010) Nontraditional families and childhood progress through school. *Demography*, 47, 755-775.
- Rosenfeld, M. J. (2014) Couple longevity in the era of same-sex marriage in the United States. *Journal of Marriage and Family*, 76, 905-918.
- Rosnow, R. L., & Rosenthal, R. (1996) Computing contrasts, effect sizes, and counternulls on other people's published results: general procedures for research consumers. *Psychological Methods*, 1, 331-340.
- Sarantakos, S. (1996a) Children in three contexts: family, education, and social development. *Children Australia*, 21, 23-31.

- Sarantakos, S. (1996b) Same-sex couples: problems and prospects. *Journal of Family Studies*, 2, 147-163.
- Sarantakos, S. (1998) Sex and power in same-sex couples. *Australian Journal of Social Issues*, 2, 147-163.
- Sarantakos, S. (2000) *Same-sex couples*. Sydney, Australia: Harvard Press.
- Schneider, R., Jr. (2013) BTW. *The Gay & Lesbian Review*, 20(2: March/April), 8-9.
- Schumm, B., Jr. (1966) A mathematical model for wake-type fluid flow from turbine-type impellers. Unpublished doctoral dissertation, Univer. of Rochester, Rochester, NY. (UMI No. 6610824)
- Schumm, W. R. (1982) Integrating theory, measurement, and statistical analysis in family studies survey research. *Journal of Marriage and the Family*, 44(4), 983-998.
- Schumm, W. R. (1993) On publishing family research using "sophisticated" quantitative methodologies. *Marriage and Family Review*, 18(1/2), 171-175.
- Schumm, W. R. (2002a) Commentary on Morgan-Miller's "An exploratory study of different types of violence presented in early Christian and Islamic historical documents." *Psychological Reports*, 91, 520-524.
- Schumm, W. R. (2002b) The admiral's son with an independent spirit. In S. K. Steinmetz & G. W. Peterson (Eds.), *Pioneering paths in the study of families*. New York: Haworth. Pp. 635-659.
- Schumm, W. R. (2003a) A reanalysis of Price's "Islam and human rights: a case of deceptive first appearances." *Psychological Reports*, 93, 1335-1338.
- Schumm, W. R. (2003b) A statistical comparison of the apparent importance of various prophets within early Islamic and Christian historical documents. *Psychological Reports*, 93, 472-474.
- Schumm, W. R. (2003c) Interpreting past religious discrimination today. *Psychological Reports*, 93, 513-514.
- Schumm, W. R. (2003d) Themes observed in early Qur'anic material: a brief commentary on Michael Sells' text *Approaching the Qur'an*. *Psychological Reports*, 93, 1291-1294.
- Schumm, W. R. (2004a) A reply to Belkin's argument that ending the "Gay Ban" will not influence military performance. *Psychological Reports*, 95, 637-640.
- Schumm, W. R. (2004b) Anthrax vaccine and Gulf War illness symptoms in Captain Jean Tanner's Dover Air Force Base survey. *Medical Veritas*, 1, 163-165.
- Schumm, W. R. (2004c) Differential risk theory as a subset of social exchange theory: implications for making gay marriage culturally normative and for understanding stigma against homosexuals. *Psychological Reports*, 94, 208-210.
- Schumm, W. R. (2004d) Islam and the "universal" gender difference in religious commitment: a brief report in response to Stark (2002). *Psychological Reports*, 94, 1104-1106.
- Schumm, W. R. (2004e) Relative chronology of violent and non-violent themes in early Christian and Islamic historical documents. *Psychological Reports*, 94, 931-932.
- Schumm, W. R. (2004f) Relative proportion of ethical teaching versus historical events concerning Jesus in early Islamic and Christian historical documents. *Psychological Reports*, 94, 894-896.
- Schumm, W. R. (2004g) What was really learned from Tasker and Golombok's (1995) study of lesbian and single parent mothers? *Psychological Reports*, 94, 422-424.
- Schumm, W. R. (2005a) Empirical and theoretical perspectives from social science on gay marriage and child custody issues. *St. Thomas Law Review*, 18(2), 425-471.
- Schumm, W. R. (2005b) Variations in themes of asceticism, humility, and love among Muslim sayings attributed to Jesus. *Islamic Studies*, 44, 113-123.
- Schumm, W. R. (2005c) Was it statistically legitimate to combine data from the four textile mills in Brachman et al.'s (1962) study of the effectiveness of a human anthrax vaccine? *Medical Veritas*, 2, 342-343.
- Schumm, W. R. (2005d) Were the rights of human subjects violated at the Arms Mill in Manchester, New Hampshire in 1957 during the human anthrax vaccine trials? *Medical Veritas*, 2, 344-347.
- Schumm, W. R. (2006a) A discriminant analysis of Whissell's New Testament data: on the psychological trail of the author of Hebrews. *Psychological Reports*, 98, 274-276.
- Schumm, W. R. (2006b) Reflections from the social sciences on the possible causes of abusing Muslim prisoners of war. *American Journal of Islamic Social Sciences*, 23(4), 136-148.
- Schumm, W. R. (2006c) Neurologic adverse events associated with smallpox vaccination in the United States – response and comment on reporting of headaches as adverse events after smallpox vaccination among military and civilian personnel. *BMC Medicine*, 4:27 (online, 5 pages).
- Schumm, W. R. (2008) Re-evaluation of the "no differences" hypothesis concerning gay and lesbian parenting as assessed in eight early (1979-1986) and four later (1997-1998) dissertations. *Psychological Reports*, 103, 275-304.
- Schumm, W. R. (2009) Gay marriage and injustice. *The Therapist*, 21(3), 95-96.
- Schumm, W. R. (2010a) A comparison of citations across multidisciplinary psychology journals: case study of two independent journals. *Psychological Reports*, 106, 314-322.
- Schumm, W. R. (2010b) Children of homosexuals more apt to be homosexuals? A reply to Morrison and to Cameron based on an examination of multiple sources of data. *Journal of Biosocial Science*, 42, 721-742.
- Schumm, W. R. (2010c) Comparative relationship stability of lesbian mother and heterosexual mother families: a review of evidence. *Marriage & Family Review*, 46, 499-509.
- Schumm, W. R. (2010d) Evidence of pro-homosexual bias in social science: citation rates and research on lesbian parenting. *Psychological Reports*, 106, 374-380.
- Schumm, W. R. (2010e) How science is done. *Marriage & Family Review*, 46, 323-326.
- Schumm, W. R. (2010f) Statistical requirements for properly investigating a null hypothesis. *Psychological Reports*, 107, 953-971.
- Schumm, W. R. (2011a) Are two lesbian parents better than a mom and dad? Logical and methodological flaws in recent studies affirming the superiority of lesbian parenthood. *Ave Maria Law Journal*, 10, 79-120.
- Schumm, W. R. (2011b) Child outcomes associated with lesbian parenting: comments on Biblarz and Stacey's (2010) report. *Journal of Human Sexuality*, 3, 35-80.
- Schumm, W. R. (2012a) Lessons for the "devilish statistical obfuscator" or how to argue for a null hypothesis: a guide for students, attorneys, and other professionals. *Innovative Teaching*, 1, 2 (online, 13 pages).
- Schumm, W. R. (2012b) Methodological decisions and the evaluation of possible effects of different family structures on children: the New Family Structures Survey (NFSS). *Social Science Research*, 41, 1357-1366.
- Schumm, W. R. (2012c) Re-examining a landmark research study: a teaching editorial. *Marriage and Family Review*, 48, 465-489.

- Schumm, W. R. (2013) Intergenerational transfer of parental sexual orientation and other myths. *International Journal of the Jurisprudence of the Family*, 4, 267-433.
- Schumm, W. R. (2014) Challenges in predicting child outcomes from different family structures. *Comprehensive Psychology*, 3, 10 (online, 12 pages).
- Schumm, W. R. (2015a) Same-sex marriage and negative externalities revisited. *Comprehensive Psychology*. Manuscript submitted for publication.
- Schumm, W. R. (2015b) Same-sex parenting and adoption. *Comprehensive Psychology*. Manuscript submitted for publication.
- Schumm, W. R. (2015c) Sarantakos's research on same-sex parenting in Australia and New Zealand: its significance, substance, and corroboration with research from the United States. *Comprehensive Psychology*, 4, 16 (online, 29 pages).
- Schumm, W. R. (in press) The legal recognition of same-sex marriage produces marital inequality. *International Journal of the Jurisprudence of the Family*.
- Schumm, W. R., Akagi, C. A., & Bosch, K. R. (2008) Relationship satisfaction for heterosexual women compared to lesbians and men in a sample of faith communities from Topeka, Kansas. *Psychological Reports*, 102, 377-388.
- Schumm, W. R., Bell, D. B., & Gade, P. A. (2000) Effects of a military overseas peacekeeping deployment on marital quality, satisfaction, and stability. *Psychological Reports*, 87, 815-821.
- Schumm, W. R., Bollman, S. R., & Jurich, A. P. (1981) The validity of Edmonds' Marital Conventionalization Scale. *Journal of Psychology*, 109, 65-71.
- Schumm, W. R., Bollman, S. R., & Jurich, A. P. (1982) The "marital conventionalization" argument: implications for the study of religiosity and marital satisfaction. *Journal of Psychology & Theology*, 19(3), 236-241.
- Schumm, W. R., & Brenneman, R. L. (2004) How adequate and well-controlled was the clinical trial of a human anthrax vaccine, 1955-1959? *Medical Veritas*, 1, 166-170.
- Schumm, W. R., Brenneman, R. L., Arieli, B., Mayo-Theus, S., & Muhammad, J. (2004) A statistical reanalysis of Brachman et al.'s (1962) study of a human anthrax vaccine. *Medical Veritas*, 1, 171-178.
- Schumm, W. R., & Crawford, D. W. (in press) Violations of fairness in social science research: the case of same-sex marriage and parenting. *International Journal of Jurisprudence of the Family*.
- Schumm, W. R., Ferguson, A. D., Hashmat, M. S., & New, T. L. (2005) Differences in paradox between Islam and Christianity: a statistical comparison. *Journal of Islam and Christian-Muslim Relations*, 16, 167-185.
- Schumm, W. R., Figley, C. R., & Fuhs, N. N. (1980) Similarity in self-esteem as a function of duration of marriage among student couples. *Psychological Reports*, 47, 365-366.
- Schumm, W. R., Figley, C. R., & Fuhs, N. N. (1981) Predicting self-disclosure anxiety in the marital relationship. *Journal of Psychology*, 107, 273-279.
- Schumm, W. R., Figley, C. R., & Jurich, A. P. (1979) The dimensionality of the Marital Communication Inventory: a preliminary factor analytic study. *Psychological Reports*, 45, 123-128.
- Schumm, W. R., Hess, J. L., Bollman, S. R., & Jurich, A. P. (1981) Marital conventionalization revisited. *Psychological Reports*, 49, 607-615.
- Schumm, W. R., Jeong, G. J., & Silliman, B. (1990) Protestant fundamentalism and marital success revisited. *Psychological Reports*, 66, 905-906.
- Schumm, W. R., Jurich, A. P., Bollman, S. R., & Bugaighis, M. A. (1985) His and her marriage revisited. *Journal of Family Issues*, 6, 221-227.
- Schumm, W. R., Jurich, A. P., Bollman, S. R., Webb, F. J., & Castelo, C. S. (2005) The long term safety of anthrax vaccine, pyridostigmine bromide (PB) tablets, and other risk factors among Reserve Component veterans of the first Persian Gulf War. *Medical Veritas*, 2, 348-362.
- Schumm, W. R., Jurich, A. P., Webb, F. J., Bollman, S. R., Reppert, E. J., & Castelo, C. S. (2005a) Changes in the subjective health of Reserve Component Veterans as a function of mobilization status during the first Persian Gulf War. *Medical Veritas*, 2, 336-341.
- Schumm, W. R., Jurich, A. P., Webb, F. J., Bollman, S. R., Reppert, E. J., & Castelo, C. S. (2005b) Exposure to anthrax vaccination and pyridostigmine bromide tablets as associated with geographic location during the first Persian Gulf War. *Medical Veritas*, 2, 521-525.
- Schumm, W. R., Jurich, A. P., Webb, F. J., Bollman, S. R., Reppert, E. J., Sanders, D., & Castelo, C. S. (2007) Gulf War illness classifications and changes in subjective health among Reserve Component veterans who had deployed to the Persian Gulf. *Medical Veritas*, 4, 1, 1414-1417.
- Schumm, W. R., & Kirn, J. E. (1982) Evaluating equity in the marital relationship: an alternative approach. *Psychological Reports*, 51, 759-762.
- Schumm, W. R., & Kohler, A. L. (2006) Social cohesion and the five pillars of Islam: a comparative perspective. *American Journal of Islamic Social Sciences*, 23(2), 126-136.
- Schumm, W. R., Martin, M. J., Jurich, A. P., & Bollman, S. R. (1982) Classifying family violence. *Journal of Family Issues*, 3, 319-340.
- Schumm, W. R., McCollum, E. E., Bugaighis, M. A., Jurich, A. P., Bollman, S. R., & Reitz, J. (1988) Differences between Anglo and Mexican-American family members on satisfaction with family life. *Hispanic Journal of Behavioral Sciences*, 10, 39-53.
- Schumm, W. R., & Nass, M. (2006) The FDA's acceptance of Brachman's 1950's anthrax research: good politics? Maybe. Good science? No. *Medical Veritas*, 3, 1, 747-752.
- Schumm, W. R., Nazarinia, R. R., & Bosch, K. R. (2009) Unanswered questions and ethical issues concerning U.S. biodefence research. *Journal of Medical Ethics*, 35, 594-598.
- Schumm, W. R., Nazarinia Roy, R. R., & Theodore, V. P. (2012) Separation and divorce. In B. A. Moore (Ed.), *Handbook of counseling military couples*. New York: Routledge. Pp. 157-181.
- Schumm, W. R., Pratt, K. K., Hartenstein, J. L., Jenkins, B. A., & Johnson, G. A. (2013) Determining statistical significance (alpha) and reporting statistical trends: controversies, issues, and facts. *Comprehensive Psychology*, 2, 10, (online, 6 pages).
- Schumm, W. R., Reppert, E. J., Jurich, A. P., Bollman, S. R., Castelo, C. S., Sanders, D., & Webb, F. J. (2001) Pyridostigmine bromide and the long-term subjective health status of a sample of female Reserve Component Gulf War veterans: a brief report. *Psychological Reports*, 88, 306-308.
- Schumm, W. R., Reppert, E. J., Jurich, A. P., Bollman, S. R., Webb, F. J., Castelo, C. S., Stever, J. C., Kaufman, M., Deng, L., Krehbiel, M., Owens, B. L., Hall, C. A., Brown, B. F. C., Lash, J. F., Fink, C. J., Crow, J. R., & Bonjour, G. N. (2002) Pyridostigmine bromide and the long-term subjective health status of a sample of over 700 male Reserve Component Gulf War era veterans. *Psychological Reports*, 90, 707-721.
- Schumm, W. R., Reppert, E. J., Jurich, A. P., Bollman, S. R., Webb, F. J., Castelo, C. S., Stever, J. C., Sanders, D., Bonjour, G. N., Brown, B. F. C., Hall, C. A., Owens, B. L., Krehbiel, M., Deng, L., & Kaufman, M. (2002) Self-reported changes in subjective health and anthrax vaccination as reported by over 900 Persian Gulf War era veterans. *Psychological Reports*, 90, 639-653.

- Schumm, W. R., Southerly, W. T., & Figley, C. R. (1980) Stumbling block or stepping stone: path analysis in family studies. *Journal of Marriage and the Family*, 42, 251-262.
- Schumm, W. R., Turek, D., & McCarthy, D. (2003) Evaluating an all-ranks military staff ride. *Psychological Reports*, 93, 1156-1158.
- Schumm, W. R., Webb, F. J., Bollman, S. R., Jurich, A. P., Reppert, E. J., Castelo, C. S., & Stever, J. A. (2004) Predicting self-reported exposure to nerve agents by Reserve Component personnel during the first Persian Gulf War. *Psychological Reports*, 94, 989-992.
- Schumm, W. R., Webb, F. J., Castelo, C. C., Akagi, C. G., Jensen, E. J., Ditto, R. M., Spencer-Carver, E., & Brown, B. (2002) Enhancing learning in statistics classes through the use of concrete historical examples. *Teaching Sociology*, 30, 361-375.
- Schumm, W. R., Webb, F. J., Jurich, A. P., & Bollman, S. R. (2002) Comments on the Institute of Medicine's 2002 report on the safety of anthrax vaccine. *Psychological Reports*, 91, 187-191.
- Sherkat, D. E. (2012) The editorial process and politicized scholarship: Monday morning editorial quarterbacking and a call for scientific vigilance. *Social Science Research*, 41, 1346-1349.
- Siegel, B. S., Perrin, E. C., Dobbins, M. I., Lavin, A., Mattson, G., Pascoe, J., & Yogman, M. (2013) Promoting the well-being of children whose parents are gay or lesbian. *Pediatrics*, 131, 827-830.
- Sirota, T. H. (1997) A comparison of adult attachment style dimensions between women who have gay or bisexual fathers and women who have heterosexual fathers. Unpublished doctoral dissertation, School of Education, New York Univer., Garden City, New York. *Dissertation Abstracts International*, 58(01), 138B. (UMI No. 9718726).
- Stacey, J., & Biblarz, T. J. (2001) (How) does the sexual orientation of parents matter? *American Sociological Review*, 66, 159-183.
- Stanger-Hall, K. F., & Hall, D. W. (2011) Abstinence-only education and teen pregnancy rates: why we need comprehensive sex education in the U.S. *PLOS One*, 6(10), e24658 (11 pages).
- Stevens, M., Perry, B., Burston, A., Golombok, S., & Golding, J. (2003) Openness in lesbian-mother families regarding mother's sexual orientation and child's conception by donor insemination. *Journal of Reproductive and Infant Psychology*, 21, 347-362.
- Strasser, M. (2011) Same-sex marriage and the right to privacy. *Journal of Law & Family Studies*, 13, 117-150.
- Stroebe, W., Postmes, T., & Spears, R. (2012) Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, 7, 670-688.
- Studenmund, A. H. (2010) *Using econometrics: a practical guide*. (6th ed.) Upper Saddle River, NJ: Prentice Hall.
- Sullins, D. P. (2015a) Bias in recruited sample research on children with same-sex parents using the Strength and Difficulties Questionnaire (SDQ). *Journal of Scientific Research and Reports*, 5(5), 375-387.
- Sullins, D. P. (2015b) Child attention-deficit hyperactivity disorder (ADHD) in same-sex parent families in the United States: prevalence and comorbidities. *British Journal of Medicine and Medical Research*, 6(10), 987-998.
- Sullins, D. P. (2015c) Emotional problems among children with same-sex parents: difference by definition. *British Journal of Education, Society, and Behavioural Science*, 7, 99-120.
- Sullins, D. P. (2015d) The unexpected harm of same-sex marriage: a critical re-appraisal, replication, and re-analysis of Wainright and Patterson's studies of adolescents with same-sex parents. *British Journal of Education, Society & Behavioural Science*, 11(2), 1-22.
- Swank, E., Woodford, M. R., & Lin, C. (2013) Antecedents of pro-LGBT advocacy among sexual minority and heterosexual college students. *Sexuality Research and Social Policy*, 10, 317-332.
- Tasker, F. (2010) Same-sex parenting and child development: reviewing the contribution of parental gender. *Journal of Marriage and Family*, 72, 35-40.
- Tasker, F., & Golombok, S. (1995) Adults raised as children in lesbian families. *American Journal of Orthopsychiatry*, 65, 203-215.
- Tasker, F., & Golombok, S. (1997) *Growing up in a lesbian family: effects on child development*. New York: Guilford Press.
- Tornello, S. L., & Patterson, C. J. (2012) Gay fathers in mixed-orientation relationships: experiences of those who stay in their marriages and of those who leave. *Journal of GLBT Family Studies*, 8, 85-98.
- Turner, S. (2015) *Book review: The sacred project of American sociology by Christian Smith (2014)*. Oxford, UK: Oxford Univer. Press.
- Umberson, D., Thomeer, M. B., Kroeger, R. A., Lodge, A. C., & Xu, M. (2015) Challenges and opportunities for research on same-sex relationships. *Journal of Marriage and Family*, 77, 96-111.
- Vittinghoff, E., & McCulloch, C. E. (2007) Relaxing the rule of ten events per variable in logistic and Cox regression. *American Journal of Epidemiology*, 165, 710-718.
- Wainright, J. L., & Patterson, C. J. (2006) Delinquency, victimization, and substance use among adolescents with female same-sex parents. *Journal of Family Psychology*, 20, 526-530.
- Wainright, J. L., & Patterson, C. J. (2008) Peer relations among adolescents with female same-sex parents. *Developmental Psychology*, 44, 117-126.
- Wainright, J. L., Russell, S. T., & Patterson, C. J. (2004) Psychosocial adjustment, school outcomes, and romantic relationships of adolescents with same-sex parents. *Child Development*, 75, 1886-1898.
- Warner, R. M. (2013) *Applied statistics: from bivariate through multivariate techniques*. (2nd ed.) Los Angeles, CA: Sage.
- Wicherts, J. M., Bakker, M., & Molenaar, D. (2011) Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLOS One*, 6(11), e26828.
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006) The poor availability of psychological research data for reanalysis. *American Psychologist*, 61, 726-728.
- Wilkinson, L., and the Task Force on Statistical Inference, APA Board of Scientific Affairs. (1999) Statistical methods in psychology journals: guidelines and explanations. *American Psychologist*, 54, 594-604.
- Williams, J. (2011) *Muzzled: the assault on honest debate*. New York: Crown Press.
- Wood, P. (2013) The campaign to discredit Regnerus and the assault on peer review. *Academic Questions*, 26, 171-181.
- Yancey, G. (2015) Both/and instead of either/or. *Society*, 52, 23-27.
- Zhou, Q., Eisenberg, N., Wang, Y., & Reiser, M. (2004) Chinese children's effortful control and dispositional anger/frustration: relations to parenting styles and children's social functioning. *Developmental Psychology*, 40, 352-366.

APPENDIX

Sample items for three types of social desirability response bias with answers of true or false

Individual bias

I never have lost my temper.
I send money to every charity that asks me for a donation.
I am always agreeable with whatever someone else has to say.
I seldom ever make mistakes.

Relationship bias

My marriage/relationship is absolutely perfect.
My partner and I have never done anything to irritate each other.
The thought of breaking up because of a disagreement has never occurred to either of us.
My partner and I are always especially kind and thoughtful to each other, no matter how tired or frustrated we might be.

Parental bias (Parent report)

My child(ren) have never misbehaved or done anything I felt to be frustrating to me.
My relationship with my child(ren) is virtually perfect.
My child(ren) always listen to what I say and do what I want them to do.
My child(ren) have never been disrespectful in word or attitude towards their parent(s).
My child(ren) have never embarrassed me in public by their misbehavior.
No matter what I have thought, my child(ren) have always agreed with me.

Parental bias (Adult child report)

My parents were always fair to me.
My parents never argued with me or raised their voices at me in frustration or anger.
My parents were never harsh in their discipline toward me.
My parents never argued with each other in a disrespectful manner.
My parents never did anything that ever made me feel angry with them.
My parents were the most perfect parents ever.
