Journal of Scientific Research & Reports
5(5): 375-387, 2015; Article no.JSRR.2015.104
ISSN: 2320-0227

SCIENCEDOMAIN *international*
www.sciencedomain.org

Bias in Recruited Sample Research on Children with Same-Sex Parents Using the Strength and Difficulties Questionnaire (SDQ)

D. Paul Sullins[1*]

[1] *Department of Sociology, The Catholic University of America, USA.*

Author's contribution

The sole author designed, analyzed and interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/JSRR/2015/15298
Editor(s):
(1) Thomas Müller, Department of Neurology, St. Joseph Hospital Berlin-Weissensee, Berlin, Germany.
Reviewers:
(1) E. Okwaraji Friday, Department of Psychological Medicine, College of Medicine, University of Nigeria.
(2) Anonymous, Brazil.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=751&id=22&aid=7750

Original Research Article

Received 18th November 2014
Accepted 17th December 2014
Published 10th January 2015

ABSTRACT

Aims: To test for the presence of bias on use of a common psychometric instrument, the Strengths and Difficulties Questionnaire (SDQ), in studies of children with same-sex parents using a recruited convenience sample.
Study Design: Non-parametric quasi-experimental two-group comparison.
Methodology: Results from five qualifying studies, two with random samples and three with recruited samples, were compared with normative population data, assessing the percent of comparisons favorable or unfavorable to children with same-sex parents for six subscale measures.
Results: In the recruited samples 79.3 % (range: 75-83) of comparisons were favorable to children with same-sex parents, compared with no favorable comparisons (0%, range 0-0) in the random samples. Two additional random samples with related measures were also adduced, also with no favorable comparisons (0%, range 0-0).
Conclusion: Evidence suggests strong bias resulting in false positive outcomes for parent-reported SDQ in recruited samples of same-sex parents.

**Corresponding author: Email: psullins@gmail.com;*

Keywords: Same-sex parents; child emotional problems; strengths and difficulties questionnaire (SDQ); survey bias.

1. INTRODUCTION

Dozens of studies in past decades have suggested that children with same-sex parents (hereafter for brevity “SS children”; “SS” indicates “same-sex” or homosexual) suffer no significant disadvantages in well-being compared to children with opposite-sex parents (hereafter “OS children”; “OS” indicates “opposite-sex” or heterosexual; some prefer “different-sex”).Recent competing reviews of this literature, however, have agreed that the field has been dominated by small convenience samples recruited from GLBT (acronym for “Gay, Lesbian, Bi-sexual, Transgender” as a blanket term for members of minority sexual orientations; also sometimes “LGBT”) interest, advocacy or support groups. Allen, a critic of the “no differences” claim, found that only four of 49 studies published prior to 2010 drew a random or probability sample of SS parents. The remainder based their findings on conveniently available or selected groups of participants, often recruited from biased, politically aware sources such as “LGBT events, bookstore and newspaper advertisements, word of mouth, networking and youth groups” [1,2,3]. Anderssen et al. [4], reviewing an earlier set of studies, reported that for only one out of 23 studies of child outcomes with same-sex parents prior to 2000 was the aim of the study unknown to participants. Reviews by Rosenfeld [5] and Manning [3], while defending “no differences”, do not dispute these facts.

Critics have argued that the “no differences” finding may result, at least in part, from sample bias. As with any convenience sample, results are subject to accessibility bias (persons more difficult to contact are less likely to be sampled), suggested in this case by the larger proportion of urban respondents in same-sex samples. Further, since a large proportion of researchers in this area are also advocates for social change regarding LGBT issues and/or are LGBT themselves, researcher bias is distinctly possible. To date, there have been no blind studies in this field, and most research has been funded or co-funded by groups favorable to LGBT issues. Since sample members have been recruited, in almost all studies, based on an appeal to the study goals and design, sample ascertainment bias, which occurs when respondents who are aware of the purposes of a study differentially self-select into the sample, may also be present. Finally, SS parents, aware that a favorable study outcome can help certify to critics the acceptability of their stigmatized family form, may be more likely than OS parents to exhibit the social desirability bias that is an unavoidable feature of any parent-reported child measure.

Defenders of recruited samples in this literature acknowledge the possibility of bias, but point out that the difficulties of obtaining a probability sample of the small, stigmatized population involved are prohibitive [3,6]. In most populations less than three of a thousand households with children are headed by SS parents; Rosenfeld [5] characterizes the group as a needle in a haystack. Samples recruited through LGBT organizations and friendship networks may not be optimum, it is argued, but they are the best that can be reasonably attained. Indeed, almost every study in this literature using a nonprobability sample clearly admits its limitations and weaknesses, but also maintains that, as with many small or deviant populations, without the use of nonprobability samples we would be unlikely to learn anything at all about the characteristics of the group.

Is the finding of “no differences” due to bias related to sample recruitment? Two recent studies with sharply differing results provide an opportunity to examine this question in a focused way. Both studies use the identical metric—the parent-reported Strengths and Difficulties Questionnaire (SDQ)—to document contradictory findings on the effect of SS parenting on child emotional health. The first, a report by Crouch and colleagues on a large, well-crafted convenience sample of SS parents, finds “no evidence to support a difference in parent-reported child health” between OS parents and SS parents [6]. The second, an analysis by Sullins of a large multi-stage random population sample, finds that “children in same-sex families are at least twice as likely to experience serious emotional problems compared to their counterparts in opposite-sex families” [7]. These studies, which arrive at different results on the same measure in different sample types, form a kind of natural experiment to address the question of sample bias in SS parenting research.

The present study proceeds by comparing the results of these two studies with their respective reference populations to observe the possibility of bias due to sample design. The analysis is then extended to other similar studies in the SS parenting literature.

2. DATA AND METHODS

The Strength and Difficulties Questionnaire (SDQ) is a widely-used screening instrument which has been demonstrated to be a robust predictor of child mental health distress in diverse populations [8,9]. The SDQ consists of 5 items each covering five domains: emotional symptoms, conduct problems, hyperactivity-inattention, peer problems and prosocial behaviors. Each item is scored 0-2 for responses of “Not true”, “Somewhat true”, or “Certainly true”, producing a scale score ranging from 0 to 10 for each domain. The sum of the scores for the first four domains (excluding prosocial behaviors) forms a “Total Difficulties” score ranging from 0 to 40. Scores above a cutoff point indicate abnormal or elevated difficulties for each domain; a score above 17 for the total difficulties scale has been found to predict “a significantly increased probability of meeting criteria for a DSM-IV disorder” [10]. More information, including the SDQ instrument, scoring standards, validation studies and normative data for many countries is available online at www.sdqinfo.com. The present study makes use of the normative data for USA and Australia [11].

The Australian Study of Child Health in Same-sex Families (ACHESS) collected a convenience sample of SS parent reports in late 2012 with the aim of “determin[ing] the complete physical, mental and social wellbeing of Australian children with at least one same-sex attracted parent” [12]. Reportedly, efforts were made to recruit same-sex parent families, particularly male couples, who are often under-represented in such studies. Attaining 315 completed cases, the study achieves one of the largest targeted samples of this population to date. Unlike most studies in the field, ACHESS employed standard well-validated measures of child health such as the SDQ.

The United States’ National Health Interview Survey (NHIS), a project of that government’s Centers for Disease Control (CDC), has annually interviewed between 35,000 and 40,000 households selected by means of a complex multistage probability sample, yielding information on 75,000 to 100,000 individuals that are statistically representative of the civilian noninstitutionalized population of the United States. Extensive health and demographic information is collected for all household members. In addition, for each family that includes children under age 18, detailed supplemental health information, provided by a parent or other knowledgeable adult informant, is collected for one child chosen at random (the “sample child”).

The NHIS interview constructs a family roster which identifies household members who are spouses or cohabiting sexual partners, permitting the identification of same-sex partner couples. Sullins compared SS and OS children on the SDQ scales for child emotional health, following the example of prior studies using NHIS data to compare SS and OS couples regarding such health issues cigarette smoking and breast cancer risk [13-16].

3. RESULTS

The original data from both studies have been adapted to facilitate comparison of their findings (Table 1). Sullins reported on combined data from 2000-2013 based on an abbreviated version of the SDQ that was fielded in most of those years. Table 1 includes only data from the years 2001, 2003 and 2004, when the full twenty-item SDQ was used for NHIS. Crouch et al. reported mean SDQ scores from ACHESS separately for boys and girls; and do not include the OS norms for comparison but cite the original study that does report them. Both studies apply multiple statistical controls to the comparison of family types, which have been bypassed or removed to facilitate the present comparison. Table 1, therefore, permits us to observe the raw or unadjusted means for the SDQ domain scales in each study, differing only by population and sample type.

Differences between Australian and US SDQ norms are small, and largely orthogonal to the comparisons of interest in the present study. The NHIS data for OS parents in Table 1 is virtually identical to the US SDQ norms, which are derived from the 2001 NHIS. The US scale scores are systematically smaller, by about 20%, than the Australian scores, for reasons which are beyond the scope of the present study. What is pertinent here is not the absolute value of the scale scores, but the relative comparison of OS and SS parents on each survey.

Table 1. Mean (SD) SDQ domains comparing OS children and SS children in NHIS and ACHESS data

SDQ domains	NHIS			Mellor/ACHESS		
	OS children	**SS children**	**P**	**OS children**	**SS children**	**P**
Conduct	1.21(1.61)	2.06***(2.19)	.001	1.5(1.6)	1.44(1.59)	.57
Hyperactivity	2.72(2.51)	3.62**(2.96)	.01	3.1(2.4)	3.01(2.45)	.54
Emotional	1.47(1.85)	1.64(2.05)	.48	2.1(2.0)	1.61***(1.89)	<.001
Peer	1.29(1.51)	1.65*(1.63)	.03	1.6(1.9)	1.42(1.68)	.12
Pro-social (low)	1.25(1.69)	1.98**(2.16)	.01	1.7(1.7)	1.98*(1.99)	.02
Total Difficulties	6.65(5.68)	9.00***(6.68)	.001	8.18(6.1)	7.48(5.49)	.06
N	27,327	71		941	299	

* p <= .05 ** p <= .01 *** p <= .001, OS = Opposite-sex parents; SS = Same-sex parents; P = p-value of the OS/SS difference by t-test, NHIS = National Health Interview Survey (USA) ACHESS = Australian Child Health in Same Sex Families, All significance tests were independently assessed, and may not agree with reported results from the study , involved

For almost all the SDQ domains, a striking and systematic contrast can be observed between the two samples. On the NHIS, SS children have *higher* emotional problems than do OS children on every measure, a difference that is statistically significant for five of the six domains. On ACHESS, SS children have *lower* emotional problems than do OS children on every measure except one; the difference is statistically significant for emotional problems. Except as a general indicator of magnitude, statistical significance is not strictly relevant to this analysis, since here we are comparing samples, not populations; and, as discussed below, significance may not be meaningful for ACHESS generally.

Sampling differences between the OS and SS samples in the ACHESS comparison are discounted by Crouch et al. [6], who in any event adopt the Mellor data for normative comparison. Since it is highly unlikely that the NHIS, which during the years in question did not even inquire about sexual orientation, is biased with regard to SS children, the contrast in findings between NHIS and ACHESS must imply either that SS children suffer higher emotional problems than do OS children in the USA, but not in Australia; or that the ACHESS sample reflects bias in favor of reporting better outcomes for SS children than is actually the case.

The former possibility seems unlikely given the large cultural and social similarities between Australia and the USA, and in light of the fact that same-sex couples enjoy more liberal legal and political status in the USA than they do in Australia; but it cannot be dismissed out of hand. Bos suggests, for example, that higher stigmatization for SS children in the US than in Holland may account, in part, for the lower rate of problem behaviors among the Dutch SS children. However, Sullins documents that, on NHIS combined data from 2000-2013, SS children experience less overall stigmatization than do OS children [7]. If the higher emotional problems observed for SS children in the U.S. compared to those in Australia were due to stigmatization, therefore, we would expect OS children in the U.S. also to experience higher emotional problems than their Australian counterparts; but, as Table 1.shows, the opposite is the case. On all the SDQ domains, OS children in the U.S. have lower means for emotional difficulties than do OS children in Australia.

Thus, although the possibility of substantive national differences cannot be definitively ruled out, it seems much more likely that the differences observed reflect systematic differences in the two samples. Indeed, this likelihood is explicitly supported by Crouch et al. [6]: "The self-selection of our convenience sample has the potential to introduce bias that could distort results. Ifsystematic bias was at play however, it would be anticipated that all outcome variables would demonstrate [more favorable] scores across the sample". That is, in fact, precisely the pattern observed (with one exception) in the present analysis.

3.1 Broadening Scope

So far we have observed a sharp contrast between the NHIS, which reports higher difficulties for SS children on every measure based on a representative population sample, and ACHESS, which reports lower emotional difficulties for SS children than for OS children on every measure except one based on a convenience sample of recruited participants. If this pattern is due to bias in the convenience sample compared to the probability sample, one would expect other sample contrasts in this literature to show a similar pattern of differences.

In addition to Sullins and Crouch, since 2000 three other studies of SS parenting have reported SDQ scale results for SS children. One of the three made use of a random population sample comparable to NHIS [17]; the other two used a recruited sample comparable to ACHESS [18,19]. Altogether, these five studies comprise all gay parenting studies that have used the SDQ (to my knowledge). Table 2 summarizes the results. The two leftmost studies presented in the table report the two random samples; the remaining three study results report the recruited samples. The rightmost of the four columns summarizing each study's findings (labeled "F/U") reports whether each SDQ domain comparison is Favorable or Unfavorable for SS children.

The dispositive issue for sample bias is whether the pattern already observed, in which the random sample reflects unfavorable findings for SS children but the recruited sample reflects favorable results, persists across the additional studies. Inspection of Table 2 reveals that the answer to this question is an emphatic "yes". In the other random sample, reporting on a different national population (England), SS children are found to have higher emotional problems for every SDQ domain without exception, just as on NHIS. As with ACHESS, however, participants in the recruited samples report generally more favorable results for SS children than for OS children. The distinction in results by sample type is clear, even across two studies by the same lead author.

Despite Crouch's claim cited above, there is no reason to suppose that sample bias is necessarily monotonic, or affects all studies or all behavioral domains to the same degree. The exceptions to the pattern of favorable random and unfavorable non-random results, moreover, are not randomly distributed. For four of the six SDQ domains—conduct problems, hyperactivity, emotional symptoms and total difficulties—there are no exceptions to the pattern; for the other two domains—peer problems and pro-social behavior—the results are mixed. There may be many reasons for this difference among the recruited samples that are consistent with the operation of bias, including the presence, to a greater or lesser degree, of more than one form of bias. Perhaps parents of children with more externalizing problems, such as misconduct or hyperactivity, were more likely to voluntarily deselect from the recruited sample; or perhaps parents' knowledge of peer problems or pro-social behavior on the part of their children, or their understanding of what constitutes social desirability in these areas, was less complete or clear than it was for the other domains. Whatever the reasons (and it is beyond the scope of this study to determine them), there is no contradiction in the supposition that bias, if it is at work, may itself be biased in its effect on study results.

To facilitate interpretation, Table 3 summarizes the proportion of favorable comparisons from each study in Table 2, presented in the first five rows of the table. The top two rows report on the two random samples; the next three rows present findings from the three recruited samples. For further context, Table 3 also adds findings from two other random sample studies of SS parenting prior to 2010 that measured child emotional or behavioral difficulties, but did not use the SDQ. Although some other studies in this developing literature included elements of random sampling, Wainright and Patterson produced the only three studies that made use of a statistically representative "blind" population sample of SS children, identifying them in the data of the National Longitudinal Survey of Adolescent Health. Two of these studies contained measures of emotional or behavioral distress for SS children compared with OS children. In 2004 they reported comparative findings for depression, self-esteem and anxiety [20], and in 2006 reported on eleven measures of delinquency among adolescents such as binge drinking, illegal drug use or risky sexual behavior [21]. In both studies all of the above measures were less favorable for SS children in the sample, extending the pattern already observed on the NHIS and Golombok's 2003 study. At this point, Table 3 includes, to my knowledge, all the unambiguously random data on SS children gathered prior to 2010.

In Table 3, an average difference *between* sample types that is greater than the difference *within* sample types would suggest the presence of systematic bias. It is easy to see from inspection that the difference between random samples (top two lines and bottom two lines) and recruited samples (middle three lines) is not only greater than the difference within those two categories, but also that the magnitude of the comparison is extremely large. Among the recruited samples, the percent of findings favorable to SS children, compared to OS reference children, varies by only 4 percentage points from an average just over 79; among the random samples there is no variation at all; and between these two groups of findings is a difference of almost 80 percentage points. Put another way, the difference between the random and recruited samples is almost 20 times as large as the variation among the recruited samples. This robust finding strongly suggests that a large and persistent bias toward favorable findings for SS children, compared to OS reference children, is present in the recruited samples relative to the random ones.

The presence of such large bias implies that recruited samples are not very accurate or credible sources for assessing the population characteristics of SS children. Academics, policymakers, jurists and health practitioners interested in evidence-based decision-making should demand random sample evidence to inform their deliberations regarding this population.

Random sample evidence may not always be feasible to obtain, however, or accurate population inferences may not be needed as a study outcome; in which cases it would be helpful to briefly examine more closely the types and causes of bias that may be present in recruited sample studies in this area, and consider how they might be reduced, especially for those readers who may not be applied quantitative researchers. Since ACHESS is one of the largest, most recent and carefully conceived recruited sample studies of this population, it will be helpful to focus our examination on the potential sources of bias in this study. If we can understand how an exemplary study like ACHESS may have unaddressed bias, we will better understand how any study of this type may be biased.

3.2 Sources of Bias

By definition, bias denotes any form of systematic error that inaccurately prejudices findings toward one outcome over others. Every research effort is susceptible to bias; the scientific method assumes the presence of potential bias and imposes procedures designed to minimize or specify it. Bias in survey research may be present in how the respondents answer the questions (response bias), how the sample is collected (selection bias), or how the results are interpreted (interpretation bias). These elements may be expressed in this literature, as discussed above, as social desirability bias in parental response, ascertainment or self-selection bias in the composition of the sample, or researcher bias in interpreting results. Each of these will be briefly addressed in turn.

3.2.1 Social Desirability Bias

Because the ACHESS SDQ scales depend on parent reports, their findings are subject to social desirability bias, that is, the tendency for parents to report more socially desirable information about their own children (or themselves) than is objectively the case. To correct for this tendency the SDQ is often administered to children's teachers and/or the children themselves, as well as parents, and the two sets of results are compared. Both teacher and self-report ratings tend to be lower than those of parents. Since the comparative OS data for the ACHESS scales are also based on parent reports, social desirability would only bias the comparisons if the ACHESS parents were more or less prone to give desirable responses than was the OS comparison group. ACHESS surveyed parents using paper or online survey instruments, while the comparison group of OS parents received telephone interviews; social desirability bias is known to be more common in interviews, where the respondent is speaking with another human being, than on written or computer surveys, where the respondent is more anonymous. On the other hand, the fact that all the SS respondents, but none of the OS comparison group, were aware of the purposes of the ACHESS study, and stood to benefit from any increased social acceptance or support that a positive picture of SS children might help to bring about, may well have increased favorable responses. We do not know how substantial such bias may be, but it would not be difficult to determine by including teacher and/or self-reported ratings on the ACHESS SDQ scales. The ACHESS researchers acknowledge the possibility of parent self-representation bias, and indicate that future research will include "child-reported measures of health", as well as parent interviews. The absence of such measures to date leaves uncertain the extent to which social desirability bias may be present.

Table 2. SDQ mean scores (SD) across five surveys of SS parenting

	US NHIS (2000-2004)				**Golombok 2003** [1]				**Bos 2010**				**Golombok 2014**				**Mellor/ACHESS 2014**			
SDQ Domains	**OS**	**SS**	**P**	**F/U**	**OS**	**SS**	**P**	**F/U**	**OS**	**SS**	**P**	**F/U**	**OS**	**SS**	**P**	**F/U**	**OS**	**SS**	**P**	**F/U**
Conduct problems	1.21(1.61)	2.06 ***(2.19)	.001	U	12.8	13.9	.55	U	1.19 (.25)	1.18(.33)	.89	F	2.90 (1.80)	2.12 * (2.08)	.03	F	1.5 (1.6)	1.44(1.59)	.57	F
Hyperactivity-inattention	2.72(2.51)	3.62 **(2.96)	.01	U	12.8	16.7	.86	U	1.60 (.48)	1.53(.50)	.55	F	5.18 (2.79)	4.33 (2.45)	.08	F	3.1 (2.4)	3.01(2.45)	.54	F
Emotional symptoms	1.47(1.85)	1.64 (2.05)	.48	U	4.5	8.3	.36	U	1.26 (.27)	1.19(.50)	.46	F	1.98 (1.77)	1.73 (2.21)	.48	F	2.1 (2.0)	1.61 ***(1.89)	<.001	F
Peer problems	1.29(1.51)	1.65 *(1.63)	.03	U	5.3	13.9	.07	U	--	--	--	--	1.51 (1.68)	1.76 (1.71)	.42	U	1.6 (1.9)	1.42 (1.68)	.12	F
Pro-social behavior	8.75(1.69)	8.02 **(2.16)	.01	U	3.0	8.3	.15	U	2.67 (.28)	2.67(.33)	1.0	N	--	--	--	--	8.3 (1.7)	8.02 *(1.99)	.02	U
Total difficulties	6.65(5.68)	9.00***(6.68)	.001	U	6.0	8.3	.61	U	--	--	--	--	18.4%	13.7%	.46	F	8.18 (6.06)	7.48(5.49)	.06	F
Sample type	Ran	Ran			Ran	Ran			Rec	Rec			Rec	Rec			Ran	Rec		
N	27,256	71			133	36			36	36			49	81			941	299		

** p <= .05 ** p <= .01 *** p <= .001, SDQ = Strengths and Difficulties Questionnaire; NHIS = National Health Interview Survey; OS = Opposite-sex parents; SS = Same-sex parents; P = p-value of the OS/SS difference by t-test or chi-square test; F/U = Favorable/Unfavorable; Ran = Random; Rec = Recruited. All significance tests were independently assessed, and may not agree with reported results from the study involved. [1] Values reported for this study are percent abnormal, not mean; p-values reported are for chi square tests.*

Table 3. Correspondence between sample type (random or recruited) and comparison type (favorable or unfavorable)

Study/analysis	**Same-sex sample type**	**Measure**	**Percent favorable SS comparisons**
US NHIS (2000-2004)	Random (Statistically representative)	SDQ	0
Golombok 2003	Random (Statistically Representative)	SDQ	0
Bos 2010	Recruited (Non-representative)	SDQ	75
Golombok 2013	Recruited (Non-representative)	SDQ	80
Mellor/ACHESS 2014	Recruited (Non-representative)	SDQ	83
Other random-sample studies			
Wainright and Patterson 2006	Random	Eleven measures of adolescent delinquency	0
Wainright and Patterson 2004	Random	Three measures of depression, self-esteem and anxiety	0

3.2.2 Sample selection bias

There is no question that the ACHESS sample was distorted by selection bias, since respondents were invited to volunteer for the study through general announcements. As most introductory statistics texts point out, a sample comprised only of self-selected individuals is unavoidably biased, since those who choose not to participate are almost certainly different in systematic ways from those who choose to participate [22]. Smith, for example, states definitively: "*Voluntary response samples are always biased*: they only include people who choose to volunteer, whereas a random sample would need to include people whether or not they choose to volunteer" (emphasis in original) [23]. Not only was the ACHESS sample recruited, but it was recruited 1) through open public appeals for participation 2) disseminated among politically aware groups whose participants were clearly "interested", in a Marxian sense, in the outcome of the ACHESS study. These two factors undoubtedly increased bias, and could have easily been corrected, or could be corrected in similar future samples.

Prior research has directly noted the presence of bias in recruited samples of same-sex populations. Bailey, in a study of the Australian Twin Registry that led to the rejection of the conclusions of his own earlier study of homosexual men [24] which had "recruited [participants] by means of advertisements in homophile publications or by word of mouth", observed that " such sampling is likely to result in volunteer bias" leading to inaccurate results. "Furthermore," he added, "respondents with exclusively homosexual orientations may be overrepresented, and those with modest levels of homosexual attraction, underrepresented, obscuring the potentially continuous nature of sexual orientation" [25]. Cochran and colleagues likewise observed that, before the inclusion of sexual orientation measures in large public health surveys in the United States, earlier mental health research on minority sexual orientations was "plagued" by "the usual problems of sampling bias or absent heterosexual control groups" [17].

The ACHESS protocol stated: "Primarily recruitment will be through emails posted on gay and lesbian community email lists aimed at same-sex parenting" [24]. The word "lists" here does not refer to an actual sample frame. What appears to have been done is that announcements were published on the two websites named and affiliated blogs. This is consistent with the other recruitment activities called for in the protocol: advertisements, media releases, flyers and researcher appearances at public events. This procedure increased probable bias in two preventable ways. First, by making an open appeal for participation, rather than confining themselves to lists of known population or target members, the researchers can provide no assurance that respondents are actually part of the targeted population. A methodology report by the ACHESS researchers acknowledges that sample targets forwarded appeals to third parties, who joined the sample in a snowball fashion, though they do not seem to recognize the problem this creates [12]. There is no way to know whether those responding to the appeal were members or associates of the organizations targeted, or even same-sex parents. There is also no way to know whether or not some persons, highly interested in the study and its outcomes, may have responded multiple times. The possibility of such "sample crashing" is a prime reason why sampled surveys, which contact known unique members of a sample frame, are less prone to bias than are recruited samples. If the ACHESS survey had actually used member lists, from which the researchers contacted potential participants individually to invite their anonymous participation, this form of potential bias would have been avoided.

While the ACHESS respondent frame was far too wide, the appeal frame was much too narrow. Appeals for participation focused on "gay and lesbian press …gay and lesbian social and support groups, and …gay and lesbian community events."Two groups are mentioned in particular: recruitment "will include, but not be limited to, Gay Dads Australia and the Rainbow Families Council of Victoria" [26]. The Rainbow Families Council of Victoria is not a social and support group, although it may include those functions, but a political lobby. According to its published description: "Rainbow Families Council replaces the Fertility Access Rights Lobby, which has worked for the rights of same-sex parented families and prospective parents since 1999 and maintains a close relationship with the Victorian Gay & Lesbian Rights Lobby" [27]. Gay Dads Australia is a support group for male same-sex parents focusing on adoption and surrogate parenting. The organization actively promotes political causes relating to the legal status of same-sex parents, and views the ACHESS study as an ally in those efforts. The website for Gay

Dads Australia features links to the Rainbow Families Council, an LGBT political party, an advocacy group for same-sex marriage, an LGBT radio station—and the ACHESS study [28]. The post on the Gay Dads Australia blog encouraging participation in ACHESS urged that the data "will help direct health policy, political policy, anti-discrimination policy. Governments rely on good data. This is one way that we can all help get that data" [29].

Clearly, advertisements focused on lobby constituencies or those who stand to benefit politically, and explicit appeals to the policy relevance of a study, are measures that very likely will increase sample selection bias. Potential participants who are not as politically active, who support anti-discrimination policy less strongly, or for whom it is less salient of an issue, or who may judge that their own experiences with parenting are less likely to contribute to good data for that cause, will be less likely to participate in the study.

In an apparent attempt to counteract narrow recruitment among gay and lesbian interest groups, the ACHESS protocol also called for discussion pieces and interviews with mainstream media outlets in addition to the gay and lesbian press, to encourage participation by less-engaged same-sex parents. This well-intentioned attempt to widen the scope of recruitment, however, simply increases the problem of self-selection into the sample. Persons who take the trouble to respond to a passive media appeal are likely to be more engaged, not less, in the study outcomes, than are persons who respond to a strong membership appeal. Recruitment bias could have been more effectively reduced in the ACHESS study (and in future similar studies) simply by involving a much wider selection of groups with probable concentrations of same-sex parents. Groups more focused on the interpersonal rather than political aspects of same-sex attracted life, such as PFLAG Australia (Parents and Friends of Lesbians and Gays); or on the intersection of same-sex attraction and ethnic or religious identity, such as ArciLesbica Australia (support group for Italian GLBT women) or Queer Muslims Australia; or on the AIDS prevention community; just to name a few possibilities; would have added coverage and reduced bias in the sample.

Particularly striking is the absence of focus on organizations relating primarily to GLBT health and well-being. The Australian National GLBTI Health Alliance lists several dozen members geographically distributed throughout Australia which are concerned with health issues [30]. Some of the organizations have a political focus, but many do not. (Rainbow Families Council is a member, but Gay Dads Australia is not.) One of the explicit goals of the Health Alliance, moreover, is to promote better research on GLBT health. Recruited sample studies could reduce bias substantially by widening the scope of appeal to include larger organizations and networks such as this. In the case of ACHESS, however, collaboration with the National GLBTI Health Alliance may have been inhibited by a disagreement in research strategy discussed below that exemplifies the central issues of the present paper.

3.2.3 Interpretation bias

The ACHESS researchers have gone on record as supporting same-sex marriage, citing the ACHESS results in support of their position [31,32], so it is pertinent to consider whether the study's interpretation or analysis may be correspondingly biased. The conflation of the roles of researcher and advocate is hardly confined to ACHESS; it is widespread in SS parenting research. Unfortunately, this appears to be associated with another kind of conflation in interpreting the study results: although the sample is clearly acknowledged to be a non-representative nonprobability sample, it is then contradictorily treated for purposes of analysis as if it were representative for the basis of population inferences.

In the ACHESS publications this conflation is illustrated by the equivocal use of the word "representative" in describing the study sample. "Every effort was made to recruit a representative sample," states the findings report [6]. What the word "representative" means in this sentence is explained in the study protocol: "a diverse sample from the broad range of all families in the gay, lesbian, bisexual and transgender community to ensure maximum representation" [26]. This describes a sample in which the diverse range of family or sexual orientation types have the opportunity to be *represented in the sample*. But that is a very different thing from drawing a sample which is *statistically representative of a population*. As one learns in elementary probability, the latter quality requires that every member of the population have an equal chance of selection into the sample, a requirement which recruitment clearly invalidates. By conflating these two senses of the word “representative”, ACHESS fundamentally confuses the type of information that can be provided by its investigation.

Like many recruited sample studies of same-sex parents, the thinking in the ACHESS study seems to have been that if a snowball or convenience sample could be made diverse enough or large enough, then it would somehow become a statistically representative sample. The protocol informs us that the study would aim to achieve a large sample in order to combat, in part, unavoidable sample bias due to the hidden nature of the population. But larger size can increase the precision of sample representativeness, by reducing standard errors, only if it is already an unbiased representative sample to begin with. A non-random convenience sample does not become a representative population sample by making it larger. In fact, if, as appears to be the case with ACHESS, self-selected respondents are recruited from a narrow range of organizations, more vigorous or extensive recruitment may increase bias, by stimulating more highly engaged persons to respond. This may explain why, in Table 3, the observed SDQ bias in ACHESS is no smaller, in fact is slightly larger, than that of the two other recruited sample studies, even though the ACHESS sample is several times larger.

The analytical method of the ACHESS findings report compounds this confusion, demonstrating that this is not a tangential problem, but is central to the study. The mathematics of probabilistic inference that underlie scientific research require that the characteristics of a non-random sample, such as a recruited convenience sample, cannot be validly inferred to a population. Yet on the convenience sample that was recruited largely from email lists of gay interest organizations, the ACHESS researchers incomprehensibly employ an extensive battery of population analyses, including means comparisons with normative random population samples and tests of statistical significance of the coefficients of multiple regression analyses, that are only appropriate for random samples.

The authors report, for example, that on three of the nine scales on the Child Health Questionnaire SS children “demonstrated significant differences”, and that “[t] here were no significant differences identified for other CHQ scales”; and that none of the SDQ scale scores for SS children were significantly different than for a sample of OS children, listing the p-value for each comparison. At one point they even speak of a scale score that “approached significance” with a P-value of .052 [6]. These claims are incoherent. Statistical significance is, by definition, an assessment of the variation or uncertainty in measurement resulting from random population sampling. If the collected information on which the claim is based does not result from random sampling of some sort, statistical significance can have no interpretation. By using random-sample statistics to report apparent population inferences from a convenience sample, ACHESS transforms what may have been valuable non-parametric findings into unsubstantiated and possibly misleading parametric claims.

As aforementioned some researchers, including the ACHESS authors, counter concerns about bias with the objection that without recruited convenience samples there would be no way to gain information about the same-sex population at all. The findings of the present study suggest two effective responses to this concern. First, if it is a choice between having no information or having misleading information, the former is preferable. Science can be furthered much better by acknowledging uncertainty than by claiming an inaccurate certainty. Second, it is not true that recruited samples are the only way to collect information on same-sex persons or parents. As the random sample studies surveyed in this paper demonstrate, there are already random samples of this population; and the number of these sources are growing rapidly, as sexual orientation measures are increasingly being included in large-sample private and government population surveillance efforts.

3.2.3 Suggestions for Improvement

The purpose of this brief review has not been to criticize ACHESS in particular, but to review the sources and types of bias that are common, more or less, to the many SS parenting studies that rely on recruited samples. Above all, researchers in this area who wish to improve the state of information should strive to attain a genuinely random sample. If an investigator must use a recruited sample, the following practices will help to reduce and identify any bias that may be present, and so make the study findings more valuable.

1. Publish the raw data file. Many have advocated this standard for any public-

funded research, and it is particularly applicable to studies in the area of public health, where distorted findings can mislead both practitioners and public policy. With the ready availability of electronic data archives and the ability to easily de-identify sensitive records, there is little excuse for scientists to withhold their data from subsequent scholarly scrutiny. In the absence of a compelling reason to withhold, data that are not made available for critical review should be considered less credible. In the present study, every random sample of SS children reviewed made use of publicly available data, so any subsequent researcher can replicate and confirm their findings; whereas none of the recruited sample studies have (to my knowledge) made their data available.

2. Publish the uncontrolled univariate results, with uncertainty measures, and initial, uncontrolled statistical models where possible. The ACHESS report is exemplary in this respect (though it did not include the uncontrolled statistical models), but many of the other studies in the field of SS parenting have not been so forthcoming.
3. Do not treat a non-probability sample as if it were a probability sample. This practice is misleading and should be discontinued. This would include refraining to use inferential statistics or report p-values that depend on the assumption of random sampling, and refraining from making claims about population characteristics based on the sample. A recruited study should include a strong, clearly-worded disclaimer to ensure it is not confused with a statistically representative study, such as: “The sample in this study is not statistically representative and cannot be used to validly infer the characteristics of any population in the real world.”

4. CONCLUSION

This study has found strong evidence of substantial bias understating the psychological difficulties of children with same-sex parents on the Strength and Difficulties Questionnaire (SDQ) in studies using recruited convenience samples. Well-informed health or social policy in this area, this suggests, should not be based on studies with recruited samples, but on rigorous random sample research. Likewise, scientific or scholarly outlets should refrain from publishing population claims based on recruited samples. Recruited sample studies may have other value, however. A detailed analysis of one such study, the Australian Study of Children in Same-Sex Families (ACHESS) concludes in suggestions, in the form of three common-sense rules, for ameliorating bias and preventing misunderstanding.

Since all of the unambiguously random sample data on emotional problems was examined in the course of this analysis, the findings also make a strong substantive point: to date, no representative population data have found lower emotional problems among children with same-sex parents. Every random sample has observed higher emotional problems among such children; where the sample was large enough, those differences were statistically significant. The evidence examined in this study is limited in several ways. It is quite possible that other recruited sample studies, not examined here because they did not use the SDQ, have shown much less or much more bias than the recruited sample studies analyzed here. It is also possible that other measures, indicators or variables are less susceptible to bias in this area of study. A comprehensive examination or meta-analysis of all extant studies of same-sex parenting would provide helpful further evidence to confirm or rebut the suggestive conclusions of this study with greater certainty.

The more modest power appraisal and increased disclosure proposed by the three rules presented above may inhibit the political use of recruited sample studies of this population. However, research that violates the careful standards of scientific inference eventually becomes self-refuting as its bias becomes more generally known, and may actually hinder the development of more enlightened health and social policy regarding children in same-sex families. This point is made best, perhaps, by the Australian National GLBTI Health Alliance, a coalition of groups devoted to the health of the same-sex population that has already been mentioned above. This national organization complains: “An understanding as to whether LGBTI Australians are disproportionately affected by specific health issues can only at present be deduced from individual, often small, research studies which do not cover the population as a whole” [33]—such as the ACHESS study and other small recruited sample studies. As a corrective, the National GLBTI Health Alliance calls for the collection of comprehensive, large-scale random-sample data by means such as “[t]he inclusion of questions on sexual identity and gender identity in the Census, the National Health Survey and other official statistics data collection" as well as government-funded grant research and funding for a large-sample national study of same-sex Australians [33].

Like the present study, the Australian National GLBTI Health Alliance advocates representative (random-sample) population data which "would provide irrefutable evidence about whether or not sexuality is itself a social determinant of health" [33]. In so doing, they recognize that small studies with biased samples, which may tend for political purposes to understate health problems among same-sex persons, are not the best means to serve the genuine health needs of this population. Researchers as well as all parents, both OS and SS, should be able to agree that the goal of public health investigation in this area should be accurate, unbiased information that will best serve the health and welfare of all children involved.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Allen D. High school graduation rates among children of same-sex households. Rev Econ Househ. 2013;11(4):635-58.
2. Marks L. Same-sex parenting and children's outcomes: A closer examination of the American psychological association's brief on lesbian and gay parenting. Soc Sci Res. 2012;41(4):735-51.
3. Manning WD, Fettro MN, Lamidi E. Child well-being in same-sex parent families: Review of research prepared for American Sociological Association Amicus Brief. Popul Res Policy Rev. 2014;33(4):485.
4. Anderssen N, Amlie C, Ytterøy EA. Outcomes for children with lesbian or gay parents. A review of studies from 1978 to 2000. Scand J Psychol. 2002;43(4):335-51.
5. Rosenfeld MJ. Nontraditional families and childhood progress through school. Demography. 2013;50(3):963-9.
6. Crouch S, Waters E, McNair R, Power J, Davis E. Parent-reported measures of child health and wellbeing in same-sex parent families: A cross-sectional survey. BMC Public Health. 2014;14(1):635.
7. Sullins DP. Child emotional problems in non-traditional families. Internet. Rochester NY: Social Science Research Network; 2014. Cited 2014 Nov 7. Report No: ID 2500537. Available:http://papers.ssrn.com/abstract=2500537
8. Goodman R. Psychometric Properties of the Strengths and Difficulties Questionnaire. J Am Acad Child Adolesc Psychiatry. 2001;40(11):1337-45.
9. Goodman A, Goodman R. Strengths and difficulties questionnaire as a dimensional measure of child mental health. J Am Acad Child Adolesc Psychiatry. 2009;48(4):400-3.
10. He JP, Burstein M, Schmitz A, Merikangas K. The Strengths and Difficulties Questionnaire (SDQ): The factor structure and scale validation in U.S. Adolescents. J Abnorm Child Psychol. 2013;41(4):583-95.
11. Mellor D. Normative data for the strengths and difficulties questionnaire in Australia. Aust Psychol. 2005;40(3):215-22.
12. Crouch S, Waters E, McNair R, Power J, Davis E, van Mourik L. Triumphs and challenges in recruiting same-sex parent families. Aust N Z J Public Health. 2014;38(1):87-8.
13. Cochran SD, Mays VM. Risk of breast cancer mortality among women cohabiting with same sex partners: Findings from the National Health Interview Survey 1997-2003. J Womens Health. 2012;21(5):528-33.
14. Liu H, Reczek C, Brown D. Same-sex cohabitors and health: The role of race-ethnicity gender and socioeconomic status. J Health Soc Behav. 2013;54(1):25-45.
15. Reczek C, Liu H, Brown D. Cigarette smoking in same-sex and different-sex unions: The role of socioeconomic and psychological factors. Popul Res Policy Rev. 2014;33(4):527-51.
16. Reczek C, Liu H, Spiker R. A population-based study of alcohol use in same-sex and different-sex unions. J Marriage Fam. 2014;76(3):557-72.
17. Golombok S, Perry B, Burston A, Murray C, Mooney-Somers J, Stevens M, et al. Children with lesbian parents: A community study. Dev Psychol. 2003;39(1):20.
18. Bos HH. Planned gay father families in kinship arrangements. Aust NZJ Fam Ther. 2010;31(4):356-71.

19. Golombok S, Mellish L, Jennings S, Casey P, Tasker F, Lamb ME. Adoptive gay father families: parent–child relationships and children's psychological adjustment. Child Dev. 2014;85(2):456-68.
20. Wainright JL, Russell ST, Patterson CJ. Psychosocial adjustment, school outcomes and romantic relationships of adolescents with same-sex parents. Child Dev. 2004;75(6):1886-98.
21. Wainright JL, Patterson CJ. Delinquency victimization and substance use among adolescents with female same-sex parents. J Fam Psychol. 2006;20(3):526.
22. Peck R, Devore J. Statistics: The exploration &analysis of data. Cengage Learning. 2011;816.
23. Smith M. Common mistakes in using statistics. Internet. Cited 2014 Nov 11. Available:https://www.ma.utexas.edu/users/mks/statmistakes/TOC.html
24. Bailey JM, Bobrow D, Wolfe M, Mikach S. Sexual orientation of adult sons of gay fathers. Dev Psychol. 1995;31(1):124.
25. Bailey JM, Dunne MP, Martin NG. Genetic and environmental influences on sexual orientation and its correlates in an Australian twin sample. J Pers Soc Psychol. 2000;78(3):524.
26. Crouch SR, Waters E, McNair R, Power J, Davis E. ACHESS–The Australian study of child health in same-sex families: Background research, design and methodology. BMC Public Health. 2012;12(1):646.
27. Rainbow Families Council. Internet. Cited 2014 Dec 10. Available:http://www.rainbowfamilies.org.au/
28. Gay dads Australia Internet. Cited 2014 Dec 10. Available:http://www.gaydadsaustralia.com.au/
29. Gay Dads Australia Internet. Be part of the world's largest study into the health and wellbeing of our children blog announcement; 2012. Cited 2014 Nov 12. Available:http://gaydadsaustralia.blogspot.com/
30. National LGBTI Health Alliance Members | National LGBTI Health Alliance. Internet. Cited 2014 Dec 11. Available:http://lgbtihealth.org.au/members
31. Crouch S. It's time to adopt change | Election Watch. Internet. Cited 2014 Dec 10. Available:http://electionwatch.edu.au/victoria-2014/it%E2%80%99s-time-adopt-change
32. Crouch S, Block K, Davis E, Waters E, Gibbs L. How voting for equity will make life better now and in the future [Internet]. The Conversation; 2013. Cited 2014 Dec 11. Available:http://theconversation.com/how-voting-for-equity-will-make-life-better-now-and-in-the-future-17167
33. Researching LGBTI Health | National LGBTI Health Alliance. Internet. Cited 2014 Dec 11. Available:http://lgbtihealth.org.au/research

---

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=751&id=22&aid=7750