

1  
2  
3  
4  
5  
6  
7  
8  
9

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

---

10  
11

### 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 same-sex children, compared with no Favorable comparisons (0%, range 0-0) in the random samples. Two additional random samples with related measures were also added, 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.

12  
13  
14  
15  
16

*Keywords: same-sex parents, child emotional problems, Strengths and Difficulties Questionnaire (SDQ), survey bias*

17 **INTRODUCTION**

18

19 Dozens of studies in past decades have suggested that children with same-sex parents  
20 (hereafter for brevity “SS children”) suffer no significant disadvantages in well-being  
21 compared to children with opposite-sex parents (hereafter “OS children”). Recent competing  
22 reviews of this literature, however, have agreed that the field has been dominated by small  
23 convenience samples recruited from GLBT (acronym for “Gay, Lesbian, Bi-sexual,  
24 Transgender”) as a blanket term for members of minority sexual orientations; also sometimes  
25 “LGBT”) interest, advocacy or support groups. Allen, a critic of the “no differences” claim,  
26 found that only four of 49 studies published prior to 2010 drew a random or probability  
27 sample of SS parents. The remainder based their findings on conveniently available or  
28 selected groups of participants, often recruited from biased, politically aware sources such  
29 as “LGBT events, bookstore and newspaper advertisements, word of mouth, networking and  
30 youth groups” (1, see also 2,3). Anderssen et al. (4), reviewing an earlier set of studies,  
31 reported that for only one out of 23 studies of child outcomes with same-sex parents prior to  
32 2000 was the aim of the study unknown to participants. Reviews by Rosenfeld (5) and  
33 Manning (3), while defending “no differences”, do not dispute these facts.

34

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

49

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

60

61 Is the finding of “no differences” due to bias related to sample recruitment? Two recent  
62 studies with sharply differing results provide an opportunity to examine this question in a  
63 focused way. Both studies use the identical metric—the parent-reported Strengths and  
64 Difficulties Questionnaire (SDQ)—to document contradictory findings on the effect of SS  
65 parenting on child emotional health. The first, a report by Crouch and colleagues on a large,  
66 well-crafted convenience sample of SS parents, finds “no evidence to support a difference in  
67 parent-reported child health” between OS parents and SS parents. (6) The second, an  
68 analysis by Sullins of a large multi-stage random population sample, finds that “children in  
69 same-sex families are at least twice as likely to experience serious emotional problems

70 compared to their counterparts in opposite-sex families”(7). These studies, which arrive at  
71 different results on the same measure in different sample types, form a kind of natural  
72 experiment to address the question of sample bias in SS parenting research.

73

74 The present study proceeds by comparing the results of these two studies with their  
75 respective reference populations to observe the possibility of bias due to sample design.  
76 The analysis is then extended to other similar studies in the SS parenting literature. The  
77 results suggest the presence of persistent bias, toward favorable outcomes for SS children,  
78 in non-random recruited samples. The probable causes of the bias, and suggestions for  
79 reducing and identifying it, are discussed.

80

## 81 **2. DATA AND METHODS**

82

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

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

104 The United States’ National Health Interview Survey (NHIS), a project of that government’s  
105 Centers for Disease Control (CDC),annually collects informationon 75,000 to 100,000  
106 individuals from household interviews based on a complex multistate probability sample that  
107 is statistically representative of the civilian noninstitutionalized population of the United  
108 States. Extensive health and demographic information is collected for all household  
109 members. In addition, for each family that includes children under age 18, detailed  
110 supplemental health information, provided by a parent or other knowledgeable adult  
111 informant, is collected for one child chosen at random (the “sample child”).

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

117 **3. RESULTS**

118

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

129

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

137

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

146

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

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

SDQ Domains	NHIS			Mellor/ACHES		
	OS Children	SS Children	P	OS Children	SS Children	P
Conduct	1.21 (1.61)	<b>2.06</b> <sup>***</sup> (2.19)	.001	1.5 (1.6)	1.44 (1.59)	.57
Hyperactivity	2.72 (2.51)	<b>3.62</b> <sup>**</sup> (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)	<b>1.61</b> <sup>***</sup> (1.89)	<.001
Peer	1.29 (1.51)	<b>1.65</b> <sup>*</sup> (1.63)	.03	1.6 (1.9)	1.42 (1.68)	.12
Pro-social (low)	1.25 (1.69)	<b>1.98</b> <sup>**</sup> (2.16)	.01	1.7 (1.7)	<b>1.98</b> <sup>*</sup> (1.99)	.02
Total Difficulties	6.65 (5.68)	<b>9.00</b> <sup>***</sup> (6.68)	.001	8.18 (6.1)	7.48 (5.49)	.06
N	27,327	71		941	299	

\*  $p \leq .05$  \*\*  $p \leq .01$  \*\*\*  $p \leq .001$

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) ACHES = Australian Child Health in Same Sex Families  
 All significance tests were independently assessed, and may not agree with reported results from the study involved.

155  
 156 The former possibility seems unlikely given the large cultural and social similarities between  
 157 Australia and the USA, and in light of the fact that same-sex couples enjoy more liberal legal  
 158 and political status in the USA than they do in Australia; but it cannot be dismissed out of  
 159 hand. Bos suggests, for example, that higher stigmatization for SS children in the US than in  
 160 Holland may account, in part, for the lower rate of problem behaviors among the Dutch SS  
 161 children. However, Sullins documents that, on NHIS combined data from 2000-2013, SS  
 162 children experience less overall stigmatization than do OS children. If the higher emotional  
 163 problems observed for SS children in the US compared to those in Australia were due to  
 164 stigmatization, therefore, we would expect OS children also to experience higher emotional

165 problems than their Australian counterparts; but, as Table 1 shows, the opposite is the case.  
166 On all the SDQ domains, OS children in the US have lower means for emotional difficulties  
167 than do OS children in Australia.

168

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

176

### 177 **Broadening Scope**

178

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

185

186 In addition to Sullins and Crouch, since 2000 three other studies of SS parenting have  
187 reported SDQ scale results for SS children. One of the three made use of a random  
188 population sample comparable to NHIS(7); the other two used a recruited sample  
189 comparable to ACHES(18,19). Altogether, these five studies comprise all gay parenting  
190 studies that have used the SDQ (to my knowledge). Table 2 summarizes the results. The  
191 two leftmost studies presented in the table report the two random samples; the remaining  
192 three study results report the recruited samples. The rightmost of the four columns  
193 summarizing each study’s findings (labeled “F/U”) reports whether each SDQ domain  
194 comparison is Favorable or Unfavorable for SS children. Statistical significance is not strictly  
195 applicable to these comparisons and is reported only as a rough indicator of magnitude.

196

197 The dispositive issue for sample bias is whether the pattern already observed, with the  
198 random sample reflecting unfavorable findings for SS children but the recruited sample  
199 reflecting favorable results, persists across the additional studies. Inspection of Table 2  
200 reveals that the answer to this question is an emphatic “yes”. As with NHIS, SS children are  
201 consistently found to have higher emotional problems for every SDQ domain on the other  
202 random sample, but, as with ACHES, on none of the other recruited samples. This second  
203 random sample, from Golombok’s 2003 study, reports on data from England, providing  
204 further evidence that the pattern of higher emotional problems for SS children is not simply a  
205 national characteristic of the United States. Participants in the recruited samples report  
206 generally more favorable results for SS children than for OS children. The distinction in  
207 results by sample type is clear, even across two studies by the same lead author.

208

209 While the pattern of higher emotional difficulties among SS children occurs without exception  
210 in the random samples, in the recruited samples there are some exceptions to the  
211 contrasting pattern, which nonetheless predominates. Despite the claim cited above, there  
212 is no reason to suppose that sample bias is necessarily monotonic, or affects all studies or  
213 all behavioral domains to the same degree. The exceptions to the pattern of “favorable  
214 random/unfavorable non-random” results, moreover, are not randomly distributed. For four  
215 of the six SDQ domains—conduct problems, hyperactivity, emotional symptoms and total  
216 difficulties—there are no exceptions to the pattern, but the results are mixed for the other two  
217 domains—peer problems and pro-social behavior. There may be many reasons for this

218 differenceamong the recruited samples that are consistent with the operation of bias,  
219 including the presence, to a greater or lesser degree, of more than one form of bias.  
220 Perhaps parents of children with more externalizing problems, such as misconduct or

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

SDQ Domains	US NHIS (2000-2004)				Golombok 2003 <sup>1</sup>				Bos 2010				Golombok 2014				Mellor/ACHES 2014			
	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)	<b>2.06</b> *** (2.19)	.001	U	12.8	13.9	.55	U	1.19 (.25)	1.18 (.33)	.89	F	2.90 (1.80)	<b>2.12</b> * (2.08)	.03	F	1.5 (1.6)	1.44 (1.59)	.57	F
Hyperactivity-inattention	2.72 (2.51)	<b>3.62</b> ** (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)	<b>1.61</b> *** (1.89)	<.001	F
Peer problems	1.29 (1.51)	<b>1.65</b> * (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)	<b>8.02</b> ** (2.16)	.01	U	3.0	8.3	.15	U	2.67 (.28)	2.67 (.33)	1.0	N	--	--	--	--	8.3 (1.7)	<b>8.02</b> * (1.99)	.02	U
<b>Total Difficulties</b>	6.65 (5.68)	<b>9.00</b> *** (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

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  
All significance tests were independently assessed, and may not agree with reported results from the study involved.

<sup>1</sup> Values reported for this study are percent abnormal, not mean; p-values reported are for chi square tests.



223 hyperactivity, were more likely to voluntarily deselect from the recruited sample; or perhaps  
 224 parents' knowledge of peer problems or pro-social behavior on the part of their children, or their  
 225 understanding of what constitutes social desirability in these areas, was less complete or clear than  
 226 it was for the other domains. Whatever the reasons (and it is beyond the scope  
 227 of this study to determine them), there is no contradiction in the supposition that bias, if it is at work,  
 228 may itself be biased in its effect on study results.

**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

229  
 230 To facilitate interpretation, Table 3 summarizes the proportion of favorable comparisons from each  
 231 study in Table 2, presented in the first five rows of the table. The top two rows report on the two  
 232 random samples; the next three rows present findings from the three recruited samples. For further  
 233 context Table 3 also adds findings from two other random sample studies of SS parenting prior to  
 234 2010 that measured child emotional or behavioral difficulties, but did not use the SDQ. Although  
 235 some other studies in this developing literature included elements of random sampling, Wainright  
 236 and Patterson produced the only three studies that made use of a statistically representative “blind”  
 237 population sample of SS children, identifying them in the data of the National Longitudinal Survey

238 of Adolescent Health. Two of these studies contained measures of emotional or behavioral  
239 distress for SS children compared with OS children. In 2004 they reported comparative findings for  
240 depression, self-esteem and anxiety, and in 2006 reported on eleven measures of delinquency  
241 among adolescents such as binge drinking, illegal drug use or risky sexual behavior. All of the  
242 above measures in both studies were less favorable for SS children in the sample, extending the  
243 pattern already observed on the NHIS and Golombok's 2003 study, At this point, Table 3 includes  
244 all the unambiguously random data on SS children gathered prior to 2010.

245

246 In comparing the findings in Table 3, an average difference *between* sample types that is greater  
247 than the difference *within* sample types would suggest the presence of systematic bias. Here it is  
248 easy to see from inspection that, not only is the difference between random and recruited samples  
249 greater than it is within those two categories, the magnitude of the comparison is extremely large.  
250 Among the recruited samples, the percent of findings favorable to SS children, compared to OS  
251 reference children, varies by at only 4 percentage points from an average just over 79; among the  
252 random samples there is no variation at all; but between these two groups of findings is a  
253 difference of almost 80 percentage points. Put another way, the difference between the random  
254 and recruited samples is almost 20 times as large as the variation among the recruited samples.  
255 This robust finding strongly suggests that a large and persistent bias toward favorable findings for  
256 SS children, compared to OS reference children, is present in the recruited samples relative to the  
257 random ones.

258

### 259 **Sources of Bias**

260

261 The presence of such large bias implies that recruited samples are not very accurate or credible  
262 sources for assessing the population characteristics of SS children. Academics, policymakers,  
263 jurists and health practitioners interested in evidence-based decision-making should demand  
264 random sample evidence to inform their deliberations regarding this population. However, random  
265 sample evidence may not always be feasible to obtain or accurate population inferences may not  
266 be needed as a study outcome; in which cases it would be helpful to briefly examine more closely  
267 the types and causes of bias that may be present in recruited sample studies in this area, and  
268 consider how they might be reduced, especially for those readers who may not be applied  
269 quantitative researchers. Since ACHES is one of the largest, carefully conceived and most recent  
270 recruited sample studies of this population, it will be helpful to focus our examination on the  
271 potential sources of bias in this study. If we can understand how an exemplary study like ACHES  
272 may have unaddressed bias, we will better understand how any study of this type may be biased.

273

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

282

### 283 **Social Desirability Bias**

284

285 Because the ACHES SDQ scales depend on parent reports, their findings are subject to social  
286 desirability bias, that is, the tendency for parents to report more socially desirable information  
287 about their own children (or themselves) than is objectively the case. To correct for this tendency  
288 the SDQ is often administered to children's teachers and/or the children themselves, as well as

289 parents, and the two sets of results are compared. Both teacher and self-report ratings tend to be  
290 lower than those of parents. Since the comparative OS data for the ACHESS scales are also  
291 based on parent reports, social desirability would only bias the comparisons if the ACHESS parents  
292 were more or less prone to give desirable responses than is the OS comparison group. One factor  
293 that may have increased such differential bias is that ACHESS surveyed parents using paper or  
294 online survey instruments, while the comparison group of OS parents received telephone  
295 interviews; social desirability bias is known to be more common in interviews, where the  
296 respondent is speaking with another human being, than on written or computer surveys, where the  
297 respondent is more anonymous. This difference, however, would tend to lower favorable  
298 responses by the ACHESS parents, which is the opposite of the bias observed in their responses.  
299 On the other hand, the fact that all the SS respondents, but none of the OS comparison group,  
300 were aware of the purposes of the ACHESS study, and stood to benefit from any increased social  
301 acceptance or support that a positive picture of SS children might help to bring about, may well  
302 have increased favorable responses, particularly when combined with selection bias as discussed  
303 below. We do not know how substantial such bias may be, but it would not be difficult to determine  
304 by including teacher and/or self-reported ratings on the ACHESS SDQ scales. The ACHESS  
305 protocol stated that study would include child self-reports, but there is no mention of these results  
306 in the subsequent report. In the report Crouch et al. acknowledge, albeit minimally, the possibility  
307 of parent self-representation bias, and indicate that future research will include “child-reported  
308 measures of health”, as well as parent interviews. The absence of such measures to date leaves  
309 uncertain the extent to which social desirability bias may be present.

310

### 311 **Sample Selection Bias**

312

313 While it is unclear whether social desirability bias may have distorted ACHESS’s results, there is no  
314 question that that they are distorted by selection bias, since respondents were invited to volunteer  
315 for the study through general announcements. As most introductory statistics texts point out, a  
316 sample comprised only of self-selected individuals is unavoidably biased, since those who choose  
317 not to participate are almost certainly different in systematic ways from those who choose to  
318 participate (20). Smith states this basic principle clearly: “*Voluntary response samples are always*  
319 *biased: they only include people who choose to volunteer, whereas a random sample would need*  
320 *to include people whether or not they choose to volunteer*” (emphasis in original) (21).

321

322 Not only was the ACHESS sample recruited, but it was recruited 1) through open public appeals for  
323 participation 2) disseminated among politically aware groups whose participants were clearly  
324 “interested”, in a Marxian sense, in the outcome of the ACHESS study. These two factors  
325 undoubtedly increased bias, and could have easily been corrected, or could be corrected in similar  
326 future samples.

327

328 The ACHESS protocol stated: “Primarily recruitment will be through emails posted on gay and  
329 lesbian community email lists aimed at same-sex parenting.” The use of the word “lists” here is  
330 misleading. What actually appears to have been done is that announcements were published on  
331 the two websites named and affiliated blogs. This is consistent with the other recruitment activities  
332 called for in the protocol: advertisements, media releases, flyers and researcher appearances at  
333 public events. This procedure increased probable bias in two preventable ways. First, by making  
334 an open appeal for participation, rather than confining themselves to lists of known population or  
335 target members, the researchers can provide no assurance that respondents are actually part of  
336 the targeted population. There is no way to know whether those responding to the appeal were  
337 members or associates of the organizations targeted, or even same-sex parents. There is also no  
338 way to know whether or not some persons, highly interested in the study and its outcomes, may  
339 have responded multiple times. The possibility of such “sample crashing” is a prime reason why

340 sampled surveys, which contact known unique members of a sample frame, is less prone to bias  
341 than a recruited sample. If the ACHESSE survey had actually used member lists, from which the  
342 researchers contacted potential participants individually, this form of potential bias would have  
343 been avoided.

344

345 While the respondent frame was far too wide, however, the appeal frame was much too narrow.  
346 Appeals for participation focused on “gay and lesbian press ...gay and lesbian social and support  
347 groups, and ...gay and lesbian community events.” This procedure inevitably focused recruitment  
348 efforts on politically aware groups whose participants were clearly “interested”, in a Marxian sense,  
349 in the outcome of the ACHESSE study. Two groups are mentioned in particular: recruitment “will  
350 include, but not be limited to, Gay Dads Australia and the Rainbow Families Council of Victoria.”  
351 The Rainbow Families Council of Victoria is not a social and support group, although it may include  
352 those functions, but a political lobby, according to its published description: “Rainbow Families  
353 Council replaces the Fertility Access Rights Lobby, which has worked for the rights of same-sex  
354 parented families and prospective parents since 1999 and maintains a close relationship with the  
355 Victorian Gay & Lesbian Rights Lobby.” Gay Dads Australia is a support group for male same-sex  
356 parents focusing on adoption and surrogate parenting. The organization actively promotes political  
357 causes relating to the legal status of same-sex parents, and views the ACHESSE study as an ally in  
358 those efforts. The website for Gay Dads Australia features links to the Rainbow Families Council,  
359 an LGBT political party, an advocacy group for same-sex marriage, an LGBT radio station—and  
360 the ACHESSE study. The post on the Gay Dads Australia blog encouraging participation in ACHESSE  
361 urged that the data “will help direct health policy, political policy, anti-discrimination policy.  
362 Governments rely on good data. This is one way that we can all help get that data.”

363

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

370 In an apparent attempt to counteract narrow recruitment among gay and lesbian interest groups,  
371 the ACHESSE protocol also called for discussion pieces and interviews with mainstream media  
372 outlets in addition to the gay and lesbian press, to encourage participation by less-engaged same-  
373 sex parents. This creditable attempt to widen the scope of recruitment, however, simply increases  
374 the problem of self-selection into the sample. Persons who take the trouble to respond to a  
375 passive media appeal are likely to be more engaged, not less, in the study outcomes, than are  
376 persons who respond to a strong membership appeal.

377

378 Recruitment bias could have been more effectively reduced in the ACHESSE study (and in future  
379 similar studies) simply by involving a much wider selection of groups with probable concentrations  
380 of same-sex parents. Groups more focused on the interpersonal rather than political aspects of  
381 same-sex attracted life, such as PFLAG Australia (Parents and Friends of Lesbians and Gays); or  
382 on the intersection of same-sex attraction and ethnic or religious identity, such as ArciLesbica  
383 Australia (support group for Italian GLBT women) or Queer Muslims Australia; or on the AIDS  
384 prevention community; just to name a few possibilities, would have added coverage and reduced  
385 bias in the sample.

386 Particularly striking is the absence of focus on organizations relating primarily to GLBT health and  
387 well-being. The Australian National GLBTI Health Alliance lists several dozen members  
388 geographically distributed throughout Australia which are concerned with health issues. Some of  
389 the organizations have a political focus, but many do not. (Rainbow Families Council is a member,

390 but Gay Dads Australia is not.) One of the goals of the Health Alliance, moreover, is to promote  
391 better research on GLBT health. Recruited sample studies could reduce bias substantially by  
392 widening the scope of appeal to include larger organizations and networks such as this. In the  
393 case of ACHES, however, collaboration with the National GLBTI Health Alliance may have been  
394 inhibited by a disagreement in research strategy discussed below that exemplifies the central  
395 issues of the present paper.

### 396 **Interpretation Bias**

397  
398 Interpretation bias can occur when researchers are personally sympathetic to a particular study  
399 outcome, which then influences their interpretation of its findings. The ACHES researchers have  
400 gone on record as supporting same-sex marriage, citing the ACHES results in support of their  
401 position, so it is appropriate to consider whether the study's interpretation or analysis may be  
402 correspondingly biased. The conflation of the roles of researcher and advocate is hardly confined  
403 to ACHES; it is widespread in SS parenting research. Unfortunately, this appears to be  
404 associated with another kind of conflation in interpreting the study results: although the sample is  
405 clearly acknowledged to be a non-representative nonprobability sample, it is then treated for  
406 purposes of analysis as if it were representative for the basis of population inferences.

407  
408 In the ACHES publications this conflation is illustrated by the equivocal use of the word  
409 "representative" in describing the study sample. "Every effort was made to recruit a representative  
410 sample," states the findings report. What the word "representative" means in this sentence is  
411 explained in the study protocol: "a diverse sample from the broad range of all families in the gay,  
412 lesbian, bisexual and transgender community to ensure maximum representation". This describes  
413 a sample in which the diverse range of family or sexual orientation types have the opportunity to be  
414 *represented in the sample*. But that is a very different thing from drawing a sample which is  
415 *statistically representative of a population*. As we learn in elementary probability, the latter quality  
416 requires that every member of the population have an equal chance of selection into the sample, a  
417 requirement which recruitment clearly invalidates. By conflating these two senses of the word  
418 "representative", ACHES fundamentally confuses the type of information that can be provided by  
419 its investigation.

420  
421 Like many recruited sample studies of same-sex parents, the thinking in the ACHES study seems  
422 to have been that if a snowball or convenience sample could be made diverse enough or large  
423 enough, then it would somehow become a statistically representative sample. The protocol informs  
424 us that the study would aim to achieve a large sample in order to combat, in part, unavoidable  
425 sample bias due to the hidden nature of the population. But (as almost any introductory survey  
426 statistics text states) *increasing sample size does not reduce bias*. A larger sample size can  
427 increase statistical power, by reducing random sampling error, only if the sample is representative  
428 and unbiased—in this case, random—to begin with. A non-random convenience sample does not  
429 become a representative population sample by making it larger. In fact, if, as appears to be the  
430 case with ACHES, self-selected respondents are recruited from a narrow range of organizations,  
431 more vigorous or extensive recruitment may increase bias, by stimulating more highly engaged  
432 persons to respond. This may explain why, in Table 3, the observed SDQ bias in ACHES is no  
433 smaller, in fact is slightly larger, than that of the two other recruited sample studies, even though  
434 the ACHES sample is several times larger.

435  
436 The analytical method of the ACHES findings report compounds the confusion, demonstrating  
437 that this is not a tangential problem, but is central to the study. The mathematics of probabilistic  
438 inference that underlie scientific research require that the characteristics of a non-random sample,  
439 such as a recruited convenience sample, cannot be validly inferred to a population. Yet on the



440 convenience sample that they recruited largely from email lists of gay interest organizations, the  
441 ACHESS researchers incomprehensibly employ an extensive battery of population analyses,  
442 including means comparisons with normative random population samples and tests of statistical  
443 significance of the coefficients of multiple regression analyses that are only appropriate for random  
444 samples. They report, for example, that on three of the nine scales on the Child Health  
445 Questionnaire SS children “demonstrated significant differences”, and that “[t]here were no  
446 significant differences identified for other CHQ scales”; and that none of the SDQ scale scores for  
447 SS children were significantly different than for a sample of OS children, listing the p-value for each  
448 comparison. At one point they even speak of a scale score that “approached significance” with a  
449 P-value of .052. What can this language possibly mean for a convenience sample? Statistical  
450 significance is an assessment of the variation or uncertainty in measurement resulting from random  
451 population sampling. If the collected information on which the claim is based does not result from  
452 random sampling, statistical significance can have no interpretation. What can it mean to say that  
453 a result “approached significance”, that is, the magnitude and/or precision of the finding almost  
454 exceeded the variability in the sample due to random sampling at the conventional level for  
455 concluding that the sample result exists in the population, when there is no random sampling  
456 involved? In using random-sample statistics to report apparent inferences from a convenience  
457 sample, ACHESS transforms what may have been valuable non-parametric findings into  
458 unsubstantiated and possibly misleading parametric claims.  
459

460 As aforementioned some researchers, including the ACHESS authors, counter concerns about  
461 bias with the objection that without recruited convenience samples there would be no way to gain  
462 information about the same-sex population at all. There are two effective responses to this  
463 concern. First, if it is a choice between having no information and having misleading information,  
464 the former is the preferable option. Science can be furthered much better by acknowledging  
465 uncertainty than by claiming an inaccurate certainty. Second, it is not true that recruited samples  
466 are the only way to collect information on same-sex persons or parents. As the random sample  
467 studies surveyed in this paper demonstrate, there are already random samples of this population;  
468 and the number of these sources are growing rapidly, as sexual orientation measures are  
469 increasingly being included in large-sample private and government population surveillance efforts.  
470

471 As noted above, the purpose of this brief review has not been to criticize ACHESS in particular, but  
472 to review the sources and types of bias that are common, more or less, to the many SS parenting  
473 studies that rely on recruited samples. Above all, researchers in this area who wish to improve the  
474 state of information should strive to attain a genuinely random sample. If an investigator must use a  
475 recruited sample, the following practices will help to reduce and identify any bias that may be  
476 present, and so make the study findings more valuable.  
477

- 478 1. Publish the raw data file. Many have advocated this standard for any public-funded  
479 research, and it is particularly applicable to studies in the area of public health, where  
480 distorted findings can mislead both practitioners and public policy. With the ready  
481 availability of electronic data archives and the ability to easily de-identify sensitive records,  
482 there is little excuse for scientists to withhold their data from subsequent scholarly scrutiny.  
483 In the absence of a compelling reason to withhold, data that are not made available for  
484 critical review should be considered less credible. In the present study, every random  
485 sample of SS children reviewed made use of publicly available data, so any subsequent  
486 researcher can replicate and confirm their findings; whereas none of the recruited sample  
487 studies have made their data available.  
488
- 489 2. Publish the uncontrolled univariate results, with uncertainty measures, and initial,  
490 uncontrolled statistical models where possible. The ACHESS report is exemplary in this

491 respect (though it did not include the uncontrolled statistical models), but many of the other  
492 studies in the field of SS parenting have not been so forthcoming.

493  
494 3. Do not treat a non-probability sample as if it were a probability sample. Many of the  
495 recruited sample studies in this field, including ACHES, make a formulaic disclaimer  
496 about the representativeness of their sample, and then proceed to make population claims  
497 anyway, as if the sample were statistically representative. This practice is misleading and  
498 should be discontinued. This would include refraining to use inferential statistics or report  
499 p-values that depend on the assumption of random sampling, and refraining from making  
500 claims about population characteristics based on the sample. A recruited study should  
501 include a strong, clearly-worded disclaimer to ensure it is not confused with a statistically  
502 representative study, such as: "The sample in this study is not statistically representative  
503 and cannot be used to infer the characteristics of any population in the real world."

504

#### 505 **4. CONCLUSION**

506

507 This study has found strong evidence of substantial bias understating the psychological difficulties  
508 of children with same-sex parents on the Strength and Difficulties Questionnaire (SDQ) in studies  
509 using recruited convenience samples. This suggests that studies with recruited samples should  
510 not form the basis for health or social policy in this area, in favor of rigorous random sample  
511 research. Likewise, scientific or scholarly outlets should refrain from publishing population claims  
512 based on recruited samples, although recruited sample studies may have other value. A detailed  
513 analysis of one such study, the Australian Study of Children in Same-Sex Families (ACHES)  
514 offers suggestions, concluding in three common-sense rules, for ameliorating bias and preventing  
515 misunderstanding.

516

517 The evidence here is limited in several ways. It is quite possible that other recruited sample  
518 studies, not examined here because they did not use the SDQ, have shown much less or much  
519 more bias than the recruited sample studies analyzed here. It is also possible that other measures,  
520 indicators or variables are less susceptible to bias in this area of study. A comprehensive  
521 examination or meta-analysis of all extant studies of same-sex parenting would provide helpful  
522 further evidence to confirm or rebut the suggestive conclusions of this study with greater certainty.

523

524 Since all of the unambiguously random sample data on emotional problems was examined in the  
525 course of this analysis, the findings also make a strong substantive point: to date, no representative  
526 population data have found lower emotional problems among children with same-sex parents.  
527 Every random sample has observed higher emotional problems among such children; where the  
528 sample was large enough, those differences were statistically significant.

529

530 The more modest power appraisal and increased disclosure proposed by the three rules may  
531 inhibit the political use of such studies. However, research that violates the careful standards of  
532 scientific inference eventually becomes self-refuting as its bias becomes more generally known,  
533 and may actually hinder the development of more enlightened health and social policy regarding  
534 children in same-sex families.

535 This point is made best, perhaps, by the Australian National GLBTI Health Alliance, a coalition of  
536 groups devoted to the health of the same-sex population that has already been mentioned above.  
537 This national organization devoted to the health of same-sex persons complains: "An  
538 understanding as to whether LGBTI Australians are disproportionately affected by specific health  
539 issues can only at present be deduced from individual, often small, research studies which do not

540 cover the population as a whole”—such as the ACHESSE study and other small recruited sample  
541 studies. As a corrective, the National GLBTI Health Alliance calls for the collection of  
542 comprehensive, large-scale random-sample data by means such as “[t]he inclusion of questions on  
543 sexual identity and gender identity in the Census, the National Health Survey and other official  
544 statistics data collection” as well as government-funded grant research and funding for a large-  
545 sample national study of same-sex Australians. Cochran and colleagues likewise observe that,  
546 before the inclusion of sexual orientation measures in large public health surveys in the United  
547 States, earlier mental health research on minority sexual orientations was “plagued” by “the usual  
548 problems of sampling bias or absent heterosexual control groups”(17).

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

557

558



## 559 REFERENCES

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

1. Allen D. High school graduation rates among children of same-sex households. *Rev Econ Househ.* 2013 Dec 1;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. *SocSci Res.* 2012 Jul;41(4):735–51.
3. Manning WD, Fetto 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 Aug 1;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 Jun;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 Oct [cited 2014 Nov 7]. Report No.: ID 2500537. Available from: <http://papers.ssrn.com/abstract=2500537>
8. Goodman R. Psychometric Properties of the Strengths and Difficulties Questionnaire. *J Am Acad Child Adolesc Psychiatry.* 2001 Nov;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 Apr;48(4):400–3.
10. He J-P, 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 May 1;41(4):583–95.
11. Mellor D. Normative data for the strengths and difficulties questionnaire in Australia. *Aust Psychol.* 2005 Nov;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 May;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 SocBehav.* 2013 Mar 1;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 Aug 1;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 N Z J FamTher.* 2010;31(04):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.

- 609  
610  
611  
612  
613
20. Peck R, Devore J. Statistics: The Exploration & Analysis of Data. Cengage Learning; 2011. 816 p.
  21. Smith M. Common Mistakes in Using Statistics [Internet]. [cited 2014 Nov 11]. Available from: <https://www.ma.utexas.edu/users/mks/statmistakes/TOC.html>