

1
2
3
4
5
6
7
8
9
10
11

12
13
14
15
16

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

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.

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

35

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

50

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

61

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

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

74

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

78

79 **2. DATA AND METHODS**

80

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

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

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

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

116 **3. RESULTS**

117

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

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

SDQ Domains	NHIS			Mellor/ACHESSE		
	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	

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

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) ACHESSE = Australian Child Health in Same Sex Families

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

128

129 Differences between Australian and US SDQ norms are small, and largely orthogonal to the
 130 comparisons of interest in the present study. The NHIS data for OS parents in Table 1 is
 131 virtually identical to the US SDQ norms, which are derived from the 2001 NHIS. The US
 132 scale scores are systematically smaller, by about 20%, than the Australian scores, for

133 reasons which are beyond the scope of the present study. What is pertinent here is not the
134 absolute value of the scale scores, but the relative comparison of OS and SS parents on
135 each survey.

136
137 For almost all the SDQ domains, a striking and systematic contrast can be observed between
138 the two samples. On the NHIS, SS children have *higher* emotional problems than do OS
139 children on every measure, a difference that is statistically significant for five of the six
140 domains. On ACHESS, SS children have *lower* emotional problems than do OS children on
141 every measure except one; the difference is statistically significant for emotional problems.
142 Except as a general indicator of magnitude, statistical significance is not strictly relevant to
143 this analysis, since here we are comparing samples, not populations; and, as discussed
144 below, significance may not be meaningful for ACHESS generally.

145
146 Sampling differences between the OS and SS samples in the ACHESS comparison are
147 discounted by Crouch et al. (6), who in any event adopt the Mellor data for normative
148 comparison. Since it is highly unlikely that the NHIS, which during the years in question did
149 not even inquire about sexual orientation, is biased with regard to SS children, the contrast
150 in findings between NHIS and ACHESS must imply either that SS children suffer higher
151 emotional problems than do OS children in the USA, but not in Australia; or that the
152 ACHESS sample reflects bias in favor of reporting better outcomes for SS children than is
153 actually the case.

154
155 The former possibility seems unlikely given the large cultural and social similarities between
156 Australia and the USA, and in light of the fact that same-sex couples enjoy more liberal legal
157 and political status in the USA than they do in Australia; but it cannot be dismissed out of
158 hand. Bos suggests, for example, that higher stigmatization for SS children in the US than in
159 Holland may account, in part, for the lower rate of problem behaviors among the Dutch SS
160 children. However, Sullins documents that, on NHIS combined data from 2000-2013, SS
161 children experience less overall stigmatization than do OS children. If the higher emotional
162 problems observed for SS children in the US compared to those in Australia were due to
163 stigmatization, therefore, we would expect OS children also to experience higher emotional
164 problems than their Australian counterparts; but, as Table 1 shows, the opposite is the case.
165 On all the SDQ domains, OS children in the US have lower means for emotional difficulties
166 than do OS children in Australia.

167
168 Thus, although the possibility of substantive national differences cannot be definitively ruled
169 out, it seems much more likely that the differences observed reflect systematic differences in
170 the two samples. Indeed, this likelihood is explicitly supported by Crouch et al.: "The self-
171 selection of our convenience sample has the potential to introduce bias that could distort
172 results. ... If systematic bias was at play however, it would be anticipated that all outcome
173 variables would demonstrate [more favorable] scores across the sample" (6). That is
174 precisely the pattern observed (with one exception) in the present analysis.

175 176 **Broadening Scope**

177
178 So far we have observed a sharp contrast between the NHIS, which reports higher difficulties
179 for SS children on every measure based on a representative population sample, and
180 ACHESS, which reports lower emotional difficulties for SS children than for OS children on
181 every measure except one based on a convenience sample of recruited participants. If this
182 pattern is due to bias in the convenience sample compared to the probability sample, one
183 would expect other sample contrasts in this literature to show a similar pattern of differences.

184

185 In addition to Sullins and Crouch, since 2000 three other studies of SS parenting have
186 reported SDQ scale results for SS children. One of the three made use of a random
187 population sample comparable to NHIS(7); the other two used a recruited sample
188 comparable to ACHES(18,19). Altogether, these five studies comprise all gay parenting
189 studies that have used the SDQ (to my knowledge). Table 2 summarizes the results. The
190 two leftmost studies presented in the table report the two random samples; the remaining
191 three study results report the recruited samples. The rightmost of the four columns
192 summarizing each study's findings (labeled "F/U") reports whether each SDQ domain
193 comparison is Favorable or Unfavorable for SS children.

194
195 The dispositive issue for sample bias is whether the pattern already observed, in which the
196 random sample reflects unfavorable findings for SS children but the recruited sample reflects
197 favorable results, persists across the additional studies. Inspection of Table 2 reveals that
198 the answer to this question is an emphatic "yes". In the other random sample, reporting on a
199 different national population (England), SS children are found to have higher emotional
200 problems for every SDQ domain without exception, just as on NHIS. As with ACHES,
201 however, participants in the recruited samples report generally more favorable results for SS
202 children than for OS children. The distinction in results by sample type is clear, even across
203 two studies by the same lead author.

204
205 Despite Crouch's claim cited above, there is no reason to suppose that sample bias is
206 necessarily monotonic, or affects all studies or all behavioral domains to the same degree.
207 The exceptions to the pattern of "favorable random/unfavorable non-random" results,
208 moreover, are not randomly distributed. For four of the six SDQ domains—conduct
209 problems, hyperactivity, emotional symptoms and total difficulties—there are no exceptions
210 to the pattern, but the results are mixed for the other two domains--peer problems and pro-
211 social behavior. There may be many reasons for this difference among the recruited samples
212 that are consistent with the operation of bias, including the presence, to a greater or lesser
213 degree, of more than one form of bias. Perhaps parents of children with more externalizing
214 problems, such as misconduct or hyperactivity, were more likely to voluntarily deselect from
215 the recruited sample; or perhaps parents' knowledge of peer problems or pro-social behavior
216 on the part of their children, or their understanding of what constitutes social desirability in
217 these areas, was less complete or clear than it was for the other domains. Whatever the
218 reasons (and it is beyond the scope of this study to determine them), there is no
219 contradiction in the supposition that bias, if it is at work, may itself be biased in its effect on
220 study results.

221
222 To facilitate interpretation, Table 3 summarizes the proportion of favorable comparisons from
223 each study in Table 2, presented in the first five rows of the table. The top two rows report
224 on the two random samples; the next three rows present findings from the three recruited
225 samples. For further context Table 3 also adds findings from two other random sample
226 studies of SS parenting prior to 2010 that measured child emotional or behavioral difficulties,
227 but did not use the SDQ. Although some other studies in this developing literature included
228 elements of random sampling, Wainright and Patterson produced the only three studies that
229 made use of a statistically representative "blind" population sample of SS children,
230 identifying them in the data of the National Longitudinal Survey of Adolescent Health. Two
231 of these studies contained measures of emotional or behavioral distress for SS children
232 compared with OS children. In 2004 they reported comparative findings for depression, self-
233 esteem and anxiety(20), and in 2006 reported on eleven measures of delinquency among
234 adolescents such as binge drinking, illegal drug use or risky sexual behavior(21). All of the
235 above measures in both studies were less favorable for SS children in the sample, extending

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

SDQ Domains	US NHIS (2000-2004)				Golombok 2003 ¹				Bos 2010				Golombok 2014				Mellor/ACHESS 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)	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		

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

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.

237 the pattern already observed on the NHIS and Golombok's 2003 study. At this point, Table 3
 238 includes, to my knowledge, all the unambiguously random data on SS children gathered prior to
 239 2010.
 240

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

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

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

266 **Sources of Bias**

267
268 **By definition, bias denotes any form of** systematic error that inaccurately prejudices findings toward
269 one outcome over others. Every research effort is susceptible to bias; the scientific method
270 assumes the presence of potential bias and imposes procedures designed to minimize or specify it.
271 Bias in survey research may be present in how the respondents answer the questions (response
272 bias), how the sample is collected (selection bias), or how the results are interpreted (interpretation
273 bias). These elements may be expressed in this literature, as discussed above, as social
274 desirability bias in parental response, ascertainment or self-selection bias in the composition of the
275 sample, and researcher bias in interpreting results. Each of these will be briefly addressed in turn.
276

277 **Social Desirability Bias**

278
279 Because the ACHESS SDQ scales depend on parent reports, their findings are subject to social
280 desirability bias, that is, the tendency for parents to report more socially desirable information
281 about their own children (or themselves) than is objectively the case. To correct for this tendency
282 the SDQ is often administered to children's teachers and/or the children themselves, as well as
283 parents, and the two sets of results are compared. Both teacher and self-report ratings tend to be
284 lower than those of parents. Since the comparative OS data for the ACHESS scales are also
285 based on parent reports, social desirability would only bias the comparisons if the ACHESS parents
286 were more or less prone to give desirable responses than was the OS comparison group. ACHESS
287 surveyed parents using paper or online survey instruments, while the comparison group of OS
288 parents received telephone interviews; social desirability bias is known to be more common in
289 interviews, where the respondent is speaking with another human being, than on written or
290 computer surveys, where the respondent is more anonymous. On the other hand, the fact that all
291 the SS respondents, but none of the OS comparison group, were aware of the purposes of the
292 ACHESS study, and stood to benefit from any increased social acceptance or support that a
293 positive picture of SS children might help to bring about, may well have increased favorable
294 responses. We do not know how substantial such bias may be, but it would not be difficult to
295 determine by including teacher and/or self-reported ratings on the ACHESS SDQ scales. The
296 ACHESS researchers acknowledge the possibility of parent self-representation bias, and indicate
297 that future research will include "child-reported measures of health", as well as parent interviews.
298 The absence of such measures to date leaves uncertain the extent to which social desirability bias
299 may be present.

300 301 **Sample Selection Bias**

302

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

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

326 The ACHESSE protocol stated: "Primarily recruitment will be through emails posted on gay and
327 lesbian community email lists aimed at same-sex parenting"(24).The word "lists" here does not
328 refer to an actual sample frame. What actually appears to have been done is that announcements
329 were published on the two websites named and affiliated blogs. This is consistent with the other
330 recruitment activities called for in the protocol: advertisements, media releases, flyers and
331 researcher appearances at public events. This procedure increased probable bias in two
332 preventable ways. First, by making an open appeal for participation, rather than confining
333 themselves to lists of known population or target members, the researchers can provide no
334 assurance that respondents are actually part of the targeted population. A methodology report by
335 the ACHESSE researchers acknowledges that sample targets forwarded appeals to third parties,
336 who joined the sample in a snowball fashion, though they do not seem to recognize the problem
337 this creates (12). There is no way to know whether those responding to the appeal were members
338 or associates of the organizations targeted, or even same-sex parents. There is also no way to
339 know whether or not some persons, highly interested in the study and its outcomes, may have
340 responded multiple times. The possibility of such "sample crashing" is a prime reason why
341 sampled surveys, which contact known unique members of a sample frame, are less prone to bias
342 than are recruited samples. If the ACHESSE survey had actually used member lists, from which the
343 researchers contacted potential participants individually to invite their anonymous participation, this
344 form of potential bias would have been avoided.
345

346 While the respondent frame was far too wide, the appeal frame was much too narrow. Appeals for
347 participation focused on "gay and lesbian press ...gay and lesbian social and support groups, and
348 ...gay and lesbian community events."Two groups are mentioned in particular: recruitment "will
349 include, but not be limited to, Gay Dads Australia and the Rainbow Families Council of
350 Victoria"(26). The Rainbow Families Council of Victoria is not a social and support group, although

351 it may include those functions, but a political lobby. According to its published description:
352 "Rainbow Families Council replaces the Fertility Access Rights Lobby, which has worked for the
353 rights of same-sex parented families and prospective parents since 1999 and maintains a close
354 relationship with the Victorian Gay & Lesbian Rights Lobby"(27). Gay Dads Australia is a support
355 group for male same-sex parents focusing on adoption and surrogate parenting. The organization
356 actively promotes political causes relating to the legal status of same-sex parents, and views the
357 ACHES study as an ally in those efforts. The website for Gay Dads Australia features links to the
358 Rainbow Families Council, an LGBT political party, an advocacy group for same-sex marriage, an
359 LGBT radio station—and the ACHES study(26). The post on the Gay Dads Australia blog
360 encouraging participation in ACHES urged that the data "will help direct health policy, political
361 policy, anti-discrimination policy. Governments rely on good data. This is one way that we can all
362 help get that data"(29).

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

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

385 Bailey, in a study of the Australian Twin Registry that led to the rejection of the conclusions of his
386 own earlier study of homosexual men which had "recruited [participants] by means of
387 advertisements inhomophile publications or by word of mouth", observed that "[s]uch sampling is
388 likely to result in volunteer bias" that skews results. "Furthermore," he added, "respondents with
389 exclusively homosexual orientations may be overrepresented, and those with modest levels of
390 homosexual attraction, underrepresented, obscuring the potentially continuous
391 nature of sexual orientation. Cochran and colleagues likewise observe that, before the inclusion of
392 sexual orientation measures in large public health surveys in the United States, earlier mental
393 health research on minority sexual orientations was "plagued" by "the usual problems of sampling
394 bias or absent heterosexual control groups"(17).

395 Particularly striking is the absence of focus on organizations relating primarily to GLBT health and
396 well-being. The Australian National GLBTI Health Alliance lists several dozen members
397 geographically distributed throughout Australia which are concerned with health issues(30). Some
398 of the organizations have a political focus, but many do not. (Rainbow Families Council is a

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

405 **Interpretation Bias**

406

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

415

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

428

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

442

443 The analytical method of the ACHES findings report compounds this confusion, demonstrating
444 that this is not a tangential problem, but is central to the study. The mathematics of probabilistic
445 inference that underlie scientific research require that the characteristics of a non-random sample,
446 such as a recruited convenience sample, cannot be validly inferred to a population. Yet on the
447 convenience sample that was recruited largely from email lists of gay interest organizations, the
448 ACHES researchers incomprehensibly employ an extensive battery of population analyses,

449 including means comparisons with normative random population samples and tests of statistical
450 significance of the coefficients of multiple regression analyses, that are only appropriate for random
451 samples. They report, for example, that on three of the nine scales on the Child Health
452 Questionnaire SS children “demonstrated significant differences”, and that “[t]here were no
453 significant differences identified for other CHQ scales”; and that none of the SDQ scale scores for
454 SS children were significantly different than for a sample of OS children, listing the p-value for each
455 comparison. At one point they even speak of a scale score that “approached significance” with a
456 P-value of .052(6). These claims are incoherent. Statistical significance is, by definition, an
457 assessment of the variation or uncertainty in measurement resulting from random population
458 sampling. If the collected information on which the claim is based does not result from random
459 sampling, statistical significance can have no interpretation. By using random-sample statistics to
460 report apparent population inferences from a convenience sample, ACHESSE transforms what may
461 have been valuable non-parametric findings into unsubstantiated and possibly misleading
462 parametric claims.

463

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

475

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

482

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

498

499 3. Do not treat a non-probability sample as if it were a probability sample. Many of the
500 recruited sample studies in this field, including ACHESSE, make a formulaic disclaimer
501 about the representativeness of their sample, and then proceed to make population claims
502 anyway, as if the sample were statistically representative. This practice is misleading and
503 should be discontinued. This would include refraining to use inferential statistics or report
504 p-values that depend on the assumption of random sampling, and refraining from making
505 claims about population characteristics based on the sample. A recruited study should
506 include a strong, clearly-worded disclaimer to ensure it is not confused with a statistically
507 representative study, such as: “The sample in this study is not statistically representative
508 and cannot be used to validly infer the characteristics of any population in the real world.”
509

510 **4. CONCLUSION**

511
512 This study has found strong evidence of substantial bias understating the psychological difficulties
513 of children with same-sex parents on the Strength and Difficulties Questionnaire (SDQ) in studies
514 using recruited convenience samples. Well-informed health or social policy in this area, this
515 suggests, should not be based on studies with recruited samples, but on rigorous random sample
516 research. Likewise, scientific or scholarly outlets should refrain from publishing population claims
517 based on recruited samples. Recruited sample studies may have other value, however. A detailed
518 analysis of one such study, the Australian Study of Children in Same-Sex Families (ACHESSE)
519 offers suggestions, concluding in three common-sense rules, for ameliorating bias and preventing
520 misunderstanding.
521

522 Since all of the unambiguously random sample data on emotional problems was examined in the
523 course of this analysis, the findings also make a strong substantive point: to date, no representative
524 population data have found lower emotional problems among children with same-sex parents.
525 Every random sample has observed higher emotional problems among such children; where the
526 sample was large enough, those differences were statistically significant.
527 The evidence examined in this study is limited in several ways. It is quite possible that other
528 recruited sample studies, not examined here because they did not use the SDQ, have shown much
529 less or much more bias than the recruited sample studies analyzed here. It is also possible that
530 other measures, indicators or variables are less susceptible to bias in this area of study. A
531 comprehensive examination or meta-analysis of all extant studies of same-sex parenting would
532 provide helpful further evidence to confirm or rebut the suggestive conclusions of this study with
533 greater certainty.
534

535 The more modest power appraisal and increased disclosure proposed by the three rules presented
536 above may inhibit the political use of recruited sample studies of this population. However,
537 research that violates the careful standards of scientific inference eventually becomes self-refuting
538 as its bias becomes more generally known, and may actually hinder the development of more
539 enlightened health and social policy regarding children in same-sex families.

540 This point is made best, perhaps, by the Australian National GLBTI Health Alliance, a coalition of
541 groups devoted to the health of the same-sex population that has already been mentioned above.
542 This national organization devoted to the health of same-sex persons complains: “An
543 understanding as to whether LGBTI Australians are disproportionately affected by specific health
544 issues can only at present be deduced from individual, often small, research studies which do not
545 cover the population as a whole”(33)—such as the ACHESSE study and other small recruited
546 sample studies. As a corrective, the National GLBTI Health Alliance calls for the collection of
547 comprehensive, large-scale random-sample data by means such as “[t]he inclusion of questions on

548 sexual identity and gender identity in the Census, the National Health Survey and other official
549 statistics data collection” as well as government-funded grant research and funding for a large-
550 sample national study of same-sex Australians(33).

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

559
560

561 **REFERENCES**

562

563

- 564 1. Allen D. High school graduation rates among children of same-sex households. *Rev Econ*
565 *Househ.* 2013 Dec 1;11(4):635–58.
- 566 2. Marks L. Same-sex parenting and children’s outcomes: A closer examination of the
567 American psychological association’s brief on lesbian and gay parenting. *SocSci Res.* 2012
568 Jul;41(4):735–51.
- 569 3. Manning WD, Fetto MN, Lamidi E. Child Well-Being in Same-Sex Parent Families: Review
570 of Research Prepared for American Sociological Association Amicus Brief. *Popul Res Policy Rev.*
571 2014 Aug 1;33(4):485.
- 572 4. Anderssen N, Amlie C, Ytterøy EA. Outcomes for children with lesbian or gay parents. A
573 review of studies from 1978 to 2000. *Scand J Psychol.* 2002;43(4):335–51.
- 574 5. Rosenfeld MJ. Nontraditional families and childhood progress through school.
575 *Demography.* 2013 Jun;50(3):963–9.
- 576 6. Crouch S, Waters E, McNair R, Power J, Davis E. Parent-reported measures of child
577 health and wellbeing in same-sex parent families: a cross-sectional survey. *BMC Public Health.*
578 2014;14(1):635.
- 579 7. Sullins DP. Child Emotional Problems in Non-Traditional Families [Internet]. Rochester,
580 NY: Social Science Research Network; 2014 Oct [cited 2014 Nov 7]. Report No.: ID 2500537.
581 Available from: <http://papers.ssrn.com/abstract=2500537>
- 582 8. Goodman R. Psychometric Properties of the Strengths and Difficulties Questionnaire. *J Am*
583 *Acad Child Adolesc Psychiatry.* 2001 Nov;40(11):1337–45.
- 584 9. Goodman A, Goodman R. Strengths and Difficulties Questionnaire as a Dimensional
585 Measure of Child Mental Health. *J Am Acad Child Adolesc Psychiatry.* 2009 Apr;48(4):400–3.
- 586 10. He J-P, Burstein M, Schmitz A, Merikangas K. The Strengths and Difficulties Questionnaire
587 (SDQ): the Factor Structure and Scale Validation in U.S. Adolescents. *J Abnorm Child Psychol.*
588 2013 May 1;41(4):583–95.
- 589 11. Mellor D. Normative data for the strengths and difficulties questionnaire in Australia. *Aust*
590 *Psychol.* 2005 Nov;40(3):215–22.
- 591 12. Crouch S, Waters E, McNair R, Power J, Davis E, van Mourik L. Triumphs and challenges
592 in recruiting same-sex parent families. *Aust N Z J Public Health.* 2014;38(1):87–8.
- 593 13. Cochran SD, Mays VM. Risk of Breast Cancer Mortality Among Women Cohabiting with
594 Same Sex Partners: Findings from the National Health Interview Survey, 1997–2003. *J Womens*
595 *Health.* 2012 May;21(5):528–33.
- 596 14. Liu H, Reczek C, Brown D. Same-Sex Cohabitors and Health: The Role of Race-Ethnicity,
597 Gender, and Socioeconomic Status. *J Health SocBehav.* 2013 Mar 1;54(1):25–45.
- 598 15. Reczek C, Liu H, Brown D. Cigarette Smoking in Same-Sex and Different-Sex Unions: The
599 Role of Socioeconomic and Psychological Factors. *Popul Res Policy Rev.* 2014 Aug 1;33(4):527–
600 51.
- 601 16. Reczek C, Liu H, Spiker R. A Population-Based Study of Alcohol Use in Same-Sex and
602 Different-Sex Unions. *J Marriage Fam.* 2014;76(3):557–72.
- 603 17. Golombok S, Perry B, Burston A, Murray C, Mooney-Somers J, Stevens M, et al. Children
604 with lesbian parents: a community study. *Dev Psychol.* 2003;39(1):20.
- 605 18. Bos HH. Planned gay father families in kinship arrangements. *Aust N Z J FamTher.*
606 2010;31(04):356–71.
- 607 19. Golombok S, Mellish L, Jennings S, Casey P, Tasker F, Lamb ME. Adoptive Gay Father
608 Families: Parent–Child Relationships and Children’s Psychological Adjustment. *Child Dev.*
609 2014;85(2):456–68.
- 610 20. Wainright JL, Russell ST, Patterson CJ. Psychosocial adjustment, school outcomes, and

611 romantic relationships of adolescents with same-sex parents. *Child Dev.* 2004;75(6):1886–98.
612 21. Wainright JL, Patterson CJ. Delinquency, victimization, and substance use among
613 adolescents with female same-sex parents. *J Fam Psychol.* 2006;20(3):526.
614 22. Peck R, Devore J. *Statistics: The Exploration & Analysis of Data.* Cengage Learning; 2011.
615 816 p.
616 23. Smith M. Common Mistakes in Using Statistics [Internet]. [cited 2014 Nov 11]. Available
617 from: <https://www.ma.utexas.edu/users/mks/statmistakes/TOC.html>
618 24. Bailey JM, Bobrow D, Wolfe M, Mikach S. Sexual orientation of adult sons of gay fathers.
619 *Dev Psychol.* 1995;31(1):124.
620 25. Bailey JM, Dunne MP, Martin NG. Genetic and environmental influences on sexual
621 orientation and its correlates in an Australian twin sample. *J PersSoc Psychol.* 2000;78(3):524.
622 26. Crouch SR, Waters E, McNair R, Power J, Davis E. ACHES—The Australian study of child
623 health in same-sex families: background research, design and methodology. *BMC Public Health.*
624 2012;12(1):646.
625 27. Rainbow Families Council [Internet]. [cited 2014 Dec 10]. Available from:
626 <http://www.rainbowfamilies.org.au/>
627 28. Gay Dads Australia [Internet]. [cited 2014 Dec 10]. Available from:
628 <http://www.gaydadsaustralia.com.au/>
629 29. Gay Dads Australia [Internet]. Be Part of the World’s Largest Study into the Health and
630 Wellbeing of our Children [blog announcement]. 2012 [cited 2014 Nov 12]. Available from:
631 <http://gaydadsaustralia.blogspot.com/>
632 30. National LGBTI Health Alliance Members | National LGBTI Health Alliance [Internet]. [cited
633 2014 Dec 11]. Available from: <http://lgbtihealth.org.au/members>
634 31. Crouch S. It’s time to adopt change | Election Watch [Internet]. [cited 2014 Dec 10].
635 Available from: <http://electionwatch.edu.au/victoria-2014/it%E2%80%99s-time-adopt-change>
636 32. Crouch S, Block K, Davis E, Waters E, Gibbs L. How voting for equity will make life better
637 now and in the future [Internet]. *The Conversation.* 2013 [cited 2014 Dec 11]. Available from:
638 <http://theconversation.com/how-voting-for-equity-will-make-life-better-now-and-in-the-future-17167>
639 33. Researching LGBTI Health | National LGBTI Health Alliance [Internet]. [cited 2014 Dec 11].
640 Available from: <http://lgbtihealth.org.au/research>
641