Exhibit C

Court File No. 684/00

ONTARIO
SUPERIOR COURT OF JUSTICE
(DIVISIONAL COURT)

BETWEEN:

HEDY HALPERN and COLEEN ROG
MICHAEL LESHNER and MICHAEL S
MICHELLE BRADSHAW and REBEKAH
ALOYSIUS PITTMAN and THOMAS ALI
DAWN ONISHENKO and JULIE ERBI
CAROLYN ROWE and CAROLYN MOI
BARBARA McDOWALL and GAIL DON
ALISON KEMPER and JOYCE BARI

-and-

THE ATTORNEY GENERAL OF CANADA, et al

Respondents

and

EGALE CANADA INC. et al

Intervenors

-AND-

Court File No. 39/2001

BETWEEN:
METROPOLITAN COMMUNITY CHURCH OF TORONTO

Applicant

-and-

THE ATTORNEY GENERAL OF CANADA, et al

Respondents

and

HEDY HALPERN, et al

Intervenors

AFFIDAVIT OF STEVEN LOWELL NOCK

ONTARIO
SUPERIOR COURT OF JUSTICE
(DIVISIONAL COURT)

BETWEEN:

**HEDY HALPERN and COLEEN ROGERS,
MICHAEL LESHNER and MICHAEL STARK,
MICHELLE BRADSHAW and REBEKAH ROONEY,
ALOYSIUS PITTMAN and THOMAS ALLWORTH,
DAWN ONISHENKO and JULIE ERBLAND,
CAROLYN ROWE and CAROLYN MOFFATT,
BARBARA McDOWALL and GAIL DONNELLY,
ALISON KEMPER and JOYCE BARNETT**

Applicants

-and-

THE ATTORNEY GENERAL OF CANADA, et al

Respondents

and

EGALE CANADA INC. et al

Intervenors

-AND-

BETWEEN:
METROPOLITAN COMMUNITY CHURCH OF TORONTO

Applicant

-and-

THE ATTORNEY GENERAL OF CANADA, et al

Respondents

and

HEDY HALPERN, et al

Intervenors

<u>AFFIDAVIT OF STEVEN LOWELL NOCK</u>

I, Steven Lowell Nock, of the City of Charlottesville, in the State of Virginia, in the United States of America, make oath and say as follows:

1.   I have been asked by the Attorney General of Canada to apply my expertise in research methodology and evaluate the scientific literature concerning the effect of legal recognition of the marriages of gay and lesbian couples on their children, cited in the affidavit of Professor Jerry Bigner, sworn November 11, 2000, and filed on behalf of the Applicants in this case. The articles upon which Professor Bigner's opinion rests are contained in his Exhibit "B" and were previously relied on in the Brief to the court in Vermont, in the case of *Baker v. Vermont.* I have read and evaluated each of those articles.

2.   My affidavit is divided into two main segments. In the first, I explain the principles of sound social science research methodology. I describe the characteristics of good research design and highlight the pitfalls that result from the failure to apply proper design techniques. Clearly, where the design of research is substandard, it is dangerous to rely on the conclusions reached if they are intended as truths.

3.   In the second segment of this affidavit I analyze the studies presented by Professor Bigner for their value and reliability in supporting the assertions that Professor Bigner says they support. I do this analysis with reference to the accepted methodological techniques and terms described in the first segment of this affidavit. Through this analysis I draw my conclusions that 1) all of the articles I reviewed contained at least one fatal flaw of design or execution; and 2) not a single one of those studies was conducted according to general accepted standards of scientific research.

4.   The task I undertook was to evaluate the relevant studies simply from the standpoint of whether or not they provide reliable answers to the questions or hypotheses their authors intended to address. As a result, my analysis is made solely from the perspective of a research methodologist. I do not make any claim regarding the inherent

2

truth or falsity of any of the hypotheses proposed to be tested in the studies, or of any converse hypotheses. It is the policy maker who depends on the truth value alleged in the results and conclusions reached through social science. With this in mind, only objective and sound methodological analysis can fulfill the need.

I. Qualifications

5.   I am currently a Professor of Sociology at the University of Virginia where I have taught since 1978. I teach both undergraduate and graduate courses. At the undergraduate level, I teach Research Methods, The Family, and Family Policy. At the graduate level, I teach Research Design, Intermediate Graduate Statistics, and Family Research.

6.   I am co-founder of the Center for Children, Families, and the Law at the University of Virginia, a multi-disciplinary center to foster collaborative research and teaching on issues involving children and families.

7.   My research focuses primarily on households and families. I am concerned with the causes and consequences of changes in family organization and structure. Thus, I have investigated marriage, divorce, and cohabitation by focusing on the factors that lead individuals into these statuses and the consequences of entering them. I am the author of six books and over 50 articles and chapters that are detailed in my *curriculum vitae*, attached as Exhibit "1" to this affidavit. Almost everything I have published relies on quantitative analysis of large, nationally representative samples of adults. My most recent book (*Marriage in Men's Lives*) was based on a statistical analysis of 6,000 men interviewed annually from 1979 through 1993. The book was the recipient of the 1999 American Sociological Association William J. Goode Book Award for the most outstanding contribution to family scholarship.

8.	I am also Director of the Marriage Matters Project which is a five-year research effort supported by the National Science Foundation and the Smith Richardson Foundation. This research investigates the legal innovation known as Covenant Marriage in Louisiana. It is a quantitative effort involving approximately 1,200 individuals interviewed repeatedly over the course of five years.

9.	I currently serve as Associate Editor for *Journal of Marriage and the Family* and *Social Science Research.*

II. Relevant Issues Of Research Design
a. Introduction

10.	Before evaluating the specific claims made by Professor Bigner in his affidavit, I first want to outline the strategies that would produce scientifically acceptable research results concerning the effect of legal recognition of the marriages of gay and lesbian couples for the children in such unions. These strategies are the basis of my evaluation of the articles contained in Professor Bigner's brief as they conform to accepted standards for scientific research.

11.	Let me begin by noting that the central question, that is, what effect does gay and lesbian marriage have on children in such unions, cannot be answered at the moment. With the exception of the extremely recent change in the Netherlands, no jurisdiction has yet to recognize the unions of gays and lesbians as marriages. As a result, it is clearly impossible to evaluate how such a change has affected the children involved.

12.	Since it is not possible to consider this research question (i.e., would the legal recognition of the marriages of gay and lesbian couples affect the children in such unions), we are left to consider a related question. As I see the issue, there are actually two such questions, only one of which can be answered. First, and most importantly, does a homosexual union of adults cause the children to develop differently than they

would have if they had heterosexual parents (or some other arrangement)? This is a researchable question that can, in principle, be answered. However, the simple fact is, to date, this question has not been answered. Second, does marriage change the behavior of gay or lesbian parents toward their children or toward each other (i.e., does marriage cause relationships to be more stable, cause parents to treat children differently etc.)? While this second question has been addressed with respect to heterosexuals, it cannot be answered with respect to homosexual parents because there has never been a legal marriage of homosexuals. Any answer to this question in regard to homosexual marriage is purely hypothetical.

13.　　　　In the comments that follow, I have assumed that the following statement, found in the affidavit of Dr. Jerry Bigner, guides the research: Is it true that "The children of gay and lesbian parents are as healthy and well adjusted as those of their heterosexual counterparts?" (Bigner affidavit, page 6)

b. Correlation and Causation

14.　　　　Before discussing how we might address such a question, I want to distinguish between "correlation" and "causation." When two things are correlated, we can show that they tend to vary together. That is, different levels of one tend to be associated with different levels of the other. A well-known example of correlation is the relationship between educational attainment and income. Those with higher levels of educational attainment have higher average incomes. Another well-known example of a correlation is the relationship between divorce and children's educational attainment. Children who experience their parent's divorce before age 16 complete fewer years of schooling, on average. Both of these correlations are well known, and have been replicated enough times to confirm their existence.

15.　　　　Correlation, of course, does not necessarily imply causation. That is, in trying to understand what relationship one factor has to the other, it is very unsound to assert that the correlation between educational attainment and income reflects a causal path between the first and the second. Nor is it sound to assert that the correlation

between divorce and educational attainment means that divorce is the cause of children completing fewer years of schooling. From the perspective of a research methodologist, it would be foolish – and, indeed, unsound – to make such causal assertions without more evidence than a simple correlation.

16. To determine that a causal connection exists between any two factors X and Y requires three things:

- X and Y must be correlated;
- X must precede Y temporally; and
- No third factor Z can explain the relationship between X and Y.

17. In the case of educational attainment and income, for example, there is no question that the two are correlated. Nor is there much question that educational attainment typically precedes the earning of income. But what about the existence of a possible third factor? What if high intelligence is the true cause of both higher educational attainment *and* higher income? If so, then the correlation between education and income is spurious. It exists only because the two items share a common cause. We can apply the same logic to the divorce-education example. If poverty is a primary cause of divorce and of poor educational attainment, the correlation between divorce and education is spurious.

18. The primary question that has been asked in the research referred to in the case at hand is, in my opinion, causal in nature. "Does having gay or lesbian parents cause children to differ (from others) in consistent ways. I address how we might answer this and related questions in a way that produces reliable results from the perspective of sound research methodology.

19. To show that having gay/lesbian parents *causes* children to differ, we would need to do three things. First, we would need to show that there is a correlation between living with gay/lesbian parents and some outcome in the lives of children. Second, we would need to show that exposure to gay/lesbian parents happened before the

outcome did. And finally, we would need to show that there is no other factor that is a common cause of both.

20.   In a related way, how would we show that there is *no* causal relationship between gay/lesbian parents and children's well being? This requires somewhat less evidence. To establish the validity of such a claim would require only that no correlation be found between the sexual orientation of parents and the child's well being once all other factors have been controlled. If a valid and scientifically adequate study were to show that there is no correlation between having gay or lesbian parents and a child's well being, based on a comparison of representative groups of each type of parent, and differing only on sexual orientation, then most scientists would accept that there is no causal link between the two.

III. The Design Of The Study

a. Introduction

21.   In the following section, I discuss the relevant issues required to conduct a study to answer the question being asked in this case. Several methodological issues must be satisfied before one may attempt to investigate the relationship being discussed. In the following sections, I summarize and explain these issues as they pertain to the case at hand. Once I have done that, I turn to the evidence included in Professor Bigner's affidavit. I evaluate that evidence on the various design and sampling criteria I discuss below.

b. Sampling.

22.   First and foremost, the ability of any social-science evidence to apply to a larger group depends on the way the sample of cases was obtained. A "probability sample" is one in which every member of a *definable population* has a known probability

of being included in the study. A probability sample is always necessary in order to generalize one's results. The simplest form of probability sampling is known as "simple random sampling" (SRS). In SRS, a researcher first defines some population to which she or he wishes to generalize the results of the study. This may be a population as large as all voting adults in Canada, all adults in Canada, all children in primary grades, or as small as all patients with newly diagnosed breast cancer. Regardless of the population of interest, the researcher must be able to define it. Once defined, every member of the population must have an equal chance of being selected for participation in the study. If, for example, the population was defined as the 480,000 (1996) residents age 25 and older in the geographic limits of the city of Toronto (at that time), then every single one of these 480,000 residents must have the same chance of being selected into the sample. Simple random sampling guarantees that the chances of selection (from the defined population) are equal for all cases. A detailed explanation of how simple random sampling is achieved is contained in the paragraphs I have written in Appendix I to this affidavit.

23.   As indicated, a probability sample is required whenever a researcher wishes to make claims about the larger population from which the sample was drawn. If the goal is to make general claims about same-sex parental relationships and the children who might be affected by them, then we must have a probability sample drawn from the larger population of homosexual parents and children.

24.   A probability sample does not guarantee that the results will fairly and accurately describe the larger population. Indeed, it is possible for such a sample to err in large and important ways. For example, imagine drawing a simple random sample of 1,000 from all employed persons aged 15 and older with reported incomes in the Toronto metropolitan area. We know that the average (annual, 1995) income reported by Statistics Canada for this group of Toronto residents is $28,980 [1]. But is it possible that our

---

[1]

http://CEPS.statcan.ca/english/profil/Details/details1inc.cfm?PSGC=35&SGC=53500&A=&LANG=E&Province=All&PlaceName=toronto&CSDNAME=Toronto&CMA=535&DataType=1&TypeNameE=Census%20Metropolitan%20Area&ID=605

random sample could produce an average of $38,980, an average that is $10,000 higher than the actual value at that time? It is quite possible. Since the sample was drawn randomly, it is possible that an unrepresentative group of 1,000 people was selected. But it is not *probable*. In fact, such a result would be extraordinarily unlikely. And that is the important point about probability samples; we are able to calculate how unlikely such a result would be.

c. Probability Theory

25.　　　In practice, we cannot know if our particular probability sample is a fair and representative reflection of the population from which it was drawn. As a consequence, we apply probability theory to the results obtained from such samples. Rather than claim that our results do, in fact, reflect the true situation in the population, we attach a probability of error to any such claims. This is what is meant by "statistical significance."　The statistical significance of any sample result refers to the probability that the true (but unknown) value in the population differs from that result.

26.　　　There is no alternative to the use of probability theory when the goal is to generalize from a sample to a larger population. And there is no alternative to a probability sample when one applies probability theory. Without a probability sample, a researcher cannot use statistics that are designed to generalize from samples to populations (i.e., inferential statistics). Though this is sometimes done, the researcher who does so has violated the most basic premise of inferential statistics.

d. Variations in Sample Quality

27.　　　The quality of samples varies enormously in social science research. Deviations from pure random sampling are not uncommon. But the quality of the sample is directly related to the intended use of the information obtained from it. At one extreme there is exploratory data gathering that is merely intended to generate ideas and hypotheses for more systematic analysis at a later stage. Examples of such samples

include undergraduate students taking a course from a professor, or "mall-intercept" interviews (where a researcher recruits people as they walk by in a shopping mall). At the other extreme are large-scale continuing studies that are used to supply information for policy decisions of the federal government. A good example is the Current Population Survey (CPS) conducted by the U.S. Bureau of the Census for the Bureau of Labor Statistics every month. The CPS has been conducted for 50 years, and provides information about consumer behavior, income trends, and related economic indicators.

28.   Particularly relevant to the current issue are instances where a population is difficult to define or identify. Such rare populations present problems since no lists are available to identify them. Locating these populations then requires a search for a probability sample of the general population (i.e., a screening of the general population to identify the members of the rare population). Appropriate techniques exist for such problematic cases, and typically require screening. For example, if a researcher is interested in obtaining a sample of individuals who smoke pipes, a large general population sample would be contacted, and each respondent asked whether he or she smokes a pipe. Sometimes, such screening is made more efficient when the researcher is able to identify geographic clusters (regions) that have higher rates of the rare cases. It is also more efficient if the researcher is able to identify those clusters with no rare cases.

e. Sampling Issues for Research in this Case

29.   We do not have a precise estimate of the prevalence of homosexuality in the general population. And sampling is complicated by the stigma associated with the issue. Still, no published estimate that I know of has placed the prevalence above 10%. The most-cited source for the 10% estimate of "more or less exclusively homosexual males" is the work of Kinsey and associates from the late 1940s.[2] Unfortunately, Kinsey's research did not use a probability sample. Moreover, we do not have an agreed-upon definition of homosexuality. Is a homosexual a person whose erotic interests are

10

focused on those of the same sex? Is a homosexual a person who sometimes engages in sexual acts with a member of the same sex? Is a homosexual a person who thinks of himself or herself as a homosexual? Does a single sexual act with a person of the same sex define a person as a homosexual? Also important in the case is how to define "bisexual?" Are bisexuals to be treated as homosexuals, heterosexuals, or both? And how does one decide? Is homosexuality "learned" (i.e., socially constructed), or is it transmitted genetically? Finally, is male homosexuality the same phenomenon as female homosexuality? Answers to such questions have direct and important consequences for how one investigates the topics in this case.

30.   Unless the researcher is able clearly to define what "homosexual" means, he or she is forced to let subjects define the terms as they wish. In the research relied on by Professor 7Bigner, which I reviewed for my opinion on its validity and reliability, this is what was done. Researchers allowed subjects to define themselves as homosexual or heterosexual without further specifications. Quite simply, by relying on volunteers (rather than a sample defined by some specific definition), the researchers cannot know what is being studied. More critically, the use of volunteers means that it will never be possible to replicate the findings of the research. Should another researcher conduct a similar study but find different results, it will be impossible to know why.

31.   Depending on how one defines the term homosexual (or gay, or lesbian), different estimates of the prevalence are obtained. The work of Laumann, Gagnon, Michael, and Michaels (1994)[3] was based on personal (face-to-face) interviews with a probability sample of 3,432 adults and is probably the best source of information currently available on the prevalence of homosexuality in the United States. The population to which this sample may be generalized includes all English-speaking adults between the ages of 18 and 59 who resided in households (i.e., not institutions) at the time of the study. Using various definitions of homosexuality, these researchers found

---

[2] Sexual behavior in the human male by Alfred C. Kinsey, Wardell B. Pomeroy and Clyde E. Martin. (1948); Sexual behavior in the human female, by the staff of the Institute for Sex Research, Indiana University: Alfred C. Kinsey and others (1953).

that rates varied somewhat by sex when the question pertained to sexual behavior with a person of the same sex, as seen below:

1. Any same sex partners in the past 12 months?   1.3% women, 2.7% men

2. Any same sex partner since puberty?   3.8% women, 7.1% men

32.   When the researchers asked about attraction to members of the same sex, or sexual desire for members of the same sex (alternative definitions of homosexuality), somewhat different values were obtained, with higher rates of "desire" and "attraction" than observed for behavior. And when asked about sexual identity (how one thinks of oneself), rates were different yet, with 1.4% of women, and 2.8% of men identifying with a label denoting same-sex sexuality.

33.   Recently published research based on several large, nationally representative probability samples of all English-Speaking non-institutionalized adults age 18 and over[4] produced comparable rates of prevalence. Most of the data in this study were obtained with anonymous, self-administered questionnaires rather than face-to-face interviews. By combining years of the General Social Survey from 1988-1991, 1993, 1994, and 1996, as well as evidence from the Laumann, et. al. study just described, these authors report that 3.6% of women, and 4.7% of men have had at least one same-sex partner since age 18. Only 1.5% of women and 2.6% of men had exclusively same-sex partners in the last 5 years.

34.   I was unable to locate any probability samples of Canadian homosexuals and will, therefore, use U.S. estimates in this section.[5] How rare is the homosexual population in the United States? If we take the studies just mentioned as the best evidence, we would conclude that somewhere between 1% and 4% percent of adult

---

[3] E.O. Laumann, J.H. Gagnon, R.T. Michael, and S. Michaels. The Social Organization of Sexuality: Sexual Practices in the United States. 1994. Chicago: University of Chicago Press. Chapter 8.

[4] D. Black, G. Gates, S. Sanders, and L. Taylor. "Demographics of the Gay and Lesbian Population in the United States: Evidence from Available Systematic Data Sources." Demography, 37 (No. 2) 2000: pp139-154.

[5] However, based on my understanding that, to a large degree, the populations of the United States and Canada share common roots and cultural, at present I have no reason to believe that the results would be radically different.

American women, and between 3% and 7% of adult American men are homosexual by at least one definition of that term. Surely this is a relatively rare population, yet one sufficiently large to allow researchers to rely on probability samples for analysis. Still, even when large, nationally representative samples are used, the proportion of homosexuals who might be parents will be smaller, clearly, than these low figures. In sum, the population of homosexual adults is small. An adequate probability sample of such a population, would requires a large amount of screening to produce as many as 500 homosexual parents.

35.   If we take a midpoint estimate as the best guess of prevalence, then we would expect approximately 2.5% (halfway between 1% and 4%) of female and 5% of male (halfway between 3% and 7%) subjects to be identified as homosexuals by at least one definition of the term. If a researcher screened 20,000 individuals for study, hoping to generate a probability sample of homosexuals, we would expect to obtain approximately 500 female and 1,000 male subjects for analysis. Of these, only a fraction would be parents. As a very crude estimate of that fraction, we might consider the fraction of couples living in common-law relationships in Canada who live with children, or the fraction of married couples in Canada that live with children. (I use these two groups on the assumption that, at this point in time, the vast majority of homosexual parents bore their children in marriages or common-law heterosexual unions.) The 1996 Canadian Census found that 47% of Common-law Couples, and 61% of Married Couples have children at home. Therefore, I would expect that homosexual adults would fall midway between these two values. Assuming that 54% of homosexuals are, at present, parents, this means that about half of any sample of homosexuals would initially qualify for our study. Of the 1,500 homosexuals identified by our screening methods, we would expect 810 currently to be parents. Further qualifications would likely reduce this number further, because not all of these homosexual parents would be living, or have lived, with their children. I have no evidence that would allow me to estimate that fraction. For the sake of illustration, however, let us assume that the fraction of homosexual adults who currently live with their children is 50%. Now our sample has been cut to only 405.

36.   With current statistical methods, such samples would be adequate for preliminary research. Samples of twice this size would be adequate for almost any statistical purposes. If our goal is to produce a nationally representative sample of homosexuals sufficient to support most multivariate statistical techniques of the type needed to answer the questions at hand, we would probably need to screen about 40,000 individuals. This is not a particularly large screening task, however. For example, the Current Population Survey (U.S. Bureau of the Census) interviews (not simply screens) approximately 50,000 individuals every month. Still, the sampling task is challenging, and very expensive. But most importantly, in relation to the issues at hand, no one has done this to date.

37.   To put the sampling problem in perspective, 2.8% of Canadians are members of an Aboriginal group, 2.5% of Canadians are Baptists, and 5.6% of Canadians are at least 75 years old[6] The sampling task that would be involved in a study of gay and lesbian adults (of which some fraction would be parents) is comparable to the challenge faced by any researcher hoping to study one of these populations in Canada and obtain conclusive results that may be relied on to make very important, or potentially irreversible, policy decisions.

38.   Sampling rare populations is a challenge that researchers face all the time.[7] Homosexuals are probably no more difficult to locate and interview than homeless individuals, those who have been the victim of crimes in the past year (without reporting the incident to the police), or those who have had abortions. All have been the subject of scientific investigation. The crucial point is, however, that without a sample of the type just described, it is impossible to make scientifically valid claims about the population of homosexuals and their children.

---

[6] http://www.statcan.ca/english/Pgdb/People/popula.htm#oth
[7] My ongoing research about the legal innovation known as "covenant marriage" in Louisiana, focuses on a very rare population. Fewer than 5% of all new marriages in the state are celebrated as covenant marriages. Newly married people, moreover, are a small fraction of all people in the state. Still, I have been able to assemble a probability sample of approximately 600 individuals who have entered covenant marriages within the past 12 months with response rates ranging from 65% to 75% (depending on the month). It is, indeed, difficult to locate and interview people who are in rare populations.

f. 'Convenience' Methods of Sampling

39.		Before concluding, a brief note should be added about obtaining samples when probability methods are not used. All such strategies depend on various types of 'convenience' methods. Sometimes researchers will recruit subjects into a study by placing advertisements in various outlets. Sometimes researchers will resort to "snowball" sampling, where a subject mentions another, who mentions another and so on. And, sometimes researchers will use an existing group (e.g., students in a class, members of an organization). No such method is permitted by sound scientific methodology when the goal is to generalize to a population, because all such samples are biased in unknown ways. Particularly problematic with rare samples is the snowball strategy. The reason this strategy is so bad is because individuals who are well-known are more likely to be mentioned than those who are not well-known. And well-known individuals in rare populations often differ in important ways from those who are less well known. A well-known lesbian (if the individual's decision to be known as lesbian is a well-considered decision) is likely to be a different type of lesbian than is her less well-known counterpart.

40.		The simplest way to understand why a sample might be biased is to consider a convenience sample recruited from an organization devoted to seeking equal rights for gays and lesbians. Suppose that the homosexual participants in this group have higher levels of education than comparable homosexuals who are not members of this group. If this group were used for research purposes, then anything that is correlated with educational attainment would be biased. For example, we know that higher education is associated with better health. If we extrapolated (generalized) about the health of homosexuals from this sample, we would be making claims about a population based on a group that does not represent it. The reported health of this particular group would probably be better than would a representative sample.

g. Cross-sectional vs. Longitudinal Studies

41.   The conclusion must be that a scientific study of how parents' homosexuality affects children must begin with a probability sample of a well-defined population. However, once the population has been defined, and before the execution of the actual sampling, one additional issue must be resolved.

42.   Depending on the topic being studied, the researcher must decide whether to conduct the study only once, or conduct it repeatedly over time.   The former is typically known as a cross-sectional study and the latter as a longitudinal study. When the only goal is to estimate percentages, rates, and such descriptive information about a population, then a cross-sectional study is often adequate. However, when the goal is to produce evidence about cause, as in the present case, cross-sectional studies are considered especially weak. Longitudinal studies are always preferred when the issue is one of cause-effect.

43.   In a cross-sectional study, a group of individuals is contacted once (or several times in quick succession -- for example, several interviews in the course of a day). Information obtained in this way is limited in its ability to produce evidence of change. Without evidence of change, there is very little one can say about cause.

44.   The problems of cross-sectional studies are particularly severe when the temporal ordering of the phenomena in question is unclear, that is, where the cause and effect of the two correlated factors may go either way. For example, repeated studies have found that politically conservative individuals have higher incomes. If one were attempting to draw causal conclusions about this correlation, it would be impossible to conclude that higher incomes cause people to become more conservative, because, just as likely, is that holding conservative political positions causes people (for whatever reason) to earn more money. And, of course, as discussed at the outset, there may be absolutely no causal connection between political conservatism and income simply because the two factors are correlated.

45.   The second requirement for establishing causation (noted above) is that the cause must *precede* the effect in time. While this is often impossible to determine with absolute certainty, the scientific plausibility of this claim is enhanced significantly when the researcher is able to observe the same individuals repeatedly, over time.

46.   Longitudinal studies of the same individuals are known as panel studies. A panel is a group of individuals who are observed, or who answer questions repeatedly over a specified period of time. Well-known examples of large panel studies include the Panel Study of Income Dynamics (PSID), the National Survey of Families and Households (NSFH), the U.S. Census Bureau's Survey of Income and Program Participation (SIPP), and the National Longitudinal Survey of Youth (NLSY). Each of these panel studies includes at least 5,000 individuals who were studied at least twice.

47.   When interest focuses on developmental issues (phenomena that emerge over time) a panel study is particularly important. Some processes may require years to become obvious, while others may become immediately apparent. To the extent that the process being investigated develops slowly over the course of several years, then a panel study of long duration is needed to capture this event. If, for example, a researcher studied the transmission of homosexuality from parent to child, what could be learned by a study of 8-year old children? Perhaps a great deal. But more likely, such a study would need to follow these children for several years to investigate the possibility of change over time. A longitudinal study would need to be started when children are young (perhaps 2 or 3), and would need to follow children throughout a significant period of their lives to measure any possible changes.

48.   If a researcher is able to show that whenever an individual changes (over time) on one dimension, he or she also changes in predictable ways on another dimension, this is strong (though not incontrovertible) evidence of a causal connection. Thus for example, in my research on marriage in which I relied on a panel study of 6,000 men interviewed annually for 13 years, I was able to show that when men got married (i.e., changed from being single to being married), their incomes also changed by a

predictable amount and direction. I also found that a change in marital status was accompanied by a change in men's propensity to give help to others. These and similar patterns led me to suggest that the relationship was causal. Simply put, I argued that marriage causes men to change in the ways I observed. The reason I made such assertions, it is important to note, is because I had clearly satisfied two of the three logical requirements for establishing a causal connection. I had clearly established a correlation between marriage and several other phenomena. And I had clearly established a temporal order in which the change in marriage routinely came before the change in the other phenomena. The third requirement for establishing cause (no other factor responsible for the presumed cause and the presumed effect) was handled with multivariate statistical techniques. These are an approximation of an experiment[8], and cannot completely eliminate the problem. As a result, the evidence I presented in *Marriage in Men's Lives* can never be asserted to be proof of causation. It is, however, as close as we can get without conducting an experiment.

IV. Translating Concepts Into Measures

a. Introduction

49.   Before gathering a single datum from a sample, one must first translate the concepts of interest into indicators that can be measured. This is a central part of the entire process of designing the data-gathering procedure. Sometimes, the project calls for a questionnaire survey. Typically, in such cases, the concepts to be investigated are translated into specific questions on a questionnaire. In other cases, the research project calls for direct observations of individuals. When this is the method to be used, concepts are typically translated into observable behaviors that can be counted, coded, or otherwise recorded.

---

[8] An experiment is the intentional manipulation of a group of subjects. No naturally occurring phenomenon can be considered to be an experiment. (See Appendix II for further explanation.)

50.   For example, suppose a researcher is interested in the concept of generosity. Before it will be possible to investigate this concept, the researcher must arrive at some way to measure generosity that, in fact, can be measured. Strictly speaking, the concepts that are most often studied by social and behavioral scientists are not immediately apprehended. That is, there is no way to apply the five empirical senses (hear see, touch, taste, or feel) to determine their existence. One cannot see, touch, taste, hear, or feel generosity. Rather, generosity is an abstract concept that must be translated into indicators that may be discerned empirically. For example, the researcher might decide that any gift of money without direct compensation is an act of generosity. Now it becomes possible to empirically measure generosity. The researcher might ask individuals about their gifts of money in the past month, and whether there was any direct compensation. If the researcher is willing to believe the answers given to such questions, then he or she is able to measure such things as how many times an individual gave money, and how much money he/she gave. In this fashion, the researcher might make claims about the measured generosity of individuals, noting clearly how that term was defined. Regardless of whether others accept this definition of generosity as valid, the researcher has conformed to accepted scientific practice by clearly and specifically defining his concept. The simplest way to determine whether a concept has been defined is to ask if another researcher could replicate the study using the same empirical measures.

51.   Scientific evidence accumulates and gains credibility only through replication. The precise definition of all concepts to be used is crucial to the capability to replicate studies.

b. Operational Definitions

52.   In social science literature, the process of translating a concept into one or more empirical indicators is known as developing an operational definition of a concept. An operational definition of a concept is comparable to a recipe for a favorite dish. If one follows the recipe exactly without deviating from it, one will reproduce the desired

outcome. The dish can be replicated because there is a recipe for it. In social science research, the concepts used frequently come to have conventional operational definitions. Researchers using accepted operational definitions are able to replicate others' research, and build upon it. In this fashion, social science advances, as any science might.

53.   A good example is the (seemingly) simple concept of education. By convention, most social scientists accept "years of schooling completed" (or the highest degree earned) as an operational definition of education. Two people who differ in the number of years of completed schooling do not necessarily have different amounts of education in a more fundamental sense (there are, that is, obvious exceptions to the relationship). But the two people are considered to have completed differing amounts of schooling. The presumed relationship between the concept (education) and the indicator of it (years of schooling) is referred to as the validity of a measure. A valid measure is one that clearly measures the concept of interest. Most social scientists are willing to accept "years of completed schooling" as a valid indicator of the concept "education."

c. Valid Indicators

54.   The first requirement for a valid indicator is an operational definition. Technically, it is never possible to prove that an indicator is valid because no abstract concept can ever be measured. However, with repeated usage, and with repeated critiques of empirical indicators, social scientists have agreed on several strategies to gauge the presumptive validity of an indicator. For example, does one's measure of the concept correlate with the factors one would expect it to ('predictive validity')? In the case of education, we would presume that any valid measure of it would correlate with the prestige of one's occupation (i.e., we presume that people with more prestigious occupations also have more education). So the researcher would determine whether "years of completed schooling" correlates with established measures of occupational prestige. In fact, these two factors correlate positively, providing minimal assurance that the operational definition is valid. Researchers typically ask experts in their field to review their measures to check the presumptive validity ('face validity').

55.   With regard to the question at hand, we would need operational definitions of "gay", "lesbian", "bisexual, "parent", "child", "child's health", and "child's well being." Some of these present little problem (e.g., Statistics Canada has definitions of "parent" and "child", while psychologists have developed several measures of emotional and psychological health.) The operational definitions of "gay," "bisexual," and "lesbian" would be the most challenging concepts to measure, although several strategies have already been noted.

d. Reliability

56.   Once an operational definition exists, a researcher is able to establish the degree to which the measure has another desirable property, that of reliability.   A reliable measure is one that consistently reports the same value for the same magnitude of some phenomenon. An unreliable measure is one that fluctuates unpredictably in the values it produces. For example, we might ask if a particular IQ test is a reliable indicator of mental ability. To answer that question, we would need to know whether the same test, applied repeatedly to the same individual, would yield the same IQ score. If it did, then the test is reliable.

57.   A common threat to the reliability of any measurement is the use of a single observer to record the measurement. For instance, if a single researcher conducts repeated interviews, recording the warmth of parent-child relationships, for example, there is no way to estimate the observer's subjectivity. If several observers conduct the same types of interviews, however, it should be possible to make some estimates of this possibility (i.e., inter-rater reliability).

58.   Reliability is assessed in several ways. Sometimes a researcher will ask the same question, or use the same measurement strategy more than once (in surveys and various tests, slightly different wordings of the same question are typically included to tap this type of reliability). A similar strategy relies on the use of multiple measures of

21

the same concept. If a researcher is attempting to measure a subtle concept such as generosity, she might include 10 measures of it on a questionnaire. Any five such measures should classify a respondent the same way (i.e., as generous or not) as any other grouping of five measures. But the best and simplest strategy is to rely on established measures. To the extent possible, researchers rely on measures that have been used before, and for which there is general consensus among social scientists about reliability.

59.   A good indicator is one that is both valid *and* reliable. Unfortunately, reliability is not necessarily a guarantee of validity. My bathroom scales are very reliable. Every morning last week they weighed me at 76.8 kg. But when I went to my physician yesterday for a routine check-up, her more accurate scales weighed me at 78.2 kg. in exactly the same clothing. Clearly, while reliable, my bathroom scales are probably not valid (assuming that my physician's scales are). Rather, my bathroom scales are biased.

e. Bias

60.   Bias is a consistent error of measurement. A biased measure will consistently err in exaggerating or minimizing the magnitude of the issue being considered. Bias is introduced into a study in many ways. Sometimes the question asked is the problem. For example, if we simply ask people to report their age, we often find (in large surveys) that there are disproportionately large numbers of people who report being 20, 30, 40, 50, etc. years old, suggesting that people round their reports of their age to the nearest decade in many cases. The question in this case introduces a bias toward decades of age. It is for this reason that most survey researchers ask people to report their date of birth rather than their age. There does not appear to be bias in the former.

61.   Sometimes bias or unreliability is a result of the method used to obtain information. Many people are reluctant to divulge sensitive information. If we ask questions about topics such as masturbation, cheating, adultery, or lying, we know that many people will "under-report" the true incidence. It is for this reason that researchers

invest great effort to design their questions and methods to minimize such biasing tendencies.

62.	In questions about sexual behavior, or other very personally sensitive topics, researchers have found that telephone interviewing (where there is no face-to-face contact), the use of self-administered anonymous questionnaires, the use of computer-assisted-personal-interviewing (CAPI)(where the subject completes a series of questions on a lap-top computer with headphones) or very direct and blunt questions work best. Clearly, we should anticipate some problems with any question about a person's sexual orientation. Such questions, used either in screening, or in the actual study, would need to be carefully designed and tested. Studies do exist, that have investigated sexual orientation, while overcoming such problems for both adults and children (e.g., Laumann, et. al., and the Adolescent Health Panel Study).

63.	How do researchers know if their methods or questions are likely to be a source of bias? They pre-test questions and methods. Before conducting the actual project, a sound researcher conducts a small test of the procedures. The purpose of this pre-test is to ascertain whether the questions to be asked, or the methods to be used work as the researcher intends. A small (typically 5 to 15) group of individuals drawn from the population of interest is asked to complete the study. The researcher then interviews the participants (individually, or in a group) about the procedures used, and the methods for gathering information. He or she will ask about each question on a questionnaire. Did this question make sense? What did it mean to you? How did you understand the intent of this question? Did you know how to answer this question? What about the length of the task? Did it take too long? Were you tired or bored? Do you have any concerns about this study? Do you understand the purpose of it?

64.	Typically, the result of a pre-test is a minor revision of the data-gathering strategy. Some words are found to be confusing. Some questions are found to be threatening. Some projects are found to be too long, or too demanding. The researcher attempts to correct such problems before launching the full project. A pre-test is no

guarantee that the researcher has solved all the problems of potential bias associated with the instrumentation. But generally, it is regarded as necessary.

f. Assembling The Appropriate Comparison Group.

65.   There is still one critical design issue to be answered before gathering the data for a project. Recall that if we are attempting to answer the question "Are the children of gay and lesbian parents as healthy and well adjusted as those of their heterosexual counterparts?" we must be able to rule out any third factors that could conceivably mask or cloud the issue. How might this be done?

66.   What researchers have tried to do, in the studies reviewed, is determine what effect, if any, there is of having homosexual parents. To do this in a sound methodological manner they must somehow be able to compare children who differ in their circumstances on only this one dimension.

67.   Imagine, for example, that we were to compare the children of highly educated and wealthy homosexuals to the children of heterosexual parents in poverty. Imagine further that we compared the two groups of children in terms of their involvement with the juvenile justice system. Without doing this study, we can anticipate what such a project would reveal. Since we know from other research that children living in poverty are more likely to be involved in delinquent acts, the comparison between children with homosexual and heterosexual parents would undoubtedly show that the children of homosexual parents have significantly lower rates of delinquency. So the question is whether such a difference reflects the consequence of having homosexual parents, or of poverty?

68.   To make a convincing case about the consequences of having homosexual parents, a researcher would need to compare children living with homosexual and heterosexual parents but who did not differ on any other important dimension. A failure to compare children identical (or almost identical) on all important other dimensions

except the sexual orientation of their parents would be sufficient to invalidate the study. The only way possible to make two groups identical except for one factor is to use the process of a the classic experiment which is detailed in the paragraphs in Appendix II to this affidavit.

69.   The problem for most social scientists is that experimentation is neither feasible nor ethical. Quite simply, there is no feasible or ethical way to randomly assign children to living with either heterosexual or homosexual parents. And since we cannot do this, we must resort to various approximations to an experimental design. Every approximation shares the same objectives. All seek to make it possible to compare individuals on only the issue being studied; all seek to remove other factors from the study in one fashion or another.

g. Statistical Control

70.   On the matter of comparison groups, there is simply no option. A researcher must either resort to random assignment of cases, or statistical control. The latter refers to a class of statistical techniques that mathematically remove the effect of various confounding factors.

71.   For example, suppose we wished to compare a group of homosexual and heterosexual parents obtained in a probability sample of all Canadians for the purpose of investigating whether the children of one group or the other are more likely to skip school. Suppose further, that the homosexual parents were found to have much higher average incomes than the heterosexual parents. (That is, some fraction of the homosexuals has extremely high incomes, and few have very low incomes, while the reverse is true for the heterosexuals.) The researcher is interested in the effect of sexual orientation, and not the effect of income on children's truancy. Even if homosexuals do have higher average incomes than heterosexuals, the researcher will still want to know the effect of sexual orientation because many homosexuals will have incomes

comparable to many heterosexuals. How, then, does the researcher isolate the factor of interest – sexual orientation?

72.   To simplify the strategy, one can imagine that it would be possible to determine whether parental income affects truancy. Let us assume that it is found that every $1,000 less in family income is associated with a 1% increase in truancy (i.e., children from families earning $45,000 have 5% more truancy than do children from families earning $50,000.)

73.   Finally, assume that the average difference in "family" income between the homosexual and heterosexual parents is $10,000. Since every $1,000 difference in income is associated with a 1% difference in truancy, we would expect the children from the two groups of parents to differ by 10% simply due to their respective family incomes. Before we compared the two groups of children on the issue of their parents' sexual orientation, we would "adjust" for the income difference. If family income were the only difference between the two groups (except for sexual orientation), then the two groups of children must differ by more than 10% before we can begin to consider the possibility that homosexuality produces any effect on children's. truancy. Alternatively, should we find that the children of homosexuals do not differ at all from the children of heterosexuals in their truancy rates, we would probably conclude that homosexual's children actually have higher truancy rates than those of heterosexual parents. This is because we would expect an income effect absent any consequence of homosexuality. Failure to find significantly lower rates of truancy among the children of (more affluent, on average) homosexual parents, therefore, is actually evidence of a difference attributable to the sexual orientation of the parents.

74.   The example above simply illustrates that if samples that are not equivalent on all factors except one, (here, homosexuality of the parents) then finding no difference between children cannot render a scientific conclusion that the sexual orientation of parents has no consequences for children. (Indeed, such a finding may be evidence that parents' sexual orientation has enormous consequences for children.) The

important point is that the relevant question that must be asked is whether the researcher statistically controlled for all reasonable factors that might influence children other than the parent's sexual orientation. In my opinion, failure to do this invalidates any study of the consequences of a parent's homosexuality. In scientific research, a lack of correlation between two factors is sometimes the result of a failure to control for other relevant factors. This is the problem of a spurious non-correlation (a topic to be discussed later).

V. Gathering the Data.

a. Introduction

75.		A researcher with a clearly defined question (which we have in this case), who has a definable population, has developed a sampling strategy that is both feasible and scientifically defensible, who has translated all concepts into valid and reliable indicators, and who has pre-tested all instrumentation is ready to gather data.

b. Gathering Methods and Guidelines

76.		The choice of data-gathering methods will depend on many factors, including the resources available to the researcher, the topic, and the purpose of the research. Regardless of the method(s) used, however, there are several basic guidelines. First, to the extent possible, the researcher should do everything possible to minimize his or her role as a stimulus. That is, subjects should respond to the instrument rather than to the researcher. In face-to-face interviewing, for example, the researcher should be a neutral presence to the extent possible. This may require the use of different interviewers for different subjects. Dress and demeanor (including dialect or other speech patterns) are sometimes thought to influence the type of answers subjects provide. Race, similarly, may be an issue for certain topics. Again, to the extent possible, the researcher should be sufficiently familiar with the subjects and with the interview instrument to minimize his or her role in the data-collection.

77.   The presumption in social science research is that data gathering involving human subjects should be regarded as a stimulus-response situation. The desired objective is that every subject will respond to the same stimuli. Indeed, this is one of the strengths and weaknesses of self-administered survey questionnaires. Each questionnaire is identical, and the researcher is not present when it is completed. At the same time, the researcher cannot assure that the conditions under which the questionnaire was completed were identical for all subjects. Some may have discussed their answers with others. Some may have been watching TV while completing the questionnaire, and so on. Face-to-face survey interviews, on the other hand, offer the researcher the opportunity to explain issues, to observe the circumstances under which the instrument is completed, and to take notes on issues that might be relevant in the analysis of the results (e.g., the subject appeared to have been under the influence of alcohol).

78.   Another general guideline is that the researcher should use multiple methods of gathering data, if at all possible. If a project relies on both self-administered surveys and face-to-face interviews, the researcher gains the ability to compare the results of the different methods. Every method has its known weaknesses. Should two methods produce similar results, the researcher has greater confidence in her results because there has been a replication of sorts.

c. Response Rate

79.   Finally, regardless of the method used, the researcher must attend to the very important issue of response rates. Once a probability sample has been drawn, the researcher's goal is to obtain complete information from *every* member of it. To the extent that this is not done, unknown biases are introduced into the study. Consider the typical political poll done before most national elections. These rely on telephone interviews with individuals in a sample of all telephone numbers. Researchers generate random digits as part of the telephone number to insure that unlisted and listed numbers have equal probabilities of selection. Once a desired sample (typically between 800 and 1,200) is drawn, the task is to contact each of these numbers and interview a respondent

(chosen according to various strategies to randomly select one member of multiple-person households). In drawing this sample, telephone interviewers must contend with many problems relating to service (is the phone a residential line?), and eligibility (does the resident qualify for the study?). But many people cannot be easily reached by telephone. The use of answering machines, and various screening technologies (e.g., "caller ID") alert the subject to the origin of the incoming call. Many people simply will not answer calls from unknown sources. Others are unwilling to talk to someone who identifies him or herself as an interviewer, and so on.

80.   Telephone interviewers, therefore, face tremendous problems in completing interviews with all members of their original sample. Possibly, there is no great consequence. But possibly, there is enormous consequence. Which of these possibilities is more likely depends on whether the subjects who could not be interviewed resemble those who were in important respects. For example, if wealthier subjects are less likely to be interviewed, then the results of the study no longer generalize to the population from which the sample was drawn.

81.   Generally speaking, the issue of response rate pertains to self-selection. Once a random probability sample is drawn, inevitably, some members will not be contacted. To the extent that they do not differ in important ways from those who are contacted, then the scientific integrity of the sample is probably not compromised significantly. But this is not something that is easily determined. Since those who are not contacted are typically unknown, the researcher is often unable to estimate the magnitude of the self-selection bias. In sum, when some sampled subjects agree, while others disagree to participate in a study, this self-selection creates a potential source of bias in the result.

82.   If a researcher does not use a probability sampling method, but instead allows subjects to volunteer for any reason they wish (e.g., placing an ad in a newspaper to recruit subjects), then every single member of the study is self-selected. Unless the researcher can know the difference between those who do and do not volunteer, or make

some reasonable assumptions about such differences, the study cannot be treated as scientific evidence.

83.　　　　In practice, researchers almost never contact every member of the original probability sample. The fraction that is contacted and completes the instrument defines the response rate. How high should the response rate be to allow conclusions to be drawn from the results? Conventional standards in social science now regard a response rate of 80% to 95% as excellent, of 70% to 80% as very good, and of 60% to 70% as acceptable. Response rates below 60%, however, are reason to believe that the actual sample obtained differs in unknown ways from the sample initially drawn. Obtaining high response rates, in short, is crucial. It is for this reason that survey research often involves repeated attempts to contact members of the original sample (repeated telephone calls, or repeated visits to a residence, often as many as 8 or 10 times before dropping a case).

84.　　　　Once the data are obtained, the researcher is obliged to check them to verify that there are no significant and obvious errors. This is a small but important step before the analysis begins.

VI. Analyzing The Results.

a. The Research Hypothesis

85.　　　　The researcher is now ready to conduct the actual analysis of the data. Any questions about a correlation or a cause-effect relationship are stated in the form of hypotheses that are tested with statistical techniques. Generally speaking there are two types of hypotheses central to any research project of this sort. An hypothesis is defined as an assumption about the population represented by the probability sample of it.

86.　　　　The Research Hypothesis is what the researcher expects and hopes to find. The Research Hypothesis consists of the assumptions about a population that we are

willing to make and believe in. Were we testing a new vaccine against measles, our Research Hypothesis would be that the vaccine does, in fact, reduce the incidence of measles. This hypothesis is not intended to be exposed to a test with statistics. The remaining hypothesis/hypotheses to be tested (the testable hypothesis) is typically referred to as the Null Hypothesis.

b. The Null Hypothesis

87.   The Null Hypothesis is what the researcher actually tests. Usually, the Null Hypothesis consists of a statement that a certain population value (e.g., the percent of voters who will vote for candidate X) is equal to some given value. Statistically, this hypothesis is called the null hypothesis since it implies that there is no difference between the actual (true) value in the population, and that which is being hypothesized.

88.   Consider the statement that "homosexual and heterosexual parents spend an equal amount of time helping their children with homework." This can be understood as a testable hypothesis stating that the population averages of the two groups are equal. The researcher who has drawn a random probability sample of homosexual and heterosexual parents would compare the average time spent helping children with homework by the two groups. The Research (or Alternative) Hypothesis in this case would be that the two averages are not the same.

89.   Note that this Research Hypothesis actually includes several possibilities:

1. Mean for homosexuals > Mean for heterosexuals,
2. Mean for homosexual < Mean for heterosexuals, and more generally,
3. Mean for homosexuals ≠ Mean for heterosexuals

90.   Since there are several possible Research Hypotheses, the researcher must specify, in advance, which possibility is the more likely result of a rejection of the Null Hypothesis. When there is no specific prediction, a hypothesis such as # 3 (above) is

advanced. When the researcher has a-priori reason to expect one group to have higher (or lower) scores than the other, then hypotheses such as # 1 or # 2 are specified. The implications of such decisions pertain to the strength of evidence needed to reject the Null Hypothesis. It takes more evidence to reject the Null Hypothesis in favor of hypothesis # 3 than either # 1or # 2.

91.   Consider the problem facing a researcher who tests a new drug. The clear presumption is that this new drug will do better (produce more cures) than existing drugs or therapies (i.e., Research Hypothesis: New Drug > Old drug). The null hypothesis in this case would be that the new drug does no better (on some measure) than the old strategy. This is the hypothesis that is tested statistically. If the researcher is able to reject this hypothesis (by finding sample evidence in favor of better results from the new drug), then he will conclude that the new drug probably does, in fact, do a better job.

92.   This scientific practice resembles the case of an accused criminal in a court of law. The defendant is considered not guilty unless the evidence suggests beyond a reasonable doubt that he is guilty, so long as the trial was conducted fairly. A null hypothesis is considered tenable unless the evidence suggests otherwise, (beyond some reasonable doubt), so long as the test was conducted fairly. What is important to understand is that a failure to reject the Null Hypothesis, however, does not establish the absence of differences between two groups. Rather, it indicates insufficient evidence to render a verdict.

93.   Just as a court pronounces a sentence of guilty or not guilty (rather than innocent), so a statistical test of the null hypothesis leads to a verdict of reject, or fail to reject (not accept).

c. Threshold Value

94.   Setting up the Research and Null Hypotheses is the first step in dealing with a problem of hypothesis testing. The next step consists of devising a standard by

which a researcher will decide whether the Null Hypothesis is, or is not, to be rejected. Establishing a threshold value to distinguish the two possibilities does this. The researcher will calculate a statistic (e.g., an average) that may theoretically assume a wide range of values. Depending on the value that is obtained, the statistic either falls beyond the threshold for rejecting the Null Hypothesis, or doesn't. If it does, the researcher rejects the Null Hypothesis. If it does not, the researcher fails to reject the Null Hypothesis (note, the researcher never accepts the Null Hypothesis).

95.   To establish the threshold, the researcher relies on statistical theory. Based on a probability sample of homosexual and heterosexual parents, the difference in averages between the two may take an infinite number of values. But if the null hypothesis is true, then certain values are more likely than others. Simply put, if the Null Hypothesis is, in fact, true, then the difference of averages is more likely to equal zero than it is to equal any other value. But other values are possible, even if the true difference in the population represented by this one sample is zero. Due to the vagaries of random sampling, it is conceivable that the sample difference in averages would actually be some positive or negative value even if the true population difference is zero. But it would be unlikely to be vastly different than zero if the Null Hypothesis is true.

96.   Statisticians determine how unlikely it would be to find a particular result in a sample if the Null Hypothesis is true. This is how the boundary between rejecting and failing to reject the Null Hypothesis is established. If the Null Hypothesis is true, sample statistics are extremely unlikely to fall beyond the boundary and lead to rejecting the Null Hypothesis. By convention, this boundary is established so that the risk of incorrectly rejecting the Null Hypothesis when it is true is less than 5%. In sum, the Null Hypothesis is rejected when the sample evidence is convincing beyond a reasonable doubt of something less than 5% that it is true.

d. Error Types

97.   The important point is that the researcher who does, in fact, reject the Null Hypothesis is always doing so at some risk of error. If the boundary is established at 5%, then the probability of rejecting the Null Hypothesis when it is actually true, and should not have been rejected, is .05 (5%). Returning to the example of an accused criminal, making this type of error is comparable to convicting an innocent person. In research, this type of error is known as Type I error.

98.   In almost all articles reviewed for this case, the presumed Research Hypothesis is that the two groups do not differ. It is important to note that this is a very different type of test than is typically conducted, where the Null Hypothesis, that is, the two groups do not differ, is tested. Whenever the Research Hypothesis and Null Hypothesis are, essentially, switched as in this case, attention shifts from a Type I error to another type of error.

99.   There are actually two types of possible error involved in any testing of research and null hypotheses. Suppose that the statistical evidence from the sample does not fall beyond the boundary established. In this case, the researcher does not reject the Null Hypothesis. Still, we cannot rule out the possibility that the Null Hypothesis is, in fact, false. And there will always be a certain possibility of making this type of error. Were this a criminal trial, such an error would be comparable to finding a guilty person not guilty. In research, this type of error is known as Type II error.
A researcher is able to manipulate the chances of Type I error by the selection of the boundary point. It would be possible, for example, to minimize the chances of making a Type I error (the statistical significance of a test) by establishing the boundary at a point defined by a probability of, say, .001 rather than .05. Where the boundary is set depends on the seriousness of the consequences of making an error. Were we testing a critical medical product, we would probably set a .001 probability because the consequences of falsely rejecting the Null Hypothesis could be enormously important, such as putting patients on a treatment regimen that is not superior to existing protocols. But the important point about the two types of error is that by decreasing the probability of one

type of error, we increase the probability of the other type of error. The researcher who establishes a very demanding critical boundary (level of statistical significance) by setting a very low probability of Type I error thereby increases her chances of making a Type II error.

e. The Power of a Test

100.   The probability of committing a Type I error is known as the level of significance. The probability of committing a Type II error is related to the "power" of a test. In the language of statistics, the lower the probability of not rejecting the Null Hypothesis when it is false, the more powerful is the test. A powerful test, that is, is less likely to err by failing to reject the hypothesis that the two groups do not differ when, in fact, they do.

101.   The power of a statistical test may be compared to the power of a microscope. It reflects the ability of a statistical test to detect from evidence that the true situation differs from a hypothetical one. Just as a high-powered microscope lets us distinguish gaps in an apparently solid material that we would miss with low power or the naked eye, so does a high power test of the Null Hypothesis almost insure us of detecting when it is false. Further, just as any microscope will reveal gaps with more clarity the larger are those gaps, the larger the departure of the Null Hypothesis from the true situation specified by the Alternative Hypothesis, the more powerful is the test of the Null Hypothesis. In the case at hand, the larger the "effect" or the larger the difference between homosexual and heterosexual parents, the more powerful the test will be. If the actual difference is small, the test will be less powerful

102.   The power of a statistical test is defined as
     [1.0 – (probability of a Type II error)]

103.   Type I and Type II errors differ in their implications. In the present case, a failure to reject the Null Hypothesis when it is false (Type II error) would lead to the

35

erroneous conclusion that the children of homosexual and heterosexual parents are similar when, in fact, they are different. A faulty rejection of the Null Hypothesis when it is true (Type I error), however, would lead to the incorrect conclusion that children of the two types of parents differ when, in fact, they do not. Given that policy might be formulated on the basis of the results in this particular case, it is clearly more important to minimize the chances of Type II errors than Type I errors when the Research Hypothesis, rather than the Null Hypothesis, is that the two groups do not differ.

104.	For this reason, the researcher investigating the children of homosexual and heterosexual parents should accept a higher chance of Type I errors than is typically done in social science research. This will lower the chances of a Type II error. Rather than establish the boundary for rejecting the Null Hypothesis by setting .05 as the critical value, in my opinion, it would make more sense to set the level of significance for rejecting the Null Hypothesis at a higher value, perhaps .10.

105.	Another way to increase the power of the statistical comparison is to increase the size of the sample. Small samples have lower power than large samples. Given the nature of the problem, that is where the Null Hypothesis of "no difference" is actually the Research Hypothesis, research on this topic requires a large sample, especially to reliably detect small differences between groups.

106.	If we design our study in such a way to be powerful enough to detect rather small differences between the averages of two groups, we will need a sample of at least 400 cases to achieve Power of .80. Since Power = [1.0 - probability (Type II error)], our test runs the risk of Type II error of .20. This would mean that the researcher runs a 20% risk of failing to reject the Null Hypothesis when it is, in fact, not true.

107.	In sum, given the nature of the problem being considered, in my opinion, reliable research would require an increase in the level of statistical significance required to reject the Null Hypothesis from the conventional .05 to .10. Sound research would also require an increase in the sample to at least 400 cases. And even then, the power of

the test may be inadequate if there are other factors that must be controlled (e.g., income, age, education, etc.). This is why I suggest a sample of 800 gay parents (see my earlier comments on sampling).

108.   This brings me to the last point about the analysis. The skilled researcher must do everything possible to control *all* factors that might cloud the findings. The research must statistically control for all important differences between heterosexual and homosexual parents other than their sexual orientation. To do this would require the size sample just mentioned. Preliminary research might identify ten or fifteen possible factors that would need to be statistically controlled before a valid comparison of children in the two groups could be conducted.

109.   What other factors must be statistically controlled? The response is *any* factor that is correlated with both the cause and the effect. In the case at hand, this would mean that anything that is related (on average) to being in a same-sex union and is also related (on average) to the health or well being of children must be controlled. Possible candidates for such factors include parents' income, parents' education, parents' emotional and psychological health (e.g., depression), relationship quality (between adult partners), and various residential variables (e.g., neighborhood quality, etc.). Also important would be the relationship history that the child has experienced (how many changes in his/her parent's partners) or whether the children have lived in a heterosexual relationship for varying portions of their lives?

110.   An alternative to statistical control is achieved by matching cases. If every homosexual parent could be "matched " by a heterosexual parent on all relevant factors, this would allow the researcher to compare the two groups. Since no study, to date, has been able to do this, statistical control appears to have been the only feasible strategy that would permit a researcher to compare homosexual and heterosexual parents.

111.   Before moving to a specific evaluation of the evidence offered in Professor Bigner's brief, I want to conclude this section by noting that statistical control

is particularly important in this case. It is possible for two factors to appear to be uncorrelated due to their relationship to some third factor. If this third factor is positively correlated with sexual orientation, but negatively correlated with children's well being (or vice-versa), then a failure to control it may lead to a spurious non-correlation. In short, it is essential to understand that statistical control is as necessary in the presence of a trivial or zero correlation as it is in the presence of a strong and substantively large correlation.

VII. Examination of Prof. Bigner's Affidavit.

a. Introduction

112.   In this section, I set out my conclusions and analysis of my review of all evidence cited by Professor Bigner in his affidavit sworn November 15, 2000. I evaluate only published articles in professional outlets. I omit from my review all unpublished Ph.D. doctoral dissertations and materials that appeared in popular news outlets (e.g., Newsweek magazine). My review focuses solely on the scientific merit of the research. The evaluation that follows concentrates on those issues that I have discussed in the first half of my affidavit, above.

113.   Specifically, I evaluate
   - The scientific adequacy of the sample. Did the article rely on a probability sample of adequate size? Was there evidence of obvious sample bias?
   - The operationalization of key concepts
   - The adequacy of the comparison group, and
   - The appropriate use of inferential (generalizing) statistics.

114.   Professor Bigner's affidavit relies almost entirely on the Vermont brief included as Exhibit "B" to his affidavit. First I examine Professor Bigner's primary assertions, both in his affidavit of November 15, 2000 and in the Vermont brief (seriatim). I then review the evidence for those assertions found in the articles cited.

b. Opinion on Evidence Relied on by Professor Bigner

115.　　　All of the articles I reviewed contained at least one fatal flaw of design or execution. Not a single one was conducted according to generally accepted standards of scientific research.

116.　　　The studies reviewed exhibit the critical defects explained earlier, in the following ways:

- Not one study relied on probability samples of homosexuals and heterosexuals.
- The definition of "homosexual" was typically vague and poorly articulated, often no more than "self designated" or "self identified." There is no way, therefore, to know whether homosexuals who do not openly identify differ from those who do. Nor is there any way to know what "self identified" means with respect to the question at hand.
- In most cases, all data were collected by a single researcher. This makes it impossible to assess the extent of subjective bias that may have been introduced.
- Only one study relied on a longitudinal design.
- Researchers often relied on well-known and established measures, but rarely reported their reliability for the samples studied.
- The potential sources of serious bias are very clear and often acknowledged by the authors:
  - ➢ First, is the reliance on self-selected samples. When subjects are allowed to select themselves into a study without any scientific sampling used, the researcher cannot know how his or her subjects compare to those who did not select themselves into the study. This unknown bias makes it impossible to generalize the findings from any such study.
  - ➢ Second, is the fact that almost all samples of homosexuals have extremely high levels of education. In all studies reviewed (where such information

was noted), well over half of the homosexuals studied had completed college (only 23% of all adults in America have completed college)[9]

> Lastly, the researchers failed to incorporate statistical controls to deal with extraneous influences, even when their research revealed notable differences between their samples of homosexual and heterosexual subjects on such dimensions.

- Response rates, where noted, were typically low.
- Sample sizes were almost always too small to provide the statistical power needed to confidently fail to reject the hypothesis of 'no differences' between groups.

117.   This last point should be stressed. The researchers typically found "no differences" between their homosexual and heterosexual subjects. The tests that were conducted (even though inappropriate) relied on samples too small to allow the researcher to make this conclusion without risking a very high probability of error. 118.

        Stated most simply, the articles cited in Professor Bigner's affidavit did not rely on samples of sufficient size to provide the statistical power needed to reach the conclusions they did.

119.   My conclusion, based on the analysis that follows, is that we simply do not yet know how the children of homosexual and heterosexual parents compare at this point in time. To know this, we would need to conduct the type of project I outlined in the first half of my comments. Such a study is not a particularly large undertaking. There are many examples of social science projects that are more complex and challenging than this one would be. However, based on the studies reviewed and my own search of the literature, this research has not yet been done. Given the potential consequences of an incorrect conclusion, such research seems warranted before any body, legislatures or courts, come to any conclusion about domestic arrangements with unknown consequences for children.

---

[9] http://www.census.gov/prod/2/pop/p20/p20-489.pdf

120.		The final portion of Professor Bigner's affidavit is aimed at supporting a hypothetical argument about the benefits of legal marriage for children of same-sex couples. I am familiar with this literature and stipulate that, with few exceptions, [10] it conforms to the standards of acceptable scientific research that I established at the outset of my comments.

121.		I believe it is true, as Professor Bigner claims in his paragraph 14, that, at least with respect to heterosexual couples:

1) Children benefit from living in a healthy, loving home with both parents in the context of a healthy, happy intact family;

2) civil marriage, and the protections, supports, and obligations that accompany that status, can fortify committed relationships between parents;

3) the community and social supports that accompany civil marriage, including enhancing the strength of relationship between spouses, can promote even better parenting.

122.		The problem, in my opinion, is that there is an important, yet unanswered, question about the benefits of marriage. While it is generally true that marriage confers numerous advantages, it is unknown whether those advantages are the result of marriage, per se, or heterosexual marriage. To assume, as Professor Bigner does, that marriage has the same consequences *regardless* of the sexual orientation of the parents is pure speculation. We simply have no basis, at this point, on which to make an assumption that legal recognition of the relationships such as same-sex marriages, would eliminate the social prejudice or stigma associated with homosexuality.

123.		Professor Bigner concludes, at paragraph 15, that the evidence reviewed establishes the claim: "where children of gay and lesbian parents may have difficulties, those difficulties stem from the lack of social and legal support for their family structures rather than any intrinsic shortcoming of the family structure itself. To the extent that

---

[10] Blumstein and Schwartz, 1983; Grissett and Furr,1994; Solomon and Rothblum, 1986; Crockenberg (1982).

41

some children may experience difficulties as a result of societal reactions to their lesbian mothers or gay fathers, those difficulties could only be alleviated by legal recognition of those family structures."

124.   My opinion, based on my own reading of the literature, is that, undoubtedly, teasing, ostracism, or other forms of social prejudice are a large part of the story of the lives of children living with gay or lesbian parents. But equally pertinent are any other factors inherent in the family relationships of same-sex partners, at least to the extent that the evidence is cited by Professor Bigner. Qualitative research referred to by Professor Bigner addresses this point clearly (Blumstein and Schwartz, 1983). Surely, the question that should be asked is whether same-sex partners have different rates of break-ups than opposite sex cohabiting (unmarried) parents.

125.   If, for example, gays and/or lesbian relationships exhibited higher rates of break-up than unmarried or married heterosexual relationships, this should be known and investigated, for this factor may have effects on children. The point, however, is that this aspect has not yet been addressed. More generally, to assert that the only difficulties faced by the children of gay and lesbian parents are the result of social forces (prejudice, etc.) and not any factors related to the particular family structure, presumes that we have tested this basic idea. In my view, the accumulated evidence does not speak to this issue. If, indeed, sound scientific research were to confirm the closing assertions made by Professor Bigner, I would be pleased to agree with his opinion. In my own professional opinion, however, such research remains to be conducted and the issues remain unresolved.

c. Analysis

126.   Before addressing the issue of how children of gay and lesbian parents compare with those of heterosexual parents, Professor Bigner offers several preliminary assertions that have no proper foundation in the scientific research he relies on. While these claims may very well be true, the issue is simply whether they are supported

scientifically by the studies Professor Bigner relies on to make those claims. In my analysis, I address these claims by examining each of the studies cited by Professor Bigner and describing the crucial weaknesses the studies display.

127.   The preliminary assertions made by Professor Bigner are:
   1) About one third of lesbians and about 10% of gay men are parents.
   2) Increasing numbers of lesbian and gay couples are rearing their own children.
   3) The reasons why gay men and lesbian women become parents are no different from those motivations that prompt heterosexual men and women to become parents.
   4) Gay and lesbian parents possess parenting skills and abilities comparable to their heterosexual counterparts

128.   With respect to the first and second assertions, there are two primary sources cited: Bell (1978) and Patterson (1992). The first of these studies did not attempt to estimate the prevalence of homosexuals and the second relied on the claims of others who make the assertion that it is cited for by Professor Bigner. All of the sources cited from the Vermont Brief on this issue either did not conduct the research to make the claim, or did not claim, that the number of gay and lesbian parents is increasing. My conclusion is that none of the sources cited by Professor Bigner contains evidence about the prevalence of homosexuality, or the change in prevalence. None of the studies makes any claims about such matters (except to quote others who make such claims without evidence). In short, there is absolutely no evidence about how many homosexual parents there are, nor whether their numbers are increasing or decreasing. I have reached this opinion based on my detailed examination of each of the studies, as described in Appendix III to this affidavit

129.   The third assertion, regarding the reasons gay and lesbian men and women decide to become parents, is held to be supported by a number of studies authored by Professor Bigner himself, in collaboration with others. The first two studies, Bigner and Jacobsen (1989b; 1989a) suffer from the inappropriate application of statistical

techniques, the failure to control for extraneous factors, poor sample size, and inadequate sample, among other flaws, making it impossible to draw general conclusions from this research. In the third study, Siegenthalor and Bigner, (2000), the authors claim their research found that the reasons heterosexuals and homosexuals become parents are, indeed, different, in direct contrast to the assertion for which this article is cited. In my opinion, none of the studies reported for this assertion is sound enough methodologically to permit the claim to be made. The details of my analysis of the studies referred to is contained in Appendix IV to this affidavit.

130.   The last assertion, that gay and lesbian parents have the same parenting skills as heterosexual parents, is another one we might like to assume. However, Professor Bigner's claim is that this assertion is scientifically supported by the studies cited for it. In my opinion, the collection of these sources cited about lesbian mothers is inadequate to permit any conclusions to be drawn. None had a probability sample. All used inappropriate statistics given the samples obtained. All had biased samples. Sample sizes were consistently small, and in almost all cases inadequate to permit the researchers to draw conclusions about their failure to reject the null hypothesis (even when not stated, the presumption in all these studies is that there are no significant differences between the groups). And despite the use of good measures in many cases, there was no way to ascertain how the researchers insured that their samples of "lesbians" satisfied any definition of that term, nor of whether the samples of heterosexuals were, in fact, heterosexuals. There is no way to generalize the results of these studies beyond the peculiar and unusual samples used in them. I do not believe this collection of articles indicates that lesbian and heterosexual mothers are similar.

131.   In respect of gay men, the last assertion exhibits the same frailties if, as Bigner claims, the studies cited are considered scientific support for the claim made. In sum, the evidence contained in the Vermont brief, in regards to the parenting behaviors of gay men, rests on three studies that are all based on non-probability samples of a size that is inadequate to provide the power needed to fairly test the hypotheses involved. Other problems noted for the individual studies in Appendix V, also render their conclusions

questionable. I do not believe these articles offer the support claimed for the assertion made about the parenting skills of gay men. In fact, from a scientific perspective, the evidence confirms nothing about the quality of gay parents.

d. Principal Assertions

132.　　　　Professor Bigner makes several principal assertions that form the core of his opinion. The first is that the children of gay and lesbian or same-sex parents are as well adjusted as those of their counterparts who have heterosexual or different sex parents. Further, Professor Bigner makes the claim that the evidence also indicates that there are no differences between the children of gay parents and the children of heterosexual parents in terms of gender identity or sexual orientation, based on the studies presented in the Vermont Brief.

133.　　　　Professor Bigner says that the first assertion is supported by approximately 50 published studies, including a meta-analysis of 18 studies previously published on the subject of the impact of homosexual and heterosexual parents on children (Allen and Burrell, 1996). Many of the articles included in the meta-analysis are ones that I reviewed for earlier portions of this affidavit.

134.　　　　Meta analysis is a statistical method used to combine comparable studies when each, by itself, has inadequate sample sizes to provide needed power. The meta-analysis is able to provide more power by combining the results of many smaller studies (thereby producing a larger sample). The process of selecting appropriate studies and coding their information is fraught with its own biases and pitfalls. When the original cases are properly evaluated for quality, and weighted accordingly, such an analysis is able to correct for small samples so long as the other requirements for inferential statistics were satisfied. In the present meta-analysis, the studies that were combined suffered from the flaws already noted. As such, combining many poorly done studies, each of which has peculiar non-probability samples and unknown biases, cannot and does not provide any greater evidence than the individual studies do, taken separately.

45

135.   In Appendix VI to this affidavit I have set out my comments resulting from my detailed analysis of each study cited in support of Professor Bigner's principal assertions. The conclusion can be summarized very succinctly: all of these studies exhibit flaws that make the conclusions drawn by Professor Bigner unsupportable. However, considering that Professor Bigner's main assertion is made from these studies, I thought it would be helpful to include in the body of this affidavit a detailed analysis of the study that I view as one of the most rigorous studies among all those reviewed: Golombok and Tasker (1996).

136.   My view that these authors conducted one of most rigorous studies is because they employed a longitudinal design. A non-probability sample of 27 self-selected lesbian mothers and their 39 children, and a control group of 27 self-selected heterosexual single mothers and their 39 children were first studied in 1976-1977 when the average age of the children was approximately 10 years. Subjects were recruited with advertisements in lesbian and single-parent publications and contacts with lesbian and single parent organizations. "Lesbian" was defined as a women who regarded herself as wholly or predominately lesbian in her sexual orientation. The definition of "heterosexual" was behavioral. Members of the control group had their most recent sexual relationship with a man. Importantly, all children in the study were conceived and born into heterosexual relationships.

137.   In 1992-1993 when the children were about 24 years old, they were seen again. Of the original 54 mothers, 51 were traced. This produced an effective pool of 37 of the children of lesbians. Of these, 25 were interviewed (68%). 21 of 39 children of heterosexual mothers (54%) were also interviewed. The two groups were compared and found to be similar in terms of education, age, gender, or ethnicity. The authors investigated the reasons for panel attrition (drop outs between waves). The only notable difference between groups in attrition was that lesbians in relationships high in conflict were less likely to remain in the panel.

138.		The instrumentation is described in detail. Reliability of measures, and inter-rater reliability of raters are reported. Although this study is strong, it still suffered from the weakness that no statistical controls were employed to compensate for extraneous factors.

139.		Findings indicated that at least one difference existed between the two groups of children, contrary to the assertion that the study is supposed to support. The children raised by lesbians were more likely to have experienced a same-sex sexual relationship than young adults raised by heterosexual mothers (though this appeared most true for sons rather than daughters.). This may or may not be a true difference due to the additional weaknesses identified in the sampling (i.e. non-probability and self-selection).

140.		In sum, all the articles offered by Professor Bigner, including the study considered the most rigorous, cannot be taken as establishing the claim that scientific research shows no differences between the children of gay parents and the children of heterosexual parents in terms of gender identity or sexual orientation.

141.		Professor Bigner is correct to state that the "weight of published evidence" suggests that this is so. From a sound methodological perspective, the results of these studies can be relied on for one purpose – to indicate that further research regarding his hypothesis is warranted. However, in my opinion, the only acceptable conclusion at this point is that the literature on this topic does not constitute a solid body of scientific evidence.

VIII. Appendices

Simple Random Sampling

A simple way to envision simple Random Sampling (SRS) is to imagine writing the names of each member of some population on a card. Suppose there are 500 individuals in the population and we want a sample of 50. There would be 500 cards, each with a name on it. If all 500 cards were placed in a large box and shuffled, we could draw the first card with assurance that it has no greater or lesser chance of being drawn than any other card in the box. The chance of drawing this one name is simply 1/500. Once we draw the first case, we write the name of the person on a sheet, and place the card back into the box. It is essential that the card be returned to the box. If we did not return the card to the box, then the next name drawn would have a 1/499 chance of selection because there would only be 499 cards remaining in the box. Since 1/499 does not equal 1/500, we would have violated the primary assumption of SRS. Following in this manner, we would continue drawing a card, writing the name down, returning the card to the box, and drawing another name until we had our desired 50 cases (returning any name that has already been drawn before). At this point, we would have a pure random sample. Any results based on these 50 cases could be generalized with reasonable assurance to the entire population of 500 using standard statistical techniques.

Researchers do not, of course, use a box of cards to assemble their random samples. Rather, computer software is used to select a random sample of cases, or generate a list of random numbers. Alternatively, samples may be selected by systematically drawing every Nth case from a list (e.g., taking every 10th case from a list of 1,000 to produce a systematic random sample of 100).

In practice, researchers are sometimes unable to assemble an accurate list of all members of the population. This is true, for example, when sampling all adults in the United

States, all children in public schools, or all patients with diagnosed breast cancer. In such cases, alternative strategies are used to approximate a random sample. One common strategy is to randomly sample geographic or organizational units. For example, a researcher might randomly sample 100 U.S. Census tracts. Then, within each randomly selected Census tract, the researcher might randomly select 5 Census blocks. Within each randomly selected Census block, the researcher might randomly select 2 households. Within each randomly selected household, the researcher would interview one randomly selected individual. In all, this strategy would produce 100 X 5 X 2 = 1,000 individuals randomly selected from a total population defined as all households in U.S. Census tracts (approximately 100% of all U.S. households). A sampling statistician would calculate appropriate weights to be applied at each stage of this multi-stage sampling strategy to produce a final sample of 1,000 cases that can be treated as a random sample. A comparable strategy could be used with hospitals, schools, churches, or clubs as the initial sampling units.

The Classic Experiment

In a classic experiment, the researcher assembles a representative sample of cases and randomly assigns them to one of two groups. The 'experimental' group and the 'control' group, that is, are determined purely by chance (e.g., flipping a coin). Since there is nothing but random chance to determine which group a case ends up in, there is no logical way for the two groups to differ. Random assignment will place as many rich as poor individuals in each group, as many white or Hispanic individuals into each group, and so on. The researcher administers a test at the outset of the study to verify that the experimental and control groups do not differ. Then the researcher administers some treatment or stimulus to the experimental group that is not administered to the control group. At this point, the two groups differ only with respect to the treatment or stimulus. Logically, the two groups do not differ on any other dimension. The researcher then administers the test again. Any difference that is now found between the two groups may logically be attributed to the treatment or stimulus because it is the only thing that distinguishes the groups. (In actual practice, there are well known problems with experiments that may threaten the similarity of groups on all matters except the treatment/stimulus. These threats are dealt with by more complex experimental designs than the one just outlined)

The classic experiment comes as close as one can come to satisfying all three conditions for establishing a cause-effect relationship. And the reason it does is because it relies on random assignment of cases into the various groups to be compared. Random assignment essentially assures the researcher than all "other factors" that might confound the results are distributed evenly – one group has as many or as few as the other.

Detailed analysis of Studies Respecting Claims About Prevalence of Homosexuality and Homosexual Parents

Bell (1978)

This study of homosexuals in San Francisco, CA. is elaborate and well conceived. However, the researchers did not attempt to estimate the prevalence of homosexuality in either San Francisco or the nation. Nor is there any attempt to measure change in the homosexual population over time. The research team recruited (through self-selection) a large sample of homosexuals by distributing recruitment cards in various locations and asking respondents to volunteer to be in the study (paid advertisements, gay bars, personal contacts, gay baths, homophile organizations, private bars, public restrooms, hotels, restaurants, etc.) A heterosexual sample was obtained by probability methods developed and applied by the National Opinion Research Center. Detailed and carefully executed statistical analyses were performed, but the failings regarding prevalence and change are significant.

Patterson (1992)

This study does not make the claim Bigner attributes to its author, nor does the author offer any original research on this issue. Rather, she refers to others' claims. According to Patterson "How many children of gay and/or lesbian parents live in the United States today? No accurate answer to this question is available. … According to large-scale survey studies, about 10% of gay, and about 20% of lesbians are parents" (1992: 1026, and footnote 1).

Evidence from the Vermont Brief:

Patterson (1994).

The researcher studied 27 lesbian couples, 7 single mothers, and 4 separated lesbian mothers. She made no claims, nor conducted any research in support of the assertion that the number of gay or lesbian parents is increasing.

Pies (1990).

The author neither conducted, nor claimed to have conducted any research in support of the assertion that the number of gay or lesbian parents is increasing.

Rafkin (1990).

The author neither conducted, nor claimed to have conducted, any research in support of the assertion that the number of gay or lesbian parents is increasing.

Steckel (1987).

The author neither conducted, nor claimed to have conducted, any research in support of the assertion that the number of gay or lesbian parents is increasing.

Tasker and Golombok (1997).

The authors state in their second paragraph "It is not known how many lesbian mothers there are." (p 1). The researchers conducted a longitudinal study of 27 lesbian and 27 heterosexual single mothers. This research will be discussed in a later section. The authors neither conducted, nor claimed to have conducted, any research in support of the assertion that the number of gay or lesbian parents is increasing.

Supplementary Studies:

Bigner supplements the sources cited in the Vermont Brief with the following:

Faderman (1984)

The author of this article describes the homosexual identity formation process for lesbians. The article is based on a review of existing literature. There is no original research conducted nor reported.

Green and Bozett (1991).

The authors neither conducted, nor claimed to have conducted any original research. The authors state "Because homosexuals are an invisible population, accurate statistics on the number of gay fathers and lesbian mothers are impossible to obtain. However, based on the belief that 10% of the male population is gay, and that 20% of the gay male population has married at least once, and that 25% to 50% of this 20% have had children, the number of gay fathers in this country is likely more than two million. Add to this estimate the 6% to 7% of the female population is lesbian, and that between 1.5 and 3.3 million of them are mothers, the current estimates of children of gay fathers and lesbian mothers range between 5 million and 14 million" (198) (I omit the sources cited by the authors for these figures)

The estimates of gay fathers provided by Green and Bozett work out as follows. The lower bound estimate is 10% X 20% X 25% = 0.5% of adult males are gay fathers. The upper bound estimate is 10% X 20% X 50% = 1.0% of adult males are gay fathers. In 1990, when this article was published, there were approximately 84.5 million U.S. males over the age of 19[11]. Applying the authors' estimates, we arrive at between 422,500 and 845,000 adult gay fathers. Neither figure suggests more than 2 million such parents. (The same U.S. Census showed that there were 92.5 million females over the age of 19.

---

[11] http://www.census.gov/prod/1/gen/95statab/pop.pdf

If between 1.5 and 3.3 million of them are lesbian mothers, then between 1.6% and 3.6% of all adult females are lesbian mothers, not the 6% to 7% claimed by the author.

Flaks (1995).

The author neither conducted, nor claimed to have conducted any research in support of the assertion that the number of gay or lesbian parents is increasing.

Golombok, Tasker, and Murray (1997).

The researchers conducted an innovative project with some significant strengths. The objective was to investigate family functioning and the psychological development of children raised in fatherless families from their first year of life. The researchers assembled a non-probability sample of 30 self-selected lesbian mothers who "volunteered" for this project. They also assembled a non-probability sample of 42 heterosexual single mother "volunteers." Finally, they draw what appears to have been a probability sample of 42 heterosexual families from maternity records. The groups to be compared differed as one would expect when relying on volunteer subjects. There were significant differences in age of the mother, social class of the mother, and number of children among the groups to be compared. The authors relied on very good measures of family functioning and psychological development. Overall, the execution of the study was good (though it is not known how inter-rater reliabilities were established). There is no definition of "lesbian" or "heterosexual" provided by the researchers. Nor is there any indication of how these terms were applied to the subjects.

The authors statistically controlled for the differences among groups in mother's age, social class, and number of children in the family. Their results showed that single mothers showed greater warmth and interacted more with their child, but also reported more serious disputes. Children being reared without a father were found to be more securely attached to their mother, but perceived themselves to be less cognitively and physically competent than their peers from father-present families. Differences between lesbian and heterosexual single mothers were found only in the amount of interaction

between parent and child. Lesbian mothers interacted more frequently with their child than did heterosexual single mothers.

The sample sizes were too small to provide the statistical power needed to reliably detect no difference among groups given the statistical methods used. The reliance on "volunteer" subjects makes it impossible to estimate the biases that lead some people, but not others to volunteer for research projects. Though the authors discovered (and statistically controlled) for differences in several demographic factors, there is no way to know what other differences may also have existed, but were not discovered for failure to measure them.   This is a well-done exploratory study. It's results cannot, however, be generalized beyond the peculiar samples used in the research. There is no estimate of the number of lesbian couples, nor whether their number is changing.

Hoeffer (1981).
The researcher studied 20 lesbian and 20 heterosexual single mothers who resided in San Francisco I will discuss this research later. The author neither conducted nor claimed to have conducted any research in support of the assertion that the number of gay or lesbian parents is increasing.

Bozett (1981).
The author conducted interviews with 18 homosexual fathers in San Francisco. The author neither conducted nor claimed to have conducted any research in support of the assertion that the number of gay or lesbian parents is increasing.

Moses and Hawkins (1982)
Professor Bigner provides no citation for this reference other than the last names and date of publication. I could not locate the article in question.

Tasker, and Golombok (1995)

The authors report the results of a longitudinal study of 25 adults from lesbian families and 21 adults from heterosexual single mother families. They make no claims about the number of such families, their growth or decline, nor do they conduct or report any research relating to such claims.

Muzio (1996)

The author, a therapist, discussed one case in particular, and several others more generally in her advice to therapists treating lesbian mothers. The author notes: "Because individuals and families often seek therapy when their lived experiences contradict the dominant narrative about them, it is not unusual for lesbians to seek therapy at some point in their family building process (p. 367). This article is intended to provide advice to therapists when this happens. There is no research protocol, analysis, or comparison group involved. This is not a research article. The author makes no claims about the number of same-sex parents, or whether such numbers are changing.

Bailey, Bobrow, Wolfe, and Mikach (1995).

The authors neither conducted nor claimed to have conducted any research in support of the assertion that the number of gay or lesbian parents is increasing

Bigner (1996).

The author reviews the literature to provide guidance to therapists with gay father clients. There is no research conducted nor reported in this article.

Ricketts and Achtenberg (1990).

The case studies offered by these authors are not presented in support of the claim that the number of gay and lesbian parents is increasing.

Gay And Lesbian Parents Have the Same Motives for Becoming Parents

Bigner and Jacobsen (1989b).

The researchers rely on two samples obtained by different methods. Neither is a probability sample according to the authors. The sample of homosexual fathers was obtained from solicitations to a support group for gay fathers in Denver. The comparison group was selected from another project conducted by the senior author. The response rate for the homosexual sample was approximately 50%. There is no reported response rate for the sample of heterosexual fathers. The heterosexual fathers selected for this study were matched on age, martial status, income, ethnic identity, and education. No summary statistics are provided that would allow a comparison of the two groups on such measures. Subjects were mailed a questionnaire in most cases, though some subjects completed their questionnaires at conferences or workshops. The author acknowledges that the two samples were gathered under different conditions.

There is no operational definition of "gay." The comparison (heterosexual) sample is described as "presumed heterosexuals" because of the absence of such a definition. The researchers relied on good measures of parental behavior. The application of interferential statistics is not permitted with such a sample. The results of those statistical comparisons, however, reveal statistically significant differences between the two groups of fathers on several measures of parental behavior (limit setting, responsiveness, and reasoning/guidance).

The authors admit that the samples are biased due to high incomes. The authors also admit that the results cannot be generalized. "The sample of gay fathers is unlikely to be an accurate representation of gay fathers in the general population (p. 184). Other likely

biases are the result of different methods of recruiting the two samples, and different methods of administering the questionnaires.

Bigner and Jacobsen (1989a)

The authors rely on the same sample described above. In this article, the concern is how fathers responded to a measure referred to as the "Value of Children" questionnaire. Details of this questionnaire are not provided. Additionally, the same limitations that were described above apply to this study.

Siegenthalor and Bigner (2000)

Rather than report that there are no differences between the two groups in their motives for becoming children, the authors of this article actually report that lesbian and non-lesbian mothers differ only in their motives for becoming parents. They are not found (in the research reported) to differ in the value they place on parenthood (i.e., the satisfaction, the happiness, social status, or other benefits they derive from parenthood once children arrive) (p 84).

The authors assembled a non-probability sample of 25 self-selected lesbian and 25 self-selected non-lesbian mothers. The researchers recruited lesbians from lesbian support groups. They recruited the non-lesbians from other "parent support groups." Due to restrictions imposed by the IRB (Institutional Review Board for the Protection of Human Subjects), the researchers were unable to inquire about the sexual orientation of their subjects (p 82). As a result, they were unable to develop a definition of "homosexual" or "heterosexual" nor were they able to insure that subjects in each group met any definition of those terms. The two self-selected groups were matched on age, education, and income. Subjects rated the value of children on various dimensions. The scale used for this purpose has good reliability in repeated studies of heterosexual parents. Findings showed that lesbians differ from the non-lesbian parents in why they became parents. Lesbians were reported to be less likely to agree that "Having children gives a person a special incentive to succeed in life," "One of the highest purposes in life is to have

children," and "Having children makes a stronger bond between partners." (p 85). The use of volunteer samples, the inability to impose statistical controls to compensate for extraneous factors, and the very low power of the statistical tests make it impossible to generalize the findings of this research beyond the peculiar samples used.

Gay and Lesbian Parents Parenting Skills

Cases cited from the Vermont Brief

Green and Bozett (1991).

This is the source for the claim that "The home environments of lesbian and gay persons have been found to be as moral and as physically and psychologically healthy as those of non gays." (Vermont point 2). The authors of this (admittedly) ideological chapter neither conducted, nor claimed to have conducted, any research in support of the assertion that homosexual parents are as capable and caring as heterosexual parents. The chapter is a review of research by other authors.

Lesbian women as mothers:

Green, Mandel, Hotvedt, Gray, and Smith (1986).

These researchers relied on multiple methods. Mothers completed a self-administered questionnaire, and an interview was conducted with their children. The authors assembled two samples, neither of which is a probability sample. It is not known how many interviewers were involved, or whether inter-rater reliability was established. The first sample consisted of 50 lesbian mothers and their 56 children aged 3 to 11. The lesbian mothers were recruited through national and women's groups and through snowball sampling. The heterosexual sample was recruited through requests "for single-mother subjects" (no further details are provided). No operational definition of the term "lesbian" or "heterosexual" is provided except that lesbians were required to be "self identified" as such. . The authors administered good measures of personality and intelligence. Children were also interviewed about their peer groups, play preferences, and thoughts about life.

Inferential statistics are applied despite the fact that the samples are not probability samples. Both samples are (admittedly) biased. Though no comparison statistics (for each group) are provided, almost all subjects (86%) had completed college. There is no way to estimate the possible bias introduced by such high levels of education, nor of relying on members of women's groups. Nor are such groups described to permit the reader to assess the nature of the groups used for this project. But the authors note that 78% of the lesbians, but only 10% of the heterosexual mothers had partners living in the household. Clearly, even if no other differences existed, this simple and enormous difference invalidates any comparison between the groups without appropriate statistical controls. Such controls were not applied. The authors do not report the statistical results of their multivariate analyses, though they mention them.

Rand, Graham, and Rawlings(1982)

This research relied on a snowball sample of 25 self-selected lesbian mothers. There is no operational definition of lesbian except "self-identified." Of the 25 subjects, all but 9 had completed college, and 5 had graduate degrees. One of the measures used is highly regarded as a reliable indicator of psychological health. The other ("the affectometer") is reported to have very high reliability. The researchers compare their biased sample to national norms obtained from average samples. There was no comparison group. The most likely sources of bias are the extremely high level of education, and the fact that "all but two of the women in the present study had some degree of involvement in a lesbian community" (p 35). The authors acknowledge the bias introduced by using a snowball sample when they state "If more isolated lesbian mothers could have been included in the sample, correlations would probably have been significant." (p 35). I am unwilling to draw an conclusions from this research.

Flaks, Fisher, Masterpasqua, and Joseph (1995)

The authors rely on two non-probability samples. 15 "self identified" self-selected lesbian couples with children aged 3 to 9, and 15 heterosexual self-selected families were obtained by placing ads in lesbian newsletters, women's organizations, gay and lesbian

parenting groups, snowball sampling, and recruiting from a lesbian-mother support group. The heterosexual sample was a snowball sample. The authors used established measures with known reliability. Mothers were interviewed, in person, (it is not reported how many interviewers were involved) and reported about their children. Teachers also provided information about the children. There is no mention of response rates, and no way to calculate them from the information provided. Rather, all subjects were self-selected into the research.

The authors acknowledge the bias in their samples when they report that both groups of children (from "self identified lesbians" and presumed "heterosexual" families) differed significantly from national norms established for some of their measures. In fact, both groups of children scored higher than average on a measure of problem behaviors. As the authors acknowledge "The lesbian and heterosexual parent families studied here did not constitute random samples, and it is impossible to know what biases, if any, may have resulted as a consequence… We defined a precise and limited experimental group (i.e., lesbians)… Although the resulting sample was predominately White, highly educated, and economically privileged…"(p. 113). Indeed, 10 of the 15 lesbian mothers had graduate degrees, as did 9 of the 15 heterosexual mothers. The results of this research may not be taken as evidence in support of the assertion for which it is cited.

Miller, Jacobsen, and Bigner ((1981)
The authors rely on two non-probability samples. The lesbian sample consists of 34 self-selected mothers with custody who fit the operational definition of lesbian, i.e.,. "a woman psychologically, emotionally, and sexually attracted to another woman." (p 30). How this definition was applied is not explained. The authors refer to the sampling strategy as a "convenience sample" recruited through a feminist recreation center. The heterosexual sample was a convenience sample consisted of 47 mothers contacted at several Parent-Teacher Association meetings. Subjects completed a self-administered questionnaire, and responded to a slide show. The author notes that there was 100% inter-rater agreement in evaluating responses to the slides. All but two of the lesbians had completed college (94%). By comparison, 78% of the heterosexual subjects were

college graduates. The authors applied inferential statistics despite the samples. The author admits to no limitations on the data or the inferences drawn from them. To boost the power of the statistical tests, the authors increased the probability of a Type I error to .10 rather than .05. No statistical controls were conducted to compensate for differences between the samples. The very high level of education (especially among the lesbian sample) is one potential source of bias. The sampling methods, of course, are the most obvious problem. The results may not be generalized. This article cannot be taken as scientific evidence in support of the assertion for which it is cited.

Mucklow and Phelan (1979).

The authors describe this research as a pilot study. A purposive self-selected sample of 34 lesbian and 46 traditional mothers was located in the Denver-Fort Collins area. No details are provided on how these individuals were recruited. A lesbian mother is defined as a woman who is "psychologically, emotionally, and sexually attracted and interested in other women and who, from a previous relationship with a man, had conceived a child; or as a partner in a lesbian love relationship shared the parental role to a child" (881). The authors do not report how this definition was applied (i.e., how it was verified that all these criteria were satisfied). Members of the PTA were recruited for the heterosexual sample. No operational definition of "heterosexual" is described. One measure is reported to have high reliability. The other is reported to have low reliability. There is no way to assess the potential magnitude of bias introduced by the sampling strategies. Nor is it possible to compare the two groups on education, income, or any other measure except the two administered by the researchers. In the absence of any information about the sampling strategy, the results of this study are properly considered preliminary (a pilot study) and cannot be generalized beyond the peculiar samples used.

Lewin and Lyons (1982)

The authors assembled two non-probability, convenience samples. The first consisted of 43 self-selected divorced lesbian mothers and 37 self-selected divorced heterosexual mothers. The authors argue that there is no way to obtain representative samples of

lesbians. "obtaining a statistically representative sample of the lesbian mothers is not a realistic goal." (257). The authors recruited subjects through personal and professional referrals (snowball sampling), through publicity carried out in the local media, feminist and women's publications, newsletters published by child care and single-parent organizations, and posters. No statistical (quantitative) analysis is reported or conducted. The sample was quite biased with respect to education. Only 14% of the (combined) samples had educational levels lower than "some college." In-person, depth interviews were conducted. No report is made of the number of interviewers, nor of attempts to estimate inter-rater reliabilities. In the absence of information about the sample, the ratings of interviews, or any quantitative analysis, this study must be regarded as inadequate for purposes of the assertion it is cited to support.

Lyons (1982).

This study uses the same sample and methods described above.

Kweskin and Cook (1982).

These researchers assembled two non-probability samples by "purposive" (i.e., self-selected) means. There is no mention of how the sample of 22 lesbian mothers was recruited. The 22 heterosexual mothers were recruited from Parents Without Partners. The authors used versions of a well-known and reliable measure of gender role preferences (i.e., masculinity/femininity). Subjects completed a self-administered mailed questionnaire. No mention of the response rate is made. Without additional information about how the lesbian sample was recruited, or how the term "lesbian" was defined, it is impossible to determine the magnitude of any sampling bias. Without information about response rates, it is impossible to determine the magnitude of self-selection, even in these purposive samples.

Falk (1989)

The researcher neither conducted, nor claimed to have conducted any research in support of the assertion about lesbian mother.

Harris and Turner (1985/86)

The researchers assembled two non-probability samples. The sample of self-selected gay parents included 10 "self-described gay males, and 13 lesbian females. The sample of self-selected heterosexual parents included 2 heterosexual male single parents, and 14 heterosexual female single parents. Subjects were recruited by posters on campus and in a gay bar, advertisements in local newspapers, and an article in a gay/lesbian newsletter. Subjects were instructed to pick up questionnaires at designated locations. In addition, visits were made to meetings of a campus gay/lesbian organization, a convention of a gay/lesbian church, a Parents without Partners meeting, and several day care centers. No details are provided about the instrumentation, or reliability. It is impossible to establish response rates with samples generated by self-referral. 78% of the homosexual sample, and 87% of the heterosexual sample had college degrees. The authors do not present descriptive statistics for the heterosexual sample though they do for the homosexual sample. The sampling design makes it impossible to determine the magnitude of likely bias, though the very high levels of education are surely problematic. The authors acknowledge that their study is not representative of either gay or heterosexual parents "Thus, all generalizations must be viewed with caution." (p. 111). The sampling methods and the sample sizes were inadequate for the statistical methods used (p. 112). The results of this study do not support the assertion for which it is cited.

Lott-Whitehead and Tully (1993)

The researchers assembled a snowball sample of self-selected lesbians by using "friendship networks, word-of-mouth referrals, etc." (p 268). There was no comparison group. 187 questionnaires were distributed, of which 46 were returned (response rate = 25%). The primary method of analysis was qualitative rather than statistical. Of the 46 subjects, only 2 had less than a college education. The authors acknowledge that the research "had inherent in its design methodological flaws consistent with other similar studies…This study does not purport to contain a representative sample, and thus generalizability cannot be assumed" (p 269). In light of the very low response rate, the education bias, the lack of detail about the instrumentation, and the acknowledged flaws

in design, the results of this study cannot be used to assess the assertion for which it is cited.

Gay Men as Fathers:

The following articles are cited in support of the assertion that "Research focusing on parenting skills and attitudes of gay fathers similarly confirms that gay men are suitable, and indeed, admirable parents." (Vermont Brief).

Bozett (1989)

This is a review of the literature. The author neither conducted, nor claimed to have conducted research regarding the role of gay men as fathers.

Bigner and Jacobsen (1992).

The researchers assembled two non-probability samples. The gay sample consisted of 24 self-selected men recruited from a gay father support group. The heterosexual sample consisted of 29 self-selected fathers recruited from members of Parents without Partners. There are no statistical results presented for the substantive comparisons of the two groups. There is no operational definition of "gay" except "self identified. gay". The comparison group, therefore, was "presumed to be non-gay." (p. 103). The instrumentation consisted of slides to which men responded and a series of attitude questions. No evidence on reliability is provided. The sample size is too small to provide the power necessary for the test of the null hypothesis. I could find no evidence that the researchers controlled statistically for extraneous factors. All measurements appear to have been made by a single member of the research team, so inter-rater reliability is irrelevant. The two groups of subjects differed noticeably on educational attainment. Only 10% of the heterosexual sample had college or advanced degrees, compared with 67% of the homosexual sample. The likely sources of bias include the use of a single interviewer without attempts to establish reliability, the obvious differences in the two samples that are not dealt with by introducing statistical controls, and the unknown reliability of the instruments.

Bigner and Jacobsen (1989b)

This study was reviewed and critiqued earlier.

Scallenn (1981)

This is an unpublished Ph.D. dissertation that was not reviewed for this brief.

Harris and Turner (1985/86).

This article was reviewed and critiqued earlier.

The Children of Gay and Lesbian Parents are as Well Adjusted as Those of Their Heterosexual Counterparts.

Patterson (1992)

The researcher assembled one non-probability sample of 37 families (26 lesbian couples, 7 single lesbian mothers, and 4 separated/divorced lesbians – some of whom had partners) producing 66 self-selected lesbian subjects. All but four subjects were employed, and 44 of the 66 had at least a college education.   All children in the families were born, or adopted by lesbians, and therefore had grown up for their entire lives in such families. This design minimizes the inherent biases that would be present in studies that focus on children (of gay or lesbian parents) who were born in heterosexual relationships (as in almost all studies cited thus far). Sampling was by snowball methods. Of 39 families contacted, 37 agreed to participate. There was no comparison group. Instead, the researcher compared the children in such families to national norms established for the reliable measures used to assess children's well being. There is no report about how many researchers participated in the collection of information in face-to-face encounters in the subject's home. No statistical controls were applied to compensate for extraneous factors, though such controls would have been of little value absent a comparison group. In the end, findings about how the children from these affluent, self-selected lesbian families compare with national norms is of little statistical value because national norms are established on average, heterogeneous samples very unlike the sample used in the current study.

Patterson and Redding (1996)

This is a review of family laws relevant to lesbian and gay parents. The researchers neither conduct, nor claim to conduct any research pertaining to the parental abilities of homosexuals.

Bigner and Bozett (1990)

This is a review of the literature on gay parents. The researchers neither conduct, nor claim to conduct any research pertaining to the parental abilities of homosexuals.

Brewaeys and Van Hall (1997)

This is a balanced and reasoned review of the literature on lesbian motherhood. The authors do not conduct, nor claim to conduct any original research pertaining to the parental abilities of homosexuals.

Cramer (1986)

This is a review of the literature on gay parents. The authors do not conduct, nor claim to conduct any original research pertaining to the parental abilities of homosexuals.

Falk (1989)

The researcher neither conducted, nor claimed to have conducted any research in support of the assertion about lesbian mother.

Gottman (1990)

This is a review of the literature on gay and lesbian parents. The authors do not conduct, nor claim to conduct any original research pertaining to the parental abilities of homosexuals.

Green and Bozett (1991)

This article was reviewed and critiqued earlier.

Kirkpatrick (1987)

This is a review of several clinical cases seen for therapy by the author. There is no sampling, instrumentation, or research protocol

Kirkpatrick (1996)

This is a review of the literature on lesbian parents. The authors do not conduct, nor claim to conduct any original research pertaining to the parental abilities of homosexuals

MacCandlish (1987)

The author invited five lesbian mother families to participate in a two-hour structured interview in the subject's home. There are no details provided about how these families were recruited, their backgrounds, their motivations to participate, or the instrumentation used. There was no comparison group, and there was no analysis (quantitative) of results. No scientific inferences may be drawn from this project.

O'Connell (1993)

This was described as an "exploratory design" involving "open-ended" interviews with a questionnaire guide. It relied on a non-probability self-selected sample of 6 lesbian (age 16-23) and 5 gay (aged 19 to 23) parents obtained through snowball methods, and by placing an advertisement in two Boston gay newspapers and a local woman's newspaper. The subjects had all experienced their parents' divorce. Interviews were conducted by the researcher only. No mention of reliability is made. Instrumentation is not described sufficiently to judge its quality. There was no comparison group. No scientific inferences may be drawn from this project.

Patterson and Chan (1997)

This is a review of the literature on gay fathers. The researchers do not report any new research in this article.

Pennington (1987) # 31

The researcher describes a clinical sample of 32 children from 28 lesbian mother families that she treated since 1977. All but 2   children were seen at an outpatient psychotherapy clinic for gay men, lesbians, and their families in San Francisco. The author provides her

impressions of the issues that these children faced as they dealt with their parents' troubles. There is no sampling, instrumentation, or statistical analysis. There was no comparison made with children from heterosexual parents.

Pies (1990)

This is not a research article. It is a journalistic account of personal experiences.

Allen and Burrell (1996).

The analysis of this study is described in detail in the body of the affidavit.

Chan, Raboy and Patterson (1998).

The researchers assembled a self-selected sample by recurring families from former clients of The Sperm Bank of California. Clients who conceived and gave birth to children at least 5 years before the study was conducted were invited to participate. 195 families were so identified. The researchers were able to locate and contact 108 (55%). Of these 108, a total of 80 (74%) agreed to participate. The overall response rate, therefore, was 80/195 = 41%. Response rates, however, differed dramatically by sexual orientation of the parent. All eligible lesbian couples (100%) participated. But only 30% of lesbian single mothers, 31% of heterosexual couples and 30% of heterosexual single mothers participated.

As almost every study reviewed so far has found, these researchers note that the sample of lesbian biological mothers had significantly more education than did others. The lesbians also had higher average incomes. We cannot know about the majority of heterosexuals who decided not to participate.

The researchers administered several well-known and reliable measures of children's well being.

The authors acknowledge the limited power of their statistical analyses, and failed to incorporate statistical controls for the differences (in education and income) found among their groups.

Several potential sources of bias are acknowledged. First, the initial contact with potential subjects was from The Sperm Bank rather than the researchers. The extent to which this elicited differential participation rates is unknown. But surely, the dramatically different response rates are a critical source of concern for the results of this study. The failure to control for (acknowledged) differences among groups is also a flaw in the analysis. And probably most important, the use of women who have been artificially inseminated raises very serious questions about how representative this group of lesbians is. Due to the problems with the sample and methods of analysis, no scientific inferences may be drawn from this research.

Brewaeys, Ponjaert, Van Hall, and Golombok (1997)

This is a well-designed analysis that attempted to study entire populations rather than samples of them. The "sample" of 30 lesbian mother families with children (aged 4-8) conceived through Donor Insemination was recruited through the Fertility Department of the Brussels University Hospital. All families where the mother had attended the clinic between 1986 and 1991 were asked to participate. The agreement rate was 100%. The comparison group of 38 heterosexual Donor Insemination families and of 30 naturally conceived heterosexual families was recruited through the Fertility Department and the Obstetric Department of the University Hospital Leiden. All heterosexual families with a child born between 1986 and 1990 were asked to participate. Similar requests were made to parents whose children were born naturally. Response rates were 53% for the heterosexual Donor Insemination families, and 60% for the naturally conceived families. In-home interviews were conducted. It is fair to say that the sample may be considered broadly representative for the general population of lesbian mothers who attended a fertility clinic in order to conceive. Response rates and self-selection biases for the other groups jeopardize the degree to which each represents the relevant population, although the procedure is vastly superior to almost all others reviewed in this brief.

Comparisons of the groups revealed that they differed on educational levels with lesbians having considerably higher average educational attainments. Education, however, was not controlled in subsequent analyses. Much of the instrumentation consisted of reliable measures of child well being. The statistical analyses (though lacking needed controls) revealed no significant differences in the quality of relationships between lesbians (and their partners) and heterosexual couples. Nor was the parent-child relationship different among groups when biological mothers were compared. Unfortunately, the samples were too small to draw any conclusions about the lack of difference between groups (i.e. the study lacked sufficient statistical power). And finally, this study suffers from the same problem noted above, women who have been inseminated by artificial methods are likely to differ in important, yet unknown ways from lesbians who have conceived naturally. Still, despite the obvious limitations, this is one of the better studies among all that were reviewed.

Flacks, Ficher, Masterpasqua, and Josepth (1995)

This study was reviewed and critiqued above.

Steckel (1985)

This citation refers to an unpublished doctoral dissertation.

Golombok, Spencer, and Rutter (1983).

This study was reviewed and critiqued above.

Tasker and Golombok (1997)

This study was reviewed and critiqued above.

Gottman (1990)

This review of the literature was discussed earlier.

Schwartz (1985) Unpublished dissertation)

This citation refers to an unpublished doctoral dissertation.

Karkpatrick, Smith, and Roy (1981)

The researchers report on a study of the psychological status of a non-probability self-selected sample of 10 boys and 10 girls living full time with their "self identified" lesbian mothers. A comparison group of 10 boys and 10 girls living full-time with their single-heterosexual mothers was also evaluated. Mothers were recruited through snowball sampling and with a request in a National Organization for Women newsletter. All participants in the study, therefore, were self-selected. Each child was evaluated by several different researchers. No descriptive information is provided that would allow me to assess the differences between the two groups in terms of education, income, or other possibly extraneous influences. No information is provided about the children's backgrounds that might allow the reader to assess the findings in light of such factors. In the absence of statistical analysis, the authors conclude "lesbian mothers and heterosexual mothers were very much alike in their marital and maternal interests, current life-styles, and childrearing practices." (p 550). This is a good qualitative study, though it does not offer scientific evidence about the comparative profiles of the two groups.

Puryear (1983)

This citation refers to an unpublished doctoral dissertation.

Rees (1979) unpublished doctoral dissertation

This citation refers to an unpublished doctoral dissertation.

Barrett and Robinson (1990)

This is not a research report but a series of case studies of an unknown group of children of gay fathers. The authors raise an important point. They note: "In reviewing the impact

of gay fathering on children, it is important to acknowledge that most children who live with gay fathers are also the products of divorce and may present psychological distress that typically accompanies families experiencing marital dissolution. All too often the emotional distress of children with gay parents is solely attributed to the parents' sexual orientation rather than seen as a complex mixture of family dynamics, divorce adjustment, and incorporation of the parents' sexual coming out." (p 82). In making this point, the author reminds us that research on this subject must control for such obvious factors. Failure to do so will bias the results of any study. None of the studies reviewed controlled for such factors.

Golombok, Spencer, and Rutter (1983) # 10

The researchers report the results of studies conducted on non-probability samples of 27 lesbian families with a total of 37 children, and a comparison group of 27 heterosexual families with a total of 37 children. The definition of "lesbian" used was that a women must regard herself as predominantly or wholly lesbian and must currently be in a homosexual relationship, or have been in one in her last relationship. "Heterosexual" was defined behaviorally, by recruiting women whose last sexual relationship was with a man. Personal interviews were conducted. Instrumentation is not described in detail, reliability of indicators is not reported, nor is inter-rater reliability noted.

The two groups differed in an important way. All of the single-parents lived alone with their children. Most of the lesbians lived with a partner (only 9 of 27 lived alone with their children). Though the two groups were similar in regards age, and past marital status, they differed importantly on educational levels (67% of the lesbians and 37% of the heterosexual women had advanced education/training (p 556). The two groups of mothers also differed in their contact with their children's father.

Despite the differences between the two groups, appropriate statistical controls were not employed to adjust for these differences. The authors acknowledge the limitations of these results when they note "It is not possible to know what biases were involved in the method of sample selection." (569). Moreover, since almost all the children had been

born into a heterosexual household where they had spent at least two years, "This may be relevant in that both gender identity and sex role behavior are established early in the preschool years and the roots of sexual object choice (in so far as they are experiential) may also be found in the same age period. Accordingly, it would be unjustified to generalize our findings to rearing in a lesbian household from the outset." (p 569)

Huggins (1989) # 15

This article reports the results of a study of self-esteem among adolescents. The author assembled two non-probability samples. 36 adolescent children (13 to 19) from 32 families were divided into two groups based on their mother's sexual "object choice." The resulting samples contained nine male and female adolescents each. There is no description of how the sample was selected or obtained. The author notes that "to be asked to participate in the study, the children had to be aged 13 to 19 years and be living with their self-designated lesbian mother or self-designated heterosexual mother. The children were the biological products of a heterosexual marriage that had ended in divorce at least one year prior to the time of the study." (p 126). The author relies on a well-known measure that has established reliability in large samples. Presumably, all in-person interviews were conducted by the author (though this is not mentioned). There are no statistical controls used to compensate for potential extraneous factors. And without any information about how the sample was obtained, it is not possible to comment on the likely biases inherent in this project.

Green, Mandel, Hotvedt, Gray, and Smith (1986)

This research was reviewed and critiqued above.

Hotvedt and Mandel (1982) # 45

The authors report a study of "self designated" lesbians with custody or joint custody of at least one child (age 3-11) and a matched heterosexual sample of "self designated" heterosexual single mothers. Sample sizes are not reported. Sample recruitment strategies are not reported. The author made a good attempt to deal with extraneous

factors by matching the (unknown size) samples on age, race, and marital status of the mother, sex of the child, length of separation from the father, income of the family, education of the mother, and mother's religion as a child. Self administered questionnaires appear to have been employed. There is no description of the instrumentation except for well-known measures of mental ability. No results are presented. No response rate can be calculated. Without any description of the sample, or any statistical results, it is impossible to evaluate this study.

Lesbian and Gay Parenting at (American Psychological Association)(1995).

This is a joint publication of the American Psychological Association's Committee on Women in Psychology, Committee on Lesbian and Gay Concerns, and Committee on Children, Youth, and Families. It is written by Professor Charlotte Patterson, and is a review of the literature and annotated bibliography. It is not a research article.

Rees (1979)

This citation refers to an unpublished doctoral dissertation.

Flaks, Fischer, Masterpasqua, and Joseph (1995)

This study was reviewed and critiqued above.

Green, Mandel, Hotvedt, Gray, and Smith (1986)

This research was reviewed and critiqued above.

Kirkpatrick, Smith and Roy (1981)

This article was reviewed and critiqued above.

Golombok, Spencer and Rutter (1983)

This article was reviewed and critiqued above.

Kirkpatrick (1987)

This summary of several clinical cases was reviewed and critiqued earlier.

Patterson and Redding (1996)

This review of the literature was discussed earlier.

Bailey, Bobrow, Wolfe and Mikach (1995)

The researchers recruited a non-probability sample of 55 gay and bisexual fathers through advertisements in homophile publications. These self-selected men were asked to discuss their sons. The sons were subsequently contacted by the researchers. Of the total of 82 sons available, information was gathered from 43 (52%). Instrumentation is not described, and there are no reports on reliability. There was no comparison group. The number of interviewers is not reported, nor are inter-rater reliabilities reported. 9% of the (contacted) sons were found to be homosexuals, though no operational definition of that term is provided. Rather, both fathers and sons were asked to characterize (the sons) as homosexual, heterosexual, or bisexual, allowing the subjects to define the terms as they wished.

The authors acknowledge the most serious potential bias of the study, self-selection. "The most important potential bias is that fathers decisions to participate might depend in part on their sons' sexual orientations. .. The second limitation concerns the absence of a control group." (p 127). Most interestingly, the researchers acknowledge that the rate of homosexuality among the sons of gay men is higher than found in the general population. "It could be argued the rate of homosexuality in the sons (9%) is several time higher than that suggested by the population-based surveys and is consistent with a degree of father-to-son transmission." (p 128). The authors argue that this is not the case, however, due to the design problems of the study and the sample. The authors appear unwilling to accept

the findings of their own study and go to lengths to explain why the results should not be interpreted on their face. .

Golombok, Spencer and Rutter (1983)
This study was reviewed and critiqued above.

Golombok and Tasker (1996)
This study is reviewed in the body of my affidavit.

Gottman (1990)
This article was reviewed and discussed earlier.

Green, Mandel, Hotvedt, Gray, and Smith (1986)
This research was reviewed and critiqued above.

Green (1978)
The author reports on his study of the sexual identity of 37 children raised by homosexual or transsexual parents. The author (a psychiatrist) examined 37 children who were being raised by at least one parent who was either transsexual or homosexual. This is a clinical sample and cannot be regarded as representative of any defined population. The instrumentation (psychiatric treatments) are not detailed. There is no mention of reliability given that the author conducted all sessions. There is no comparison group.

Hoeffer (1981)
The author reports the result of a comparison of 20 lesbian and 20 heterosexual single mothers from the San Francisco Bay area and their only or oldest child, ages six through nine. The definitions of homosexual and heterosexual are "self identified." The author gives no indication of how the subjects were recruited. No comparative statistics are

provided to permit a comparison of the two groups. A modified version of a reliable measure of children's toy preferences was employed. All interviews were conducted by the author in the home of the subject. Without information about how the two groups compared (on, for example, education, age, income, race, etc.) or how the subjects were recruited for the study, it is impossible to comment on the potential biases in this study.

Kirkpatrick, Smith, and Roy (1981)
This study was reviewed and critiqued above.

Miller (1979)
The author conducted depth interviews with a snowball sample of 40 homosexual fathers and 14 of their children. No further description of the sampling is provided. No details are offered about the instrumentation. No comparison group was involved. There is no discussion of how "homosexuality" was measured. The author reports that 3 of the 14 men said they had fantasized about having sex with their sons (but none had ever acted on it)(p546). One in six sons, and one in eight daughters were homosexuals. This finding led the author to conclude "On the basis of this small, nonrandom sample, there does not appear to be a disproportionate amount of homosexuality among the children of gay fathers."(p 547) despite the absence of any comparative evidence from heterosexuals.

Schwartz (1986)
This is an unpublished doctoral dissertation. .

142.   The comments and analysis contained in the main body of this affidavit and the six Appendices that follow the main body comprise the totality of my opinion in this matter

Sworn before me at the City

of   in the State

of Virginia, in the United

States of America, this

day of March, 2001

_____

Steven L. Nock

_____

Notary Public