

How Statistics Can Mislead

The Case of “Family Caps” in State Welfare Programs

by Adam M. Zaretsky

There’s no question [that reducing welfare benefits] would be some incentive for people not to have dependent children out of wedlock.

--President Bill Clinton

Nonsense. We really don’t know what to do, and anyone who thinks that cutting benefits can affect sexual behavior doesn’t know human nature.

--Sen. Daniel Patrick Moynihan

Whenever the results of a study are released, the media, interest groups and policymakers often use the findings to declare victory or dispute the ill-founded methods by which the results were obtained. Headlines like “Reform Proves Overwhelmingly Effective” are quite adept at luring readers into an article, where terms like “statistically significant” are relegated to small notes in charts or left out altogether. When used, these terms are seldom defined, and, when they are, their definitions are often inaccurate.

Recently, announcements about an experimental welfare program in New Jersey illustrated well how jumping the gun to release preliminary findings without careful analysis of the

procedures used to obtain them can lead to a misinterpretation.

This example demonstrates clearly how injudicious interpretation of data can lead to a particular conclusion, one that was later rebutted by a more in-depth study. The reversal of direction in this instance hinged on the concept of statistical significance.

What's Significant about Statistical Significance?

In a nutshell, statistical significance refers to the probability that a relationship apparent in the data is not merely a coincidence, but, rather, is a result of a performed experiment. For example, suppose a researcher wants to determine whether a drug helps relieve the symptoms of a particular disease. He gives the drug to a randomly selected group of patients—the experimental group—while giving a placebo to another group of patients—the control group. Presumably, some patients in each group will get better. The researcher hypothesizes, though, that more patients in the experimental group will improve because of the drug.

Statistical significance describes the likelihood that the observed improvement rate for the experimental group occurred because the drug was effective. In other words, relative to the control group's outcome, is the observed outcome for the experimental group caused by the drug, or is it just a fluke? The

result is statistically significant if the researcher can determine that the outcome is not likely to be just a coincidence.

The same analysis can be applied to economic questions too. For example, determining whether the observed outcomes—say, a change in the birth rate—from the experiment of withholding money from some program participants, while giving it others, is coincidence or the result of the experiment is answered using statistical methods. The interpretation of these outcomes—or the difference between the outcomes of the two groups—should, therefore, be couched in terms of statistical significance. Otherwise, the reader is left to guess which scenario better reflects the true state of affairs. The progress reports about New Jersey’s welfare reform program illustrate this point.

New Jersey’s Experiment

In October 1992, New Jersey enacted the Family Development Program (FDP). One provision of this program is the child exclusion law, more commonly known as the “family cap.” This provision essentially stipulates that a family receiving Aid to Families with Dependent Children (AFDC) benefits will receive cash assistance for only those children born or conceived

before the mother's application for AFDC.¹ Effectively, then, the family cap did not start until August 1993. It eliminates the additional cash benefit, between \$64 and \$102 each month, upon the birth of an additional child. The infant's eligibility for Medicaid and food stamps is not affected.

It Works...Or Does It?

About three months after the family cap became effective, the administration of then-Gov. Jim Florio issued a press release claiming a 16 percent reduction in the number of children born to AFDC families as a result of the new regulations. This claim was based on a comparison of the AFDC birth rates for August 1993 and September 1993 with the same two months of 1992. Because

¹AFDC is the national welfare program for which both the federal and state governments share expenditures. The federal government sets the rules for eligibility, but the states are free to determine actual benefit levels and administer the funds. Generally, poorer states receive more funds from the federal government than wealthier states. This skewed distribution, however, does not necessarily translate into higher benefits for recipients.

of this sizable reduction in the birth rate, the family cap was declared an “obvious success.” It later became obvious, though, that this conclusion was too much, too soon. When the data were revised just four months later, the 16 percent reduction had dwindled to 9 percent.

Compounding the problem, a more relevant piece of information was ignored: the relationship between the birth rate in the project’s experimental group—those subject to the family cap—and that in the control group—those not subject to the family cap. In fact, the releases did not report if *any* change in the birth rate of the control group had occurred. This omission prevents the comparison, leaving the crucial question about the statistical significance of the difference unanswered.

More Concrete Evidence

In August 1994, a study was released that reported a substantial reduction in AFDC birth rates during the first 10 months of the FDP.² June O’Neill, currently Director of the Congressional Budget Office and formerly of Baruch College, conducted this analysis, in which she found a 19 percent reduction

²This study was conducted as a defense report in a case brought against the State of New Jersey challenging the family cap

in births, at the request of the New Jersey attorney general's office.

The accompanying table, which is Table 2 from O'Neill's study, records the birth rates for AFDC mothers both before and after the family cap became effective. The 19 percent reduction represents the ratio of the -1.29 in the "Difference" column and the 6.75 in the "Control Group" column. No statistical significance is attached to this 1.29 percentage point difference in birth rates, however, so the reader does not know whether this gap is statistically different from zero or not.

Moreover, it is not clear whether the 1.29 percentage point difference between the two groups' birth rates is the most relevant piece of information for determining the FDP's success because it does not measure the change in the birth rates over time. The more interesting comparison requires calculating the change in the birth rate *within* each of the two groups between the pre-cap and post-cap periods. These calculations are included in the table as the bottom row marked "Difference between periods." The question to ask is whether -5.96 is statistically different from -5.24; in other words, is the gap between these two numbers statistically significant? An answer of "yes" would represent evidence that the

provision of the FDP.

family cap is achieving its goal. Unfortunately, this information was not reported.

The Rutgers Study

To obtain better information about the effects of the FDP, New Jersey commissioned Rutgers University to conduct a five-year study. In June 1995, it released its preliminary findings; the actual study has not been officially released. For the period August 1993 to July 1994, the study reports that 6.7 percent of women in the control group gave birth, and that 6.9 percent of those in the experimental group gave birth. The 0.2 percentage point difference between these rates is *not* statistically significant; that is, the birth rates are essentially the same number. More important, this result was the same when controls for the pre-cap birth rate differences were included in the analysis. This last step is tantamount to testing whether one birth rate changed “significantly” more than the other across periods. It is important to remember, though, that the results do reflect only one year of available data. Further information from the study will be available with time, and it is quite possible—even probable—that these numbers will be revised again, although it is unlikely the revisions will qualitatively alter the result.

Spillover Effects

Whenever states initiate experimental or demonstration projects, their effectiveness can be gauged only after sufficient time has passed to allow for reasonable data accumulation and processing. Clearly, three months into a project the scope of New Jersey's is too early to draw any definitive conclusions. One year is probably also too soon. This is the reason five-year studies are usually conducted with experimental projects in federally funded programs.

Perhaps prematurely, other states jumped on the bandwagon to institute similar changes in their welfare laws after New Jersey's initial reports of success were announced. The release of the Rutgers study, with its more in-depth analysis, should cause these states to re-evaluate their initiatives and pay closer heed to preliminary findings in which critical bits of information—like the statistical significance of results—are glossed over or ignored.

Adam M. Zaretsky is an economist at the Federal Reserve Bank of St. Louis. Thomas A. Pollmann provided research assistance.

REFERENCES

- Camasso, Michael J. Letter to Rudolf Myers, Assistant Director,
Division of Family Development, State of New Jersey
(June 14, 1995).
- Kramer, Michael. "The Myth About Welfare Moms," *Time* (July
3, 1995).
- Laracy, Michael C. "If It Seems Too Good To Be True, It
Probably Is." The Annie E. Casey Foundation (June 21,
1995).
- O'Neill, June. "Report Concerning New Jersey's Family
Development Program." Baruch College (August 1994).

**Birth Rates of Mothers on AFDC in New Jersey between
August 1992 and June 1994**

	Experimental Group	Control Group	Difference ¹ (Experimental-Control)
Number in sample	1,777	859	
Percentage with a birth in the period August 1992 through July 1993	11.42%	11.99%	-0.57%
Percentage with a birth in the period August 1993 through June 1994	5.46%	6.75%	-1.29%
Difference between periods (percentage points) ²	-5.96	-5.24	

SOURCE: O'Neill (1994)

¹Although these percent signs are included in O'Neill's original table, it is more accurate to refer to these figures as percentage point changes rather than percent changes.

²This row is not in O'Neill's original table, but was added for illustration.