

# How Do the Foundation for Government Accountability's Evaluations of Welfare Reform Measure Up? A Report Card (Hint: The FGA Fails)

Peter Germanis<sup>1</sup>

June 24, 2018

In “They’re the think tank pushing for welfare work requirements. Republicans say they’re experts. Economists call it ‘junk science,’” Caitlin Dewey of *The Washington Post* describes the newfound influence of the Foundation for Government Accountability (FGA) in the national debate over welfare work requirements.<sup>2</sup>

House Republicans – including [Speaker] Ryan, who was introduced to the group in 2016 through Kansas Gov. Sam Brownback – have repeatedly proffered the FGA’s analysis as proof that most Americans support strict work rules in welfare programs and that such rules boost income and employment.

... And it has churned out a steady stream of infographics, opinion polls and first-person videos to promote its policies, many trumpeting a 2016 FGA study from Kansas that claimed that the reinstatement of SNAP work requirements prompted thousands of unemployed Kansans to get jobs and more than doubled their average incomes.

... On the farm bill, the FGA mobilized in a huge way to support the Republican plan, canvassing lawmakers’ offices and churning out a stream of widely disseminated “one-pagers” designed to deflect both liberal and tea party criticisms.<sup>3</sup>

While the FGA’s “studies” have support among some politicians, their work is not seen as credible by serious observers. To understand why, it is informative to compare their methodological approach for making claims about the impacts of work requirements and other welfare policies with generally accepted criteria for assessing the soundness of an evaluation.

In *Evaluating Welfare Reform: A Guide for Scholars and Practitioners*, Douglas Besharov (professor at the University of Maryland School of Public Affairs and Director of its Welfare Reform Academy), Peter Rossi (was S.A. Rice Professor Emeritus at the University of Massachusetts, Amherst, and past president of the American Sociological Association), and I (then Assistant Director of the Welfare Reform Academy) described the need for and importance of rigorous evaluations in the immediate aftermath of the 1996 welfare reform law:

Many of the evaluations will provide important information about the impact of the new welfare regime on individuals and institutions. They will identify the difficulties and successes that states have had in implementing their reforms, and estimate the impacts of such reforms on the well-being of the poor, especially on their children. These findings, in turn, can help policymakers choose between various program approaches. ...However, many of the evaluations will have such serious flaws that their utility will be sharply limited.<sup>4</sup>

We further cautioned:

The number of studies purporting to evaluate welfare reform will grow rapidly in the years to come. The challenge facing policymakers and practitioners will be to sort through the many studies and identify those that are credible. It is a task that will be complicated by the volume and complexity of the studies, and the highly charged political atmosphere that surrounds them.

...The key issue is the extent to which a discerned fault reduces the credibility of a study. Unfortunately, most policymakers and practitioners are ill-equipped to judge which faults are fatal, especially since they often must act before the traditional scholarly process can filter out invalid results. This is understandable, since assessing evaluation studies often requires both detailed knowledge of the programs involved and a high level of technical expertise. To help them better assess this research and glean the lessons it offers, this paper also describes and explains the generally accepted criteria for judging evaluations.<sup>5</sup>

The review criteria (described in more detail below) are:

- **Program Theory:** Does the program or evaluation make sense in light of existing social science knowledge?
- **Research Design:** Does the research design have both internal and external validity?
- **Program Implementation:** How does the program actually operate?
- **Data Collection:** Is the necessary data available and reliable?
- **Measurement:** Are the key variables valid and can they be measured reliably?
- **Analytical Models:** Are the data summarized and analyzed by means of appropriate statistical models?
- **Interpretation of Findings:** Are the findings interpreted objectively and do they describe the limitations of the analyses and consider alternative interpretations? Are they placed in the proper policy or programmatic context?

These criteria, of course, are not equally applicable to all evaluations, but there have been many rigorous and detailed evaluations of state welfare reforms (including mandatory welfare-to-work programs) that measure up well against these criteria and the questions posed.<sup>6</sup> In contrast, the FGA's "studies" are generally limited to a few pages of text and fail to adequately address any of these criteria, and thus their "studies" rate a grade of "F" on each criterion and overall. For a prime example of the FGA's fundamentally flawed "research," see:

- Jonathan Ingram and Nic Horton, *The Power of Work: How Kansas' Welfare Reform Is Lifting Americans Out of Poverty* (The Foundation for Government Accountability, February 16, 2016), available at: <https://thefga.org/download/PowerOfWork-KansasWelfareReform.pdf>.

The main body of the report is 19 pages; if one removes the colorful charts, the actual text is about 8 pages of misleading statements and demonstrably false conclusions.

The remainder of this paper describes the criteria used to assess the soundness of evaluations in greater detail and provides an example or two of how the FGA fails. It will focus on the aforementioned FGA “study” of a “work requirement” in Kansas. (The Appendix provides a much longer list of critiques of this and other FGA “studies.”)

### **The FGA’s Kansas Study**

Writing in *Forbes*, Nic Horton, Jonathan Ingram, and Josh Archambault of the FGA touted the findings of the Kansas “study” – one they characterized as “the first of its kind.” It examined the effects of a “work requirement” in Kansas on able-bodied adults without dependents (ABAWDs) receiving benefits from the Supplemental Nutrition Assistance Program (SNAP, formerly known as food stamps).<sup>7</sup> Specifically, SNAP limits able-bodied adults aged 18 to 49 without children to three months of benefits in any 36-month period, unless they are employed or in a work or training program for at least 20 hours a week. States do not have to offer these individuals a work or training program slot, and most do not, so this provision is more appropriately seen as a “time limit” rather than a “work requirement.”<sup>8</sup>

The FGA “researchers” summarize their results as follows:

New research from the Foundation for Government Accountability examines the results of Kansas’ welfare reforms and the findings are simply staggering. Work requirements have led to more employment, higher incomes, and less poverty. And instead of drawing millions of dollars in taxpayer-funded welfare benefits, these able-bodied adults are now contributing tens of millions of dollars to the local economy.<sup>9</sup>

Some of the specific “findings” reported in the “study” include:

- “Nearly 60 percent found employment within a year of leaving food stamps.”
- “After leaving food stamps, incomes increased 127 percent.”
- “Within a year of leaving food stamps, the number of able-bodied adults living in poverty dropped significantly and roughly half of those working climbed out of poverty entirely.”<sup>10</sup>

### **Review Criteria**

The discussion below applies an abbreviated version of the review criteria set forth by Besharov, Germanis, and Rossi (henceforth BGR) to the FGA “study” of the Kansas work requirement for ABAWDs. It is followed by a “PC Assessment.” (“PC” is short for “Peter the Citizen.”)

#### ***Program “Theory”***

**BGR:** “Underlying every program’s design is some theory or model of how the program is conceived to work and how it matches the condition it is intended to ameliorate. An evaluation of the program should describe the underlying social problem it is intended to address and how the causal processes described in the model are expected to achieve program goals.

... In assessing the adequacy of an evaluation's program theory, questions such as the following should be raised:

- Is there an adequate description of the underlying social problem the intervention is meant to address?
- Does the intervention make sense in light of existing social science theory and previous evaluations of similar interventions?
- Are the hypothesized causal processes by which the reform effort is intended to achieve its goals clearly stated?
- Have potential unwanted side effects been identified?"

**PC Assessment:** With respect to the first question about the underlying social problem the intervention is meant to address, the authors of the FGA "study" assert: "... food stamps have also become the gateway to long-term government dependency. According to the latest data, the typical enrollee now stays on food stamps for an average of more than eight years."<sup>11</sup> To support this assertion, they cite a study sponsored by the U.S. Department of Agriculture (USDA) that examined the dynamics of SNAP participation from 2008 to 2012.<sup>12</sup> This study does show that the average enrollee, *at a point in time*, received SNAP benefits for eight years, but the FGA report leaves out several important facts.

First, the finding is based on *all* SNAP recipients, including the elderly, disabled, and others who may not be able to work. The findings are not generalizable to able-bodied adults without dependents – the target group for SNAP's current work requirements and the focus of the FGA's "study." Had the FGA "researchers" looked more closely, they would have seen that nearly half of completed spells among nonelderly/nondisabled childless adults were two years or less – a far cry from the eight-year spell length they cite.<sup>13</sup>

Second, the USDA study examines "completed spell lengths" for "SNAP participants in spells underway in December 2008." This approach does not represent the duration of SNAP receipt of the "typical enrollee" over the course of a longer period, as the authors of USDA report explain:

... in the cross-sectional sample, we miss many short spells that occur within the same panel period – they are likely to end before or to begin after our sample month. Longer spells, however, are more likely to include our sample month. For this reason, the longer spells are more heavily represented in the cross-sectional sample than in the entry cohort sample.<sup>14</sup>

The USDA study also shows spell lengths for those who are "new entrants." Nearly 80 percent of nonelderly/nondisabled childless adults who started a spell exited the program within two years; indeed, over a third exited within 4 months. The median spell length was 8 months.<sup>15</sup> A more honest examination of the dynamics of SNAP receipt does not justify the conclusion that the program is a "gateway to long-term government dependency."

Third, the survey covered the 2008 to 2012 period, when the nation was in the midst of a severe economic downturn and jobs were not readily available. Understanding these and other data issues is essential in specifying the size and scope of a policy problem. The FGA presents a

grossly misleading exaggeration of the extent of long-term dependency of the “typical enrollee” as it pertains to a study about the SNAP “work requirement” for adults without dependent children.

The authors of the FGA “study” offer no evidence with respect to the second question, whether the intervention make sense in light of previous evaluations of similar interventions. The reason is that there is no credible evidence regarding the impact of the ABAWDs time limit, though one would expect, based on past research regarding the labor supply effects of benefit payments on work effort that the loss of about \$160 in monthly benefits would have a fairly negligible effect on employment and earnings. They also ignore whether the intervention could have potential unwanted side effects. This is an important omission because the FGA’s own data show that the majority of those who lost their benefits were not employed at any given point in time following exit from the program, suggesting they were adversely affected.

## **Research Design**

**BGR:** “An evaluation’s research design is crucial to its ability to answer, in credible ways, substantive questions about program effectiveness. There are two central issues in research design: (1) ‘*internal validity*,’ or the ability to rule out alternative interpretations of research findings; and (2) ‘*external validity*,’ or the ability to support generalizations from findings to larger populations of interest.”

... In assessing the adequacy of an evaluation's research design, questions such as the following should be asked:

- Are the impact estimates unbiased (internal validity)? How was bias (or potential bias) monitored and controlled for? Were these techniques appropriate?
- Are the findings generalizable to larger populations (external validity)? If not, how does this limit the usefulness of the findings?

**PC Assessment:** Internal validity refers to the extent to which a causal conclusion can be made based on a study, ruling out other possible explanatory factors. External validity refers to the degree to which the findings of a study can be generalized to other populations, settings, times, and programs. All of the “studies” by the FGA suffer from serious problems of both internal validity and external validity. (External validity will be discussed in the section dealing with “Interpretation of Findings” below.)

The authors of the FGA “study” suggest it was a “first of its kind” study. Apparently, they believe that tracking the employment and earnings outcomes of those affected by the policy is something new. They explain:

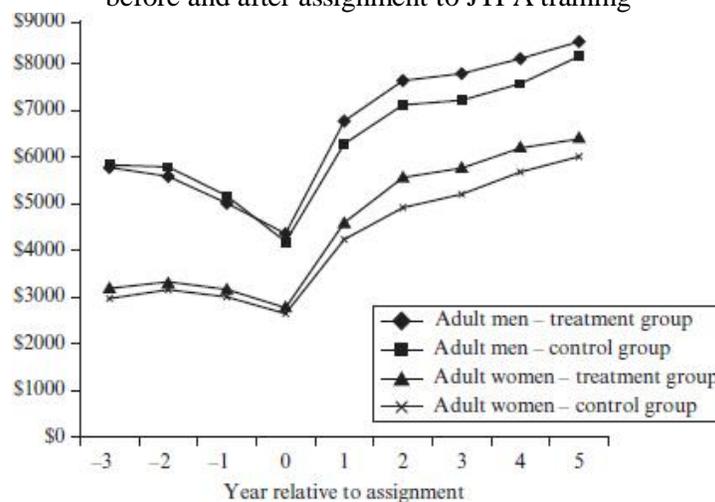
The Brownback administration also set in motion a first-of-its-kind tracking system to monitor the state’s success in moving Kansans from welfare to work. Over the course of two years, the state’s Department of Labor and its Department for Children and Families tracked earnings and employment for nearly 41,000 able-bodied adults leaving food stamps after the work requirements went into effect.

There are hundreds of evaluations that track the employment and earnings of those subject to various welfare reforms – what makes the FGA’s study somewhat unique (though unfortunately not the first of its kind)? The FGA has no comparison or control group that could serve as a counterfactual! The FGA confuses data collection with evaluation. The authors of the study simply assume that any change in the outcomes they observe are the direct result of the “work requirement.” Even worse, they make unfounded claims that may mislead and misguide policymakers.

The most credible approach for assessing the impact of work requirements is to randomly assign individuals to a treatment group that is subject to the requirements and to a control group that is not. The findings from random assignment experiments are considered the most credible, because the treatment and control groups are alike and subject to the same external conditions, with the only difference being the intervention itself. Any difference in outcomes between the groups can be attributed to the intervention – a work requirement – itself. Thus, policymakers could have confidence in whether such requirements actually reduced welfare dependency and poverty by increasing employment and earnings.

To see the need for a credible counterfactual, one can examine the employment and earnings trajectories of the control group from *any* random assignment evaluation of work requirements or training programs. For example, consider the results of the National JTPA Evaluation, as illustrated in Figure 1 below.<sup>16</sup> The vertical axis shows annual earnings and the horizontal axis reflects the time since individuals were randomly assigned, where “year 0” is the year in which individuals were randomly assigned either the treatment group or the control group. For adult men, the annual earnings are shown with a diamond for those in the treatment group and with a square for those in the control group. For adult women, a triangle and cross are used to show the same, respectively.

Figure 1: Annual earnings of adult men and women before and after assignment to JTPA training



Source: Stephen A. Woodbury, “Chapter 17: Unemployment,” in Gerrit de Geest, editor, *Encyclopedia of Law and Economics*, 2009-2017, available at: [https://www.elgaronline.com/view/nlm-book/9781782547457/b2\\_chapter17.xml](https://www.elgaronline.com/view/nlm-book/9781782547457/b2_chapter17.xml).

Before the experiment, the earnings of both treatment and control men and women dropped, which is common in evaluations of social welfare and employment training programs, as individuals who are eligible for and use these programs have often suffered a recent job loss. Many of them regain employment on their own over time. This phenomenon (known as “Ashenfelter’s Dip”) is central to understanding the main problem with the FGA’s approach. Stephen Woodbury, a professor of economics at the University of Wisconsin, explains:

Ashenfelter’s Dip is important because a researcher who ignored it might be tempted to do a ‘before–after’ evaluation of training. For example, comparing the earnings of workers in the treatment group in the year of assignment (year zero, when earnings were about \$4,500) with their earnings two years later (year two, when earnings were about \$7,800) would lead to the conclusion that JTPA increased the earnings of adult men by about \$3,300 a year – an enormous ‘effect’.

Random assignment gives a different and more convincing answer for the impact of JTPA on the earnings of adult men. Comparing the difference between the earnings of the treatment and control groups in the years following the experiment suggests that two years after training that difference was about \$500 and three years after training it was about \$700. The evidence does suggest that JTPA training improved the earnings of adult men, but the estimated effect – \$500 to \$700 a year – is substantially less than the effect suggested by a before–after comparison (\$3,300).<sup>17</sup>

This same concern would be true for able-bodied adults without dependents who enroll in SNAP – many are likely to have just lost a job or had a temporary set-back and their use of SNAP is relatively temporary. Thus, even in the absence of a “work requirement,” many would begin to work anyway.

The frequency of work among this population has been noted by Dorothy Rosenbaum and Ed Bolen of the Center for Budget and Policy Priorities (CBPP) in their critique of the FGA’s reports on this topic:

The reports claiming that terminating SNAP after three months for this population in Kansas and Maine resulted in increased work rates misleadingly present only what happened after the time limit took effect. The reports fail to consider that many SNAP recipients already worked, or would work soon even without the time limit. Looking at the same population before the time limit took effect reveals similar work rates. While challenges exist in getting unemployed childless adults back into the workforce, there’s no evidence that the time limit itself has a significant effect on work.<sup>18</sup>

## **Data Collection**

**BGR:** “Careful data collection is essential for evaluation findings to be credible.

... In assessing the adequacy of an evaluation’s data collection, questions such as the following should be asked:

- Are the data sources appropriate for the questions being studied?
- Are the data complete? What steps were taken to minimize missing data? For example, for survey-based findings, what procedures were used to obtain high response rates?
- Is the sample size sufficiently large to yield precise impact estimates, both overall and for important subgroups?
- Are the data accurate? How was accuracy verified?
- What statistical or other controls were used to correct for potential bias resulting from missing or erroneous data? Were these techniques appropriate?
- What are the implications of missing or erroneous data for the findings?

**PC Assessment:** The FGA “studies” lack transparency in their description of data sources and adjustments to data, but it is crystal clear that many of their claims are based on misleading and inappropriate data comparisons. For example, for claims about employment, the FGA use “Department of Labor” data, *presumably* from the state Unemployment Insurance (UI) data base. This is a common source of data in many evaluations, with well-known strengths and limitations. The main problem with the FGA’s conclusions about employment impacts is the misleading way in which they present the data. For example, the FGA report asserts, “The data shows that the less time these able-bodied adults spend on welfare, the quicker they can get back into the workforce once they are freed from welfare and the more money they will make. These Kansans are discovering new lives of independence and self-sufficiency that, in some cases, they haven’t known for more than two decades.”<sup>19</sup> To support this claim, they add, “Nearly 60 percent found employment within a year of leaving food stamps.”<sup>20</sup>

The employment rate is an important outcome in any welfare reform-related evaluation. Most random assignment evaluations show the employment rate for a baseline quarter (i.e., the quarter of random assignment), and then each quarter thereafter. This shows the status of program participants at a particular point in time and relative to a control group so as to permit accurate estimates of employment impacts. The authors of the FGA report did not do this. Instead, they simply reported the number “ever-employed” through a particular quarter – not the employment rate for a particular quarter.

The first three columns of Table 1 below come from Table 7 of the FGA report. The first column shows the time period and the second column shows the number of able-bodied adults with dependents exiting food stamps in December 2013 (i.e., receiving benefits in that month, but not the next), who have a “record of employment since Q4 2013” (i.e., the baseline quarter). The third column reflects a *cumulative* employment rate, that is, the percent of leavers ever-employed at any time since the fourth quarter of 2013. (The term “cumulative” is added because an employment rate presented this way does not show the percent of leavers employed in any one quarter, but rather the percent employed at any time since a baseline period – the fourth quarter of 2013). The *cumulative* employment rate in the fourth quarter of 2014 is 58.9 percent. This forms the basis of the claim that “nearly 60 percent found employment within a year of leaving food stamps.”

The FGA’s characterization of the employment rate leaves the impression that the SNAP “work requirement” *caused* the increase, because the report provides no reference point. The authors don’t include the baseline employment rate, which would have been the fourth quarter of 2013.

For a pre-post study, this is extremely sloppy and misleading. The fact that nearly 60 percent of leavers “found” employment *might* be impressive if none were employed during the baseline quarter (i.e., the fourth quarter of 2013). In addition, instead of reporting quarterly employment rates, the authors use an employment rate based on any “record of employment since Q4 2013.” Comparing a quarterly employment rate for the baseline (or “pre”) period to an “ever-employed” over the course of a year employment rate (or “post”) period would be comparing apples to oranges. Even with a simple pre-post study, what is needed is a comparison of the quarterly employment rate in the baseline quarter to a quarterly employment rate to a subsequent quarter.

Fortunately, a quarterly employment rate can be calculated from the FGA report, because it provides data on the average quarterly wages of “all disenrollees” as well as those who are “current workers.” The quarterly employment rate can be computed by dividing the second figure by the first figure and is shown in the fourth column of Table 1 below. This calculation shows that 38.9 percent of able-bodied adults without dependents who exited SNAP in December 2013 were employed during the baseline quarter.

<b>Table 1: Comparing Employment Rates: The FGA’s Big Exaggeration</b>			
	<b>Record of employment since Q4 2013</b>	<b>Cumulative Employment Rate</b>	<b>Quarterly Employment Rate (NOT reported by FGA)</b>
Q3 2013			36.4%
Q4 2013	?	?	38.9%
Q1 2014	4,920	38.8%	38.4%
Q2 2014	6,217	48.5%	42.1%
Q3 2014	7,012	54.8%	41.8%
Q4 2014	7,537	58.9%	42.1%
Q1 2015	7,870	61.5%	
Q2 2015	8,252	64.4%	

Source: Table 7 in Jonathan Ingram and Nic Horton, *The Power of Work: How Kansas’ Welfare Reform Is Lifting Americans Out of Poverty* (The Foundation for Government Accountability, February 16, 2016), available at: <https://thefga.org/download/PowerOfWork-KansasWelfareReform.pdf>. The last column is derived from Table 8 of the FGA report.

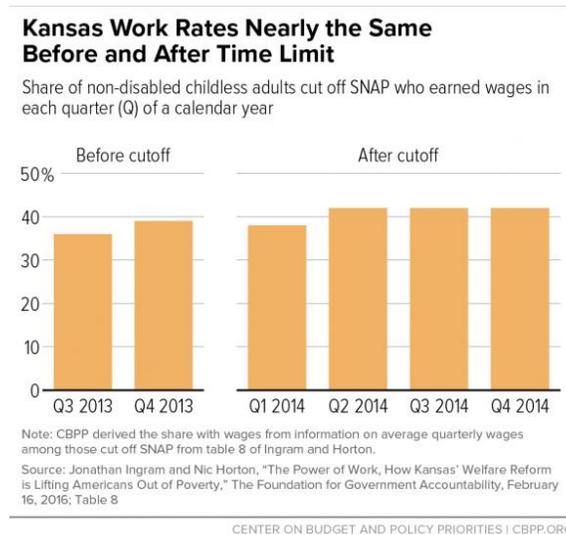
When quarter-by-quarter employment rates are examined, it is obvious that the FGA’s claims are exaggerated. By the fourth quarter, only 42.1 percent of those who exited SNAP in December 2013 were employed. So, even if one believes a simple pre-post study is appropriate, the proper comparison would be 38.9 percent at baseline to 42.1 percent at the time of follow-up a year later – an impact of 3.2 percentage points (8 percent). Does this really justify a claim like, “These Kansans are discovering new lives of independence and self-sufficiency that, in some cases, they haven’t known for more than two decades”?

Dorothy Rosenbaum and Ed Bolen of the CBPP first noted this problem in the FGA “analysis” in December 2016, stating:

... work rates before and after the time limit were very similar... Almost 40 percent of those whose SNAP was cut off already worked in each of the last two quarters before the time limit returned (the third and fourth quarters of 2013). The share working each

quarter in the year after the time limit was implemented rose slightly, to just over 40 percent. This modest increase could be explained by two factors: (1) low-wage workers are more likely to apply for and participate in SNAP when they lose a job or their incomes drop, so they often experience improvements in the future as their employment situation improves; and (2) Kansas' economy was improving over 2014, so a slightly larger share of recipients may have been able to find jobs or higher pay. The time limit does not appear to have dramatically affected work rates for the group subject to it.<sup>21</sup>

The figure below from the CBPP critique illustrates this point.



The authors of the FGA report had this data and they should have presented it. Their depiction of an increasing “ever-employed” rate is highly misleading and is not evidence of the impact of the SNAP “work requirement.”

### Program Implementation

**BGR:** “Key to understanding the success or failure of a program is how well it is implemented. Accordingly, a critical issue in evaluating programs is the degree to which they are implemented in accordance with original plans and the nature and extent of any deviations. Descriptive studies of program implementation are necessary for that understanding and for assessing the program's evaluation.

No matter how well-designed and implemented an evaluation may be, if the program was not implemented well, its impact findings may be of little use for policymaking.

In assessing the adequacy of an evaluation of a program's implementation, questions such as the following should be asked:

- Is the program or policy being evaluated fully described?

- Does the evaluation describe how the policy changes were implemented and operated?
- If defective, how did poor implementation affect estimates of effectiveness?

**PC Assessment:** Any serious evaluation would have included a process study to assess how the state had implemented the “work requirement” – to examine the disability determination process, the activities that were offered, the reasons for participation or non-participation, and perhaps most important to follow-up with a *random* sample of those who exited to find out how they were coping – particularly because the majority of exiters do not seem to have employment in any given quarter. (The report does contain a description of 8 cherry-picked examples of exiters comparing their earnings before and after implementation of the policy, but offers no detail on how those examples were selected or how the data were obtained.)

## Measurement

**BGR:** “Process and outcome variables must have reliable and valid measures. For most evaluations, the principal variables are those measuring program participation, services delivered, and outcomes achieved. An evaluation of a program that attempts to move clients to employment in the private sector clearly needs reliable and valid measures of labor force participation.

... In assessing the adequacy of an evaluation’s process and outcome measures, questions such as the following should be asked:

- Were all appropriate and relevant variables measured?
- Were the measurements affected by response and recall biases? Did subjects misrepresent data for various reasons? Were there Hawthorne effects; that is, did the act of measurement affect the outcome?”

**PC Assessment:** Despite the fact that the FGA is putatively evaluating a work requirement, it provides absolutely no information about how many of those cut-off assistance had a chance to participate in program activities and what services (if any) were provided. Indeed, it is likely the state invested little in work activities and the work requirement was really just a time limit that pushed families off the rolls.

In terms of measuring employment and welfare receipt, the FGA relied on administrative records, which were probably reasonably complete and accurate. (If the employment data came from the state UI system, some types of jobs would be missed, though this would be a factor in both the baseline quarter and any follow-up quarter.) Overall, the FGA’s data collection is limited to a few variables and the “study” provides little background information on the source of the data used and its strengths and weaknesses.

In some studies, rather than relying on actual data, the FGA has extrapolated earnings data to derive its impact estimates. Evaluations should be based on actual data, not extrapolations. For an excellent critique of one such FGA “study,” see:

- Tazra Mitchell, LaDonna Pavetti, and Yixuan Huang, “Study Praising Kansas’ Harsh TANF Work Penalties Is Fundamentally Flawed,” Center on Budget and Policy Priorities, February 20, 2018, available at: <https://www.cbpp.org/research/family-income-support/study-praising-kansas-harsh-tanf-work-penalties-is-fundamentally>.

## Analytical Models

**BGR:** “Data collected in evaluations need to be summarized and analyzed by using statistical models that are appropriate to the data and to the substantive issues of the evaluation. ... Critical characteristics of good analytic models include adequate specification (the variables included are substantively relevant) and proper functional form (the model is appropriate to the statistical properties of the data being analyzed). This is particularly important for quasi-experimental and nonexperimental evaluations.

... In assessing the adequacy of an evaluation's analytical models, questions such as the following should be asked:

- Were appropriate statistical models used?
- Were the models used tested for specification errors?”

**PC Assessment:** The FGA “studies” do not rely on statistical models, despite the fact that there are many factors that could influence the outcomes of interest.

## Interpretation of Findings

**BGR:** “No matter how well analyzed numerically, numbers do not speak for themselves nor do they speak directly to policy issues. An adequate evaluation is one in which the findings are interpreted in an even-handed manner, with justifiable statements about the substantive meaning of the findings. The evaluation report should disclose the limitations of the data analyses and present alternate interpretations.

In assessing the adequacy of an evaluation's interpretation of findings, questions such as the following should be asked:

- When alternative analysis strategies are possible, did the evaluation show how sensitive findings are to the use of such alternatives?
- Are alternative interpretations of the data discussed?
- Are important caveats regarding the findings stated?

**PC Assessment:** The “researchers” at the FGA consistently make irresponsible claims of causality based on an evaluation approach that ignores threats to internal validity; they then make inappropriate generalizations across target populations, programs, and states. Consider the following statement from FGA “researchers” Nic Horton and Jonathan Ingram:

While states like Kansas have done the hard work to prove that work requirements are critical to moving people out of dependency, state policymakers are hamstrung by federal

rules. Current rules prohibit states from expanding work requirements to other welfare programs – such as Medicaid – without first seeking special permission. Even in programs with work requirements, like TANF and food stamps, federal rules restrict who the rules can apply to and thereby hinder states from making the programs as effective as possible.

As the debate over welfare reform heats up in Washington, Congress should learn from state-led welfare reform initiatives and empower state policymakers to incorporate work requirements for able-bodied adults in as many welfare programs as possible. In so doing, they will be giving millions of American families the hand up they desperately need.<sup>22</sup>

By using the word “prove,” they are stating that their research clearly demonstrates the effectiveness of work requirements. Given the magnitude of caseload reduction, if their sole point were about cutting caseloads and costs, they would be right – there are few plausible alternative explanations. But, the very title of their original Kansas paper suggests far more, stating that the requirements lift people out of poverty – “How Kansas’ Welfare Reform Is Lifting Americans Out of Poverty.” And, within the report, they state “Kansas’ welfare reform has led to greater employment rates, higher incomes, and more hours worked for those adults who still depend on food stamps.” Without a credible counterfactual, there is no basis for these claims.

Even if the Kansas findings were valid (and they most definitely are not), it would be inappropriate to generalize (a problem of external validity) from one state (Kansas) to all other states or from one program (SNAP) to other programs like TANF and Medicaid or from one population (able-bodied adults between 18-49 without dependents) to other populations like older adults or adults with dependents. Yet, this is exactly what the FGA “researchers” do. Given the fact that there is absolutely no empirical evidence regarding the potential impact of their proposals to expand work requirements this way, the responsible approach would be to test the proposals and evaluate them rigorously with a random assignment experiment.

<b>Evaluating Welfare Reform: A Report Card for the FGA</b>		
<b>Criteria</b>	<b>Grade</b>	<b>Comment</b>
<b>Program Theory</b>	<b>F</b>	Does the program or evaluation make sense in light of existing social science knowledge?  FGA “studies” do not provide an adequate description of the problem being addressed, explain the hypothesized causal process, or identify potential unintended effects.
<b>Research Design</b>	<b>F</b>	Does the research design have both internal and external validity?  FGA “studies” do not use a research design that permits statements about causal effects; they lack internal and external validity.
<b>Program Implementation</b>	<b>F</b>	How does the program actually operate?  FGA “studies” provide virtually no description of how programs or policies are implemented.
<b>Data Collection</b>	<b>F</b>	Is the necessary data available and reliable?  FGA “studies” typically rely on a small number of outcomes and often lack transparency in how they were collected; many data comparisons are misleading (i.e., comparing apples and oranges).
<b>Measurement</b>	<b>F</b>	Are the key variables valid and can they be measured reliably?  FGA “studies” do not provide an adequate description of how data were collected and sometimes involve inappropriate adjustments or extrapolations;
<b>Analytical Models</b>	<b>F</b>	Are the data summarized and analyzed by means of appropriate statistical models?  FGA “studies” are not real evaluations; they make no attempt to adjust for potential threats to validity using conventional evaluation methods and statistical tools.
<b>Interpretation of Findings</b>	<b>F</b>	Are the findings interpreted objectively and do they describe the limitations of the analyses and consider alternative interpretations? Are they placed in the proper policy or programmatic context?  FGA “studies” are replete with erroneous and misleading statements; their conclusions are irresponsible, particularly claims of causality.
<b>Overall Grade</b>	<b>F</b>	FGA “studies” are not credible and should not be used by policymakers (or anyone else, for that matter, with the possible exception of those who teach evaluation – to illustrate how NOT to perform an evaluation).

## Appendix

The “research” by the FGA suffers from multiple problems beyond lacking a credible counterfactual. The most detailed critiques have been undertaken by “Peter the Citizen” (a conservative) and the Center on Budget and Policy Priorities (often described as a liberal organization). With respect to evidence-based policymaking, the starting point should be facts.

### Papers by “Peter the Citizen”

“Maine DOES NOT Show How to Make Welfare Work: A Response to Jared Meyer and Mary Mayhew,” June 4, 2017, available at: <http://mlwiseman.com/wp-content/uploads/2017/06/Meyer.pdf>.

“The FGA’s ‘First of Its Kind Study’ Should Have Been the Last: An Evaluation Note for Pre-Post Conservatives,” July 4, 2017, available at: [http://mlwiseman.com/wp-content/uploads/2016/05/The-FGA.FF\\_.pdf](http://mlwiseman.com/wp-content/uploads/2016/05/The-FGA.FF_.pdf).

“Using Squirrely Data is No Way to Justify Conservative Policies: A Note to AEI’s Marc Thiessen,” July 8, 2017, available at: <http://mlwiseman.com/wp-content/uploads/2016/05/Thiessen.pdf>.

“‘The Truth About Welfare Reform’ Won’t be Found at the Foundation for Government Accountability: A Response to Jonathan Ingram,” August 31, 2017, available at: <http://mlwiseman.com/wp-content/uploads/2016/05/Ingram.pdf>.

“The FGA’s Simplistic, Misleading, and Irresponsible Report on Work Requirements in Kansas: A Sentence-by-Sentence Critique,” September 2, 2017, available at: [http://mlwiseman.com/wp-content/uploads/2016/05/FGA.TANF\\_.KS\\_.pdf](http://mlwiseman.com/wp-content/uploads/2016/05/FGA.TANF_.KS_.pdf).

“A Note to Pre-Post Conservatives: You Are Not Fooling Anyone – Except Maybe Yourself (and Some Politicians),” September 2, 2017, available at: <http://mlwiseman.com/wp-content/uploads/2016/05/PrePost-Conservatives.0904.pdf>.

“The Importance of a *Credible* Counterfactual: An Explanation for Tarren Bragdon and the FGA,” May 20, 2018, available at: <http://mlwiseman.com/wp-content/uploads/2016/05/The-FGA.Counterfactual.pdf>.

### Papers by the CBPP (including critiques of FGA’s “analyses” of TANF work requirements based on the same fundamentally flawed methodology)

Dottie Rosenbaum and Ed Bolen, “SNAP Reports Present Misleading Findings on Impact of Three-Month Time Limit,” December 14, 2016, available at: <https://www.cbpp.org/research/food-assistance/snap-reports-present-misleading-findings-on-impact-of-three-month-time>.

Tazra Mitchell, “Most Kansas Families After TANF: Unsteady Work, Extremely Low Earnings,” January 24, 2018, available at: <https://www.cbpp.org/blog/most-kansas-families-after-tanf-unsteady-work-extremely-low-earnings>.

Tazra Mitchell, LaDonna Pavetti, and Yixuan Huang, “Study Praising Kansas’ Harsh TANF Work Penalties Is Fundamentally Flawed,” February 20, 2018, available at: <https://www.cbpp.org/research/family-income-support/study-praising-kansas-harsh-tanf-work-penalties-is-fundamentally>.

Tazra Mitchell, LaDonna Pavetti, and Yixuan Huang, “Life After TANF in Kansas: For Most, Unsteady Work and Earnings Below Half the Poverty Line,” February 20, 2018, available at: <https://www.cbpp.org/research/family-income-support/life-after-tanf-in-kansas-for-most-unsteady-work-and-earnings-below>.

Tazra Mitchell, “Some House Leaders Ignore Evidence, Cite Flawed Reports to Justify Taking Basic Assistance Away From Needy Individuals,” April 19, 2018, available at: <https://www.cbpp.org/research/poverty-and-inequality/some-house-leaders-ignore-evidence-cite-flawed-reports-to-justify>.

---

<sup>1</sup> The views in this document reflect my own as a citizen and do not reflect the views of any organization I am now or have ever been affiliated with. By way of background, I am a conservative and have worked on welfare issues for the Heritage Foundation, the American Enterprise Institute, and the White House under both President Reagan and President George H.W. Bush.

<sup>2</sup> Caitlin Dewey, “They’re the think tank pushing for welfare work requirements. Republicans say they’re experts. Economists call it ‘junk science,’” *The Washington Post*, May 18, 2018, available at: [https://www.washingtonpost.com/news/wonk/wp/2018/05/18/theyre-the-think-tank-pushing-for-welfare-work-requirements-republicans-say-theyre-experts-economists-call-it-junk-science/?utm\\_term=.09384a7121c2](https://www.washingtonpost.com/news/wonk/wp/2018/05/18/theyre-the-think-tank-pushing-for-welfare-work-requirements-republicans-say-theyre-experts-economists-call-it-junk-science/?utm_term=.09384a7121c2).

<sup>3</sup> Caitlin Dewey, “They’re the think tank pushing for welfare work requirements. Republicans say they’re experts. Economists call it ‘junk science,’” *The Washington Post*, May 18, 2018, available at: [https://www.washingtonpost.com/news/wonk/wp/2018/05/18/theyre-the-think-tank-pushing-for-welfare-work-requirements-republicans-say-theyre-experts-economists-call-it-junk-science/?utm\\_term=.09384a7121c2](https://www.washingtonpost.com/news/wonk/wp/2018/05/18/theyre-the-think-tank-pushing-for-welfare-work-requirements-republicans-say-theyre-experts-economists-call-it-junk-science/?utm_term=.09384a7121c2).

<sup>4</sup> Douglas J. Besharov, Peter Germanis, Peter H. Rossi, *Evaluating Welfare Reform: A Guide for Scholars and Practitioners* (College Park, MD: University of Maryland School of Public Affairs, 1997), available at: <http://www.welfareacademy.org/pubs/welfare/ewr/ewr.pdf>.

<sup>5</sup> *Ibid.*

<sup>6</sup> Jeffrey Grogger, Lynn A. Karoly, and Jacob Alex Klerman, *Consequences of Welfare Reform: A Research Synthesis*, RAND, Prepared for the Administration for Children and Families, U.S. Department of Health and Human Services, July 2002, available at: [https://www.acf.hhs.gov/sites/default/files/opre/consequences\\_of\\_welfare\\_reform.pdf](https://www.acf.hhs.gov/sites/default/files/opre/consequences_of_welfare_reform.pdf).

<sup>7</sup> Nic Horton, Jonathan Ingram, and Josh Archambault, “First-Of-Its-Kind Study Shows The Power Of Work,” *Forbes*, February 19, 2016, available at: <https://thefga.org/wp-content/uploads/2016/12/Forbes-First-Of-Its-Kind-Study-Shows-The-Power-Of-Work.pdf>. The study they refer to is: Jonathan Ingram and Nic Horton, *The Power of Work: How Kansas’ Welfare Reform Is Lifting Americans Out of Poverty* (The Foundation for Government Accountability, February 16, 2016), available at: <https://thefga.org/download/PowerOfWork-KansasWelfareReform.pdf>.

<sup>8</sup> To maintain consistency with the terminology in the FGA report, I will refer to the provision as a “work requirement,” but use quotation marks to emphasize how it has been implemented in practice.

<sup>9</sup> Nic Horton, Jonathan Ingram, and Josh Archambault, “First-Of-Its-Kind Study Shows The Power Of Work,” *Forbes*, February 19, 2016, available at: <https://thefga.org/wp-content/uploads/2016/12/Forbes-First-Of-Its-Kind-Study-Shows-The-Power-Of-Work.pdf>.

<sup>10</sup> Jonathan Ingram and Nic Horton, *The Power of Work: How Kansas’ Welfare Reform Is Lifting Americans Out of Poverty* (The Foundation for Government Accountability, February 16, 2016), p.5, available at: <https://thefga.org/download/PowerOfWork-KansasWelfareReform.pdf>.

<sup>11</sup> Jonathan Ingram and Nic Horton, *The Power of Work: How Kansas’ Welfare Reform Is Lifting Americans Out of Poverty* (The Foundation for Government Accountability, February 16, 2016), p.5, available at: <https://thefga.org/download/PowerOfWork-KansasWelfareReform.pdf>.

<sup>12</sup> Joshua Leftin, Nancy Wemmerus, James Mabli, Thomas Godfrey, and Stephen Tordella, *Dynamics of Supplemental Nutrition Assistance Program Participation from 2008 to 2012* (Arlington, VA: Decision Demographics, December 2014), available at: <https://fns-prod.azureedge.net/sites/default/files/ops/Dynamics2008-2012.pdf>.

<sup>13</sup> Joshua Leftin, Nancy Wemmerus, James Mabli, Thomas Godfrey, and Stephen Tordella, *Dynamics of Supplemental Nutrition Assistance Program Participation from 2008 to 2012* (Arlington, VA: Decision Demographics, December 2014), p.91, available at: <https://fns-prod.azureedge.net/sites/default/files/ops/Dynamics2008-2012.pdf>.

<sup>14</sup> Joshua Leftin, Nancy Wemmerus, James Mabli, Thomas Godfrey, and Stephen Tordella, *Dynamics of Supplemental Nutrition Assistance Program Participation from 2008 to 2012* (Arlington, VA: Decision Demographics, December 2014), p. 76, available at: <https://fns-prod.azureedge.net/sites/default/files/ops/Dynamics2008-2012.pdf>.

<sup>15</sup> Joshua Leftin, Nancy Wemmerus, James Mabli, Thomas Godfrey, and Stephen Tordella, *Dynamics of Supplemental Nutrition Assistance Program Participation from 2008 to 2012* (Arlington, VA: Decision Demographics, December 2014), p. 71, available at: <https://fns-prod.azureedge.net/sites/default/files/ops/Dynamics2008-2012.pdf>.

---

<sup>16</sup> This example comes from Stephen A. Woodbury, “Chapter 17: Unemployment,” in Gerrit de Geest, editor, *Encyclopedia of Law and Economics*, 2009-2017, available at: [https://www.elgaronline.com/view/nlm-book/9781782547457/b2\\_chapter17.xml](https://www.elgaronline.com/view/nlm-book/9781782547457/b2_chapter17.xml). Notably, “the treatment and the control groups had essentially similar earnings in the three years leading up to the experiment, which suggests that randomization of workers was successful in all four cases.”

<sup>17</sup> Stephen A. Woodbury, “Chapter 17: Unemployment,” in Gerrit de Geest, editor, *Encyclopedia of Law and Economics*, 2009-2017, available at: [https://www.elgaronline.com/view/nlm-book/9781782547457/b2\\_chapter17.xml](https://www.elgaronline.com/view/nlm-book/9781782547457/b2_chapter17.xml).

<sup>18</sup> Dorothy Rosenbaum and Ed Bolen, “SNAP Reports Present Misleading Findings on Impact of Three-Month Time Limit,” Center on Budget and Policy Priorities, December 14, 2016, available at: <http://www.cbpp.org/research/food-assistance/snap-reports-present-misleading-findings-on-impact-of-three-month-time>.

<sup>19</sup> Jonathan Ingram and Nic Horton, *The Power of Work: How Kansas’ Welfare Reform Is Lifting Americans Out of Poverty* (The Foundation for Government Accountability, February 16, 2016), p. 2, available at: <https://thefga.org/download/PowerOfWork-KansasWelfareReform.pdf>.

<sup>20</sup> Jonathan Ingram and Nic Horton, *The Power of Work: How Kansas’ Welfare Reform Is Lifting Americans Out of Poverty* (The Foundation for Government Accountability, February 16, 2016), p. 13, available at: <https://thefga.org/download/PowerOfWork-KansasWelfareReform.pdf>.

<sup>21</sup> Caitlin Dewey, “They’re the think tank pushing for welfare work requirements. Republicans say they’re experts. Economists call it ‘junk science,’” *The Washington Post*, May 18, 2018, available at: [https://www.washingtonpost.com/news/wonk/wp/2018/05/18/theyre-the-think-tank-pushing-for-welfare-work-requirements-republicans-say-theyre-experts-economists-call-it-junk-science/?utm\\_term=.09384a7121c2](https://www.washingtonpost.com/news/wonk/wp/2018/05/18/theyre-the-think-tank-pushing-for-welfare-work-requirements-republicans-say-theyre-experts-economists-call-it-junk-science/?utm_term=.09384a7121c2).

<sup>22</sup> Nic Horton and Jonathan Ingram, “Work Requirements are Working for Kansas Families How welfare reform increases incomes and improves lives,” The Foundation for Government Accountability, July 31, 2017, p. 11, available at: <https://thefga.org/wp-content/uploads/2017/07/Work-Requirements-are-Working-for-Kansas-Families.pdf>.