Response Document for Ms. No. 2023124 "*Sliding into Safety Net Participation: A Unified Analysis Across Multiple Programs*"

Derek Wu and Jonathan Zhang

Table of Comments **Page**

A. Summary of Key Changes

We have carefully read through the comments of the editor and referees and have taken each comment seriously and made suitable changes. We thank the editor and both referees for their thoughtful and helpful comments and believe the paper is greatly improved. Below, we summarize the key changes to the paper. We then respond to each comment point-by-point, in detail.

1. *<u>Comprehensive literature review and relative contribution</u>*
   In response to Editor's comment 1, Referee 1's comment 1, and Referee 2's comment 3 and 4, we have included an extensive literature review and succinctly situated our contribution in this literature. The key changes are:
   a) In the introduction:
   - *Expanding the first paragraph (p. 1) to better contextualize the literature that precedes our paper:*
     > *"Most of the systematic evidence comes from studies of Aid to Families with Dependent Children (AFDC) prior to welfare reform, when the safety net was more homogeneous and narrowly targeted. Yet, we know relatively little about the nature and pace of changes in household circumstances preceding program receipt in the modern (post-welfare reform) era, and specifically how these patterns differ across programs."*
   - *A succinct paragraph at the end of the introduction (p. 3) briefly outlining the literature and an overview of our contribution with the following sentence:*
     > *"Using a unified data source, we contribute to these literatures by documenting the trajectory of a diverse set of household characteristics…around the receipt of a wide range of U.S. programs in recent years."*

   b) A new 3+ page Section 2 that summarizes the prior literature and our relative contribution.
   - *Section 2.1 summarizes the prior literature (including references kindly suggested by Referee 1, along with others we surveyed), examining topics ranging from "Ashenfelter Dip" and correlates of welfare participation and poverty spells, to events that precede or predict welfare entry/exit. When*

discussing these papers, we are careful to point out defining features that are improved upon in our paper, "priming" the reader for the next subsection.

- *Section 2.2 details our contributions relative to the existing literature. Briefly, they are: i) focusing on contributing events leading up to program participation in the modern (post-welfare reform) era, which has changed drastically since pre-welfare reform; ii) providing a unified framework for comparing and contrasting patterns across multiple programs for the entire U.S., and assessing changes in characteristics like disability and marital separation that have been largely overlooked in more recent non-AFDC studies; iii) looking at quarterly changes in characteristics, allowing us to paint a richer and more complete picture of the pace and nature of events preceding program enrollment (as the older literature tends to examine "any change" over some time period); and iv) making comparisons exclusively within a participating household over time, whereas some prior studies look at patterns for both participants and non-participants.*

c) Emphasizing the difference between the modern-day safety net and pre-welfare reform safety net (i.e., our contribution) by discussing trends over time in benefit amounts for key programs–showing how TANF has fallen precipitously while other programs like SNAP have grown tremendously (new Figure A1 and discussion in Section 3, p. 7-8). This is also directly in response to Referee 2's comment 5.

2. *<u>Changes to baseline sample</u>*
   In response to the Editor's comment 2, and Referee 2's comment 1, we have changed our main baseline sample and included robustness checks with varying sample definitions. Our new baseline sample is significantly broadened, requiring only 20-24 consecutive months of SIPP participation and at least one month of data prior to program participation (previously, we required a balanced panel of at least 36 months of SIPP participation and 16 months of data pre- and post-enrollment). This increases our sample size significantly (e.g., from 4,835 Medicaid households to 19,403 and 590 TANF households to 2,721) and improves precision for many of our outcomes. Our main findings are qualitatively unchanged. We also consider different sample specifications: requiring 2 quarters and 4 quarters pre-enrollment, as well as 2 and 4 quarters both pre- and post-enrollment. With this larger sample, we also modify our empirical strategy to include household-fixed effects (previously cohort-fixed effects), which allows for a more robust examination of changes ***within*** *a household over time.*

3. *<u>Extensive robustness checks in a new subsection</u>*
   Relatedly, the new Section 6.2 now provides additional robustness checks using alternate specifications, varying sample restrictions (e.g., observing households for continuous periods pre-/post-receipt), and dropping the most recent 2014 and 2018 SIPP panels which use a recall period of 12 months as opposed to 4 months previously (in response to a suggestion by Referee 1). Our main findings remain remarkably robust.

Below, we have copied the specific comments from the editor and referees. Our responses are italicized and in blue.

B. Editor's Comments

Thank you for your submission to the National Tax Journal of your paper "Sliding into Safety Net Participation: A Unified Analysis Across Multiple Programs." I sent this paper out to two expert reviewers. Both reviewers find the topic to be very important, but both raise substantial concerns about the paper in its current form and come to somewhat different conclusions, with one referee recommending the paper be rejected and the other recommending a revise and resubmit. I have read the paper closely and think the paper is important and may be an excellent contribution to the NTJ. However, I agree with the referees that there are substantial issues in the current draft that don't allow me to commit to a revise and resubmit at this time. Therefore, I am rejecting the paper. That said, if you are able to make the changes I outline below, I'm willing to consider a new submission of this paper and intend to send it back to the same referees. If you do take me up on my reject & resubmit offer, please mention this manuscript number and provide a response to each comment made by me and the referees. Here are the key issues for you to address:

1. Connect your paper to the existing literature – This is the core reason your paper is being rejected now. As both reviewers note, you do not connect your paper well to the large existing literature on the topic and this means you are not able to make a compelling case for the specific contribution of your paper. It seems to be both about missing citations as well as not being precise/detailed enough about the specific contribution of your work relative to the literature you do cite. Until these issues are addressed, it is not possible for me to assess whether this paper makes a contribution that is appropriate for NTJ. This must be fully resolved in a revised version; please see the referees' specific comments on this.

This comment—related to Referee 1's comment 1 and Referee 2's comment 3 and 4—is one that we took the most seriously in this revision. We have made some major changes including a concise overview of the literature and our contribution in the introduction, followed by a new Section 2 entirely dedicated to the prior literature as well as our contribution. This new section is warranted given the plethora of important work that came before our paper. In Section 2, we also cite the relevant papers brought to our attention by Referee 1. Details of our changes are described above in Section A. To briefly describe our main contributions:

> *i) We study the safety net in the modern era. Much of the previous literature focuses on the context of AFDC, which targets only low-income families with children (often single mothers) and is very different from the diverse array of programs we have today. While some of these older studies focus directly on the events triggering AFDC receipt, there are also emphases on other issues—such as comparisons of static characteristics between participants and non-participants and what is correlated with getting off welfare or out of poverty. We are making comparisons between time periods within participants and looking at entry into programs.*
>
> *ii) Within this contemporary context, we use a unified framework to compare and contrast patterns around multiple programs for the entire United States. More recent studies tend to focus on a single program, a single state, and/or earnings/employment outcomes solely. We move beyond labor supply and can assess outcomes that are largely ignored by the studies using data after 2000.*

iii) We also study granular changes in outcomes at the quarterly level to paint a richer picture of the pace and nature of events compared to analyzing "any event" over some time period (as many of the prior studies do). This is in part driven by trends in social science to depict panel data with "event study"-type figures, which are more conducive for studying dynamic patterns.

2. Sample Selection – Referee 2's first two comments raise questions about sample selection that seem potentially important for the conclusions that you draw, please address these.

We have now entirely changed our baseline sample, requiring only 20-24 months of continuous SIPP participation (as opposed to 36 months) and only one month of data prior to enrollment. We have also implemented a robustness check requiring 2 quarters of observations pre- and post-enrollment, which was suggested by Referee 2's comment 1, in addition to four other sample restriction definitions. We have also addressed two related comments by Referee 1 about controlling for seasonality (in main results) and seam/recall bias in the 2014 and 2018 SIPP by excluding the last two panels (in robustness checks). The results of the robustness checks can be found in Figure A15 and are discussed in the new Section 6.2.

3. Other Comments – The referees make many other useful points. Please engage with each of them. Their reports are attached.

Below we carefully respond to both referees' comments and concerns, point-by-point.

C. Referee 1 Comments

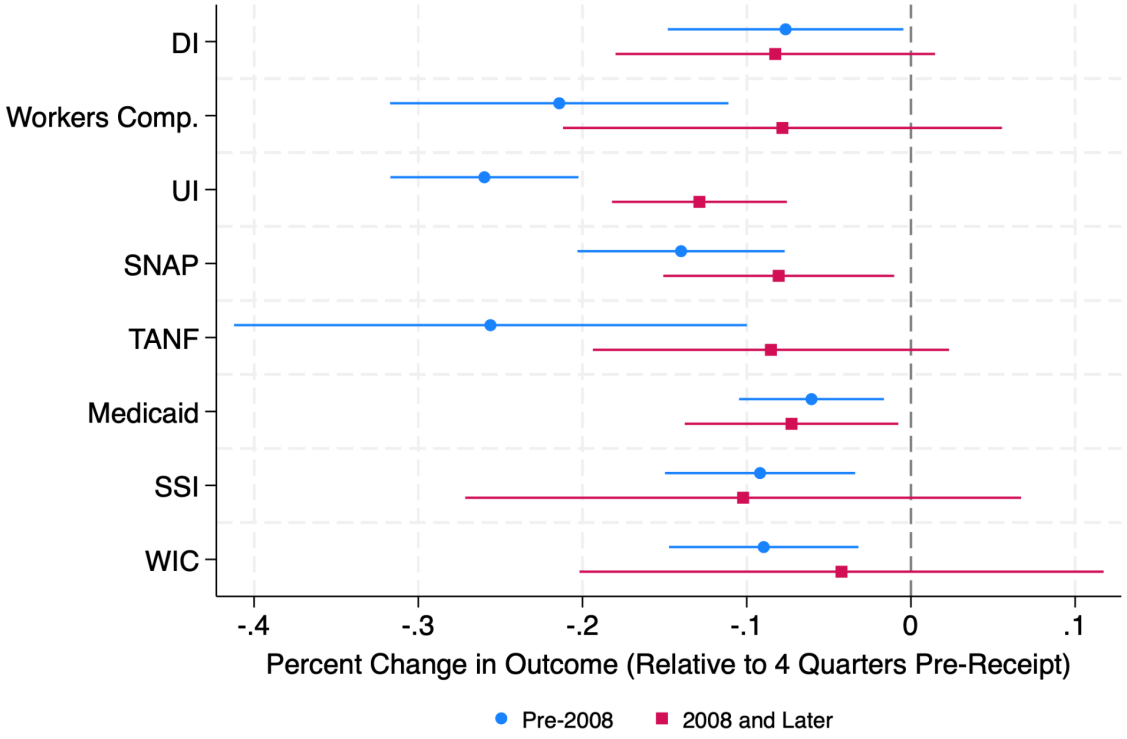
This paper uses multiple panels of the SIPP along with an administrative file on the Veterans Health program to examine the events that trigger spells of welfare and social insurance participation. The paper is descriptive, finding associations between job loss, marital disruption, children, and health and disability shocks as key events leading up to participation in transfer programs.

While the topic of the trigger events leading to participation in welfare and social insurance programs is of clear research and policy interest, the paper in its current form does not make a substantive contribution to the current literature. The paper is either dismissive of the findings from the literature (e.g. just mention of Bane and Ellwood (1986), and relegating papers like Blank and Ruggles (1996), Grogger (2004), Ham et al. (2016) to a slight mention in a footnote), or completely ignores well known papers like Huff Stevens (1999 JHR); Hoynes (2000 RESTAT); Ribar, Edelhoch, and Liu (JHR 2008) and other less well known papers but no less relevant (Boisjoly, Harris, and Duncan 1998; Gittleman 2001; McKernan and Ratcliffe 2005; Cellini, McKernan, and Ratcliffe 2008). The conclusions from this paper mirror those from the prior literature, so it is not clear what is new. It is true that much of that literature is older, but the fact that the authors skipped over recognizing the prior work means that they are unable to situate their findings. I think it is time for a look again at these programs, but the paper is missing a hook. All data are post welfare reform, so there is no pre-post welfare reform story. One option would be to add earlier panels of the SIPP to examine trigger events after welfare reform compared to before. Or barring that, another option would be to examine whether and how these associations changed after the Great Recession. At a minimum a much more comprehensive literature review is needed and how the paper contributes to said literature.

Thank you for this very helpful comment. We agree with you that the previous version was not very clear about our contribution to the existing literature. We have made major changes and believe the new manuscript is much improved. A summary of our changes can be found above in "Summary of Key Changes" and the response to Editor's comment 1, but below we have included some specific responses unique to your comment.

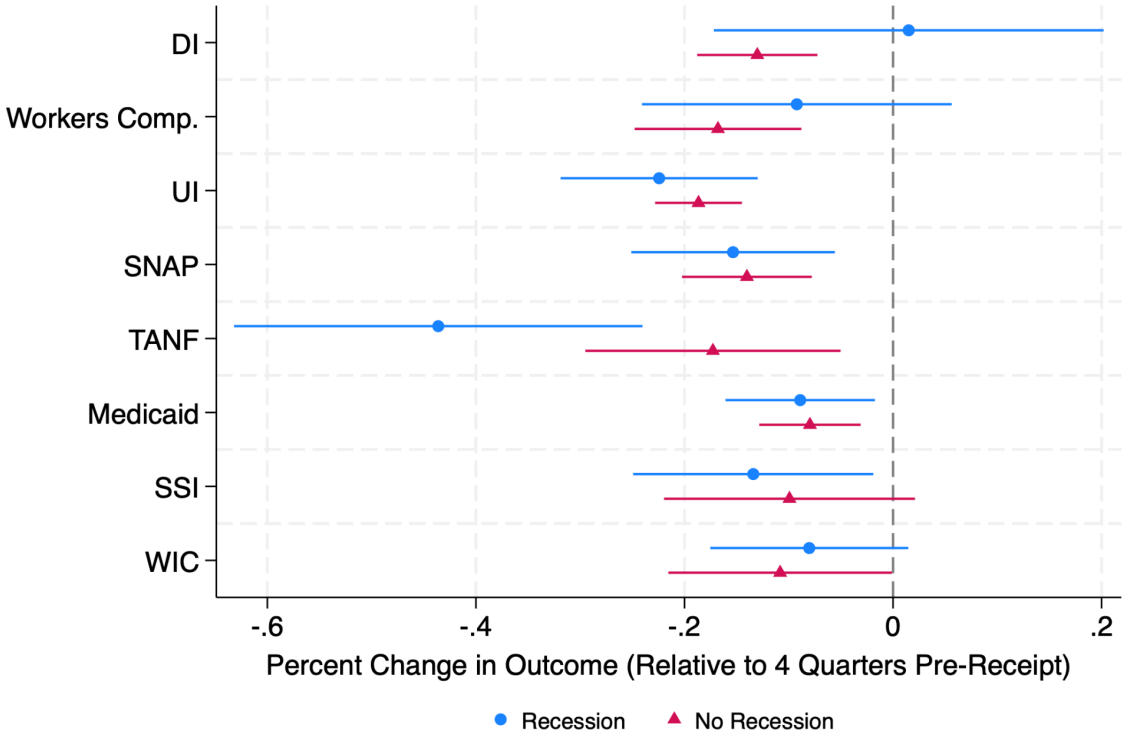
- *We have now read and cited the relevant papers you listed above (most in Section 2.1, some in the introduction), as well as additional relevant papers.*
- *Your suggestion about post-welfare reform is now reflected as one of our key contributions. As you mention, the prior literature focuses on pre-welfare reform (and mainly AFDC), where the safety net mainly targeted low-income families with children. Moreover, some of these studies are concerned with exits from welfare (whereas we are focused on entry) or differences between individuals on and not on welfare (while we are comparing across time periods within an individual).*
- *We also liked your suggestion on looking at whether the factors preceding program enrollment changed after the Great Recession. Figure A below shows changes to market income for various programs, split before and after the Great Recession. Figure B shows the same figure but for any recession vs non-recession. In both scenarios, we do not see any obvious patterns–especially with the wide confidence intervals. Therefore, we have decided to not pursue this further. However, if you feel differently, please do let us know.*

***Figure A**: Changes to Market Income Before and After the Great Recession*



Notes: This figure plots the analog of Figure 2 (in Wu and Zhang 2024), splitting the sample into two time periods, prior to the Great Recession of 2008 and after.

***Figure B**: Changes to Market Income By Whether Enrollment was During a Recession*



Notes: This figure plots the analog of Figure 2 (in Wu and Zhang 2024), splitting the sample into during any recession and during a non-recessionary period.

Specific Comments

* I appreciate the examination of a broad array of programs, and this is a contribution as most papers only look at one or a small number. But the inclusion of the VA DC program does not fit in with the others. It is true that the program is very large in terms of expenditure, but the population served is unique and lacks to broader scope of the others in the paper. Moreover, the use of the administrative data does not fit in well with the SIPP because of the differences in populations served, and sample frame of the SIPP. Thus I recommend dropping the VA DC analysis from the paper.

Thank you for this comment. We agree with you that the VA DC program is only relevant for a specific subpopulation and is less general than the other programs examined using the SIPP. We also believe that the findings from the administrative VA data with health and prescription records are consistent with our disability findings in the SIPP. We have now moved the discussion of the VA DC program, the data, and results to Appendix B. This new appendix is entirely self-contained. What used to be Figure 5 (in the previous draft) is now Appendix Figure B2. VA DC is not mentioned as a studied program anymore (i.e., we now study 8 instead of 9 programs in the main text).

Finally, we have significantly de-emphasized the VA findings in the main text. VA DC is mentioned in only two places in the main text. At the end of Section 4.2 "Outcomes Analyzed" (p. 13):

> *"... we are able to examine how health outcomes change around take-up of one particular government program, Veterans' Disability Compensation (VA DC), using administrative health records for all VA DC applicants between 2004 and 2020. Because VA DC lacks the broader scope of the other programs we examine, we leave the details of this analysis to Appendix B."*

The second place is at the end of Finding #2 (p. 20):

> *"Taken together, these findings suggest that the increases in non-employment rates preceding program receipt may reflect higher rates of disability. Using administrative health records linked to application data for another program (VA DC), we find evidence that health shocks (e.g., inpatient hospitalizations and emergency department visits) and fills of prescription painkillers spike immediately in the 3-4 months prior to application. Given that we also find statistically significant increases in non-employment and disability leading up to VA DC receipt in the SIPP, this provides suggestive evidence that the larger changes in disability observed around the receipt of other programs are related to important changes in health. The findings with regard to VA DC are discussed further in Appendix B."*

* On pp. 4-5 you claim an advantage of programs like SNAP, Medicaid, and UI is the lack of age requirement. The nuances of Medicaid eligibility (and benefit generosity) do vary widely

across states based on age of children, and as the authors know, childless adults were not eligible prior to the ACA in 2010, in other words over most of the sample period. Thus this claim about not age dependent needs some qualification.

Thank you for pointing out this qualification. We have removed the part of the sentence (previously on page 4) that states "with eligibility conditional neither on demographic characteristics nor disability." We also note in footnote 6 (p. 7) that state-specific eligibility requirements often vary by age and in Section 3.1 (p. 8) that eligibility requirements for Medicaid have changed significantly since the ACA.

* On the SIPP you need to provide the reader some important detail about changes in the sample starting in 2014. Earlier waves were conducted 3 times a year with a 4-month recall for program participation. Starting in 2014 that recall is for 12 months. This means the old rule for handling seam bias (e.g. just look at wave 1) may no longer hold. I believe that there was a NAS panel that reviewed the 2014 SIPP, finding bunching of program receipt in January of the year. If true, then does the selection of once per quarter still work for handling seam bias? This needs to be clarified.

Thank you for emphasizing this issue. We discuss this concern in footnote 9 (p. 11) of Section 4.1 "Data Sources":

> *"For most of its history (i.e., prior to 2014), the SIPP conducted interviews three times during a 12-month period, yielding a four-month recall period for monthly program participation. Starting with the 2014 Panel, however, the SIPP underwent a redesign that---among other changes---altered the interview frequency to once per year, thereby extending the recall period to twelve months. In evaluating the quality of the 2014 panel, NAS (2018) noted a number of data quality issues associated with the redesign, including greater under-reporting of government programs and increased bunching of responses at the beginning of the year. While our main estimates incorporate these more recent SIPP panels, we also calculate estimates excluding them as a result of these data quality concerns."*

We also note in footnote 14 (p. 15) of Section 5 "Empirical Methods" (referring to the choice of pooling estimates across months within a quarter):

> *"...Note that this approach would not be able to account for seam bias as adequately in the 2014 and 2018 Panels, where the reference periods are longer than a quarter. We therefore conduct robustness checks where we limit our sample to the 1996-2008 Panels."*

The robustness check dropping the two most recent SIPP Panels (2014 and 2018) can be found in Figure A15. Changes in market income around program receipt remain remarkably similar. Changes for Medicaid enrollees are slightly attenuated without the recent panels and the standard errors are also somewhat smaller.

* On p. 9 you state that you "link" electronic health records. The way this paragraph is written it almost seems like you link the VA data to the SIPP, but I don't think this is

happening. Normally I would recommend that this text be clarified, but as per my prior suggestion, I think the whole VA analysis should be cut because the population served is so different from the other programs.

Following your comment from earlier, we have now moved the VA results to the appendix. There, we have clarified the language around "link" to refer to administrative health records in the VA "linked" with VA DC application records.

* On p. 10, why only focus on 4 main demographics? This is where a review of the prior literature would be very helpful in order to understand how you selected these variables.

We focus on changes in four main characteristics: employment, disability, having a young child, and being separated. We select them because they reflect key time-varying variables that are observed monthly. There are other characteristics that are time-invariant (e.g., race) or evolve in deterministic patterns (e.g., age), and we do investigate heterogeneity along some of these characteristics in Section 6.2 "Robustness Checks and Extensions".

There are also other well-being measures in the SIPP, such as assets, savings, and living conditions that are time-varying, but are not routinely surveyed (i.e., only in specific topical modules). While older studies on AFDC and poverty triggers analyze a variety of events analogous to what we examine, more recent studies primarily focus on changes in employment and earnings.

We have added the following paragraph to Section 4.2 (p. 13):

> *"We choose these four characteristics because they reflect important life transitions that may be associated with economic vulnerability. They are also recorded in the SIPP at the monthly level, and correspond to the same individual-level events that McKernan and Ratcliffe (2005) find are associated with entries into poverty. While the SIPP does contain information on other characteristics---such as material well-being and self-reported health---that we would like to examine, they are recorded in topical modules every one or two years and are unable to be tracked over time with sufficient granularity."*

* On p. 12, the empirical model, can't you include month of year dummies to control for seasonality? This seems important for programs like school meals.

This is an important suggestion. We now include calendar month fixed effects (e.g., 12 fixed effects for January to December). This change is now reflected in Equation 1 on p. 14.

* Likewise, it is not at all clear why you use state-by-year unemployment rates when state-by-month rates are readily available at https://data.bls.gov/cgi-bin/surveymost?la It seems that you are giving up a lot of business cycle variation by using annual rates.

This was a typo in the text. We have confirmed that the unemployment rate is included at the state-month level and fixed this in the text (Section 5, p. 15).

* In describing your results you characterize them as "stylized facts". The notion of a stylized fact generally refers to a consensus from the literature, not the elevation of your own findings to "fact". Now it turns out that your results are consistent with the literature I refer to above, so perhaps what you find is something like a stylized fact, I would not characterize your own findings that way. This needs to be changed.

Thank you for pointing this out. We are using "stylized fact" incorrectly and have now changed our labelling to "Finding #1, Finding #2, Finding #3".

* On p. 14 in describing your first result, I would not call this insurance because that has a structural interpretation (i.e. there is a model-based literature on identifying the insurance value of transfers), but there is nothing structural about the model here. I'd call it income replacement.

We have now replaced "insurance" with "income replacement" for the subsection title on p. 16 (for Finding #1) as well as other relevant places in the text (a few paragraphs later and in the conclusion).

D. Referee 2 Comments

This paper considers how income (pre- and post-transfer), employment, and household structure evolve leading up to and following a household's receipt of several different types of social insurance benefits. Most of the focus is on SNAP, Medicaid, and Unemployment Insurance (UI), with some consideration of SSDI, workers' compensation, Veterans' Disability Compensation, TANF, SSI, and WIC. The paper mostly relies on household survey data from the Survey of Income and Program Participation. The findings are summarized into three stylized facts. 1) Market income declines prior to benefit receipt. There is a long, steady decline before entry into SNAP and a shorter, sharper decline before entry into UI. The decline in market income before entry into one type of safety net program is partially compensated for by transfers from other safety net programs. 2) Household-level employment falls, and incidence of work-limiting disability rises, when a household begins receiving benefits. 3) For SNAP and Medicaid, entry into the program is associated with increases in the likelihood of having a young child in the household and in the likelihood of having a separated, divorced, or widowed individual in the household.

The paper is clearly written. It is purely descriptive, and is upfront about that fact.

Major Comments

1. The paper should include more discussion of how sample selection might be influencing the patterns documented. To make it into the sample, a household must be continuously present in the SIPP for at least 36 months, which corresponds to 9 waves of the SIPP. I worry that this creates an analysis sample that is more geographically stable, with higher

and less volatile incomes, and with fewer changes in household composition, relative to the full population of recipients of safety net programs. I suggest that the authors try a version of the analysis that requires fewer observations per included household (maybe just 9 months before and 9 months after program receipt), and that they add discussion about how this affects the results.

Thank you for this suggestion. It has prompted us to re-think our sample selection criteria.

We realize that we do not require a balanced panel, especially if we control for household-fixed effects and seasonality (with the latter following a suggestion by Referee 1). We have now changed our primary baseline sample to be unbalanced, requiring only one month of data prior to program enrollment and 20-24 months (rather than 36 months) of consecutive SIPP participation. The sample size now increases significantly (e.g., from 4,835 to 19,403 for households enrolled in Medicaid). The results are qualitatively very similar. For example, using our new sample, Medicaid enrollees experience a decline in market income-to-poverty ratio of 26.03% (compared to 34.92% previously).

We have also taken your suggestion of employing a sample for whom we observe a balanced panel of 9 months before and after program receipt in Section 6.1 "Robustness and Extensions", along with other alternate sample restrictions. Specifically, we consider a sample restriction of two quarters before and after (~8 months). The results are effectively unchanged (see Figure A15).

2. Related to the concern about sample selection, some means in Table 1 seem implausible.

* The mean age seems on the high side for many columns in Table 1, but particularly for the WIC sample in column 9. As these households must contain an infant or young child to be eligible for the program, a mean household head age close to 44 seems quite high. Can the authors check to see if there are some outlier values of age that are skewing the mean?
* I also wonder if it might make sense to put some age restrictions on the sample more generally. The evolution of income and employment around program receipt might be quite different for households headed by someone under 65 vs. over 65.
* Why do only 47.3% of WIC-receiving households contain a child age 0-4? I would have expected this number to be close to 100%.

Thank you for these comments. In broadening our sample, the means in Table 1 now look more reasonable–e.g., the average household head age for WIC recipients is now 38.3 (rather than ~44) and the share of WIC households with children aged 0-4 is now 54%. The reason that this latter share is not closer to 100% is that it is measured during the first month of receipt, and WIC recipients can start receiving benefits when they are pregnant. As a result, the share increases substantially if we measure having a kid 0-4 at any point over some period (e.g., 16 months after initial receipt). That being said, the share will still not be 100% because some recipient kids may live in a different household from their mother at the time of interview. Currently, we do not have this description in the text, but we would be happy to insert it into a footnote if preferred.

We have also included a footnote (#15) in the first paragraph of Section 6 "Results" (p. 16):

"Note that the estimates in Table 1 are at the level of the household within which a program recipient resides. As a result, the average age of the household head and the average income of the household are both likely to be higher compared to what we may estimate for the individual recipients themselves."

Finally, with regard to age restrictions, we definitely take your point that households headed by an elderly vs non-elderly person may have different responses. We continue to include all households in our main analysis because many of the programs we analyze are also received by elderly households and we would like to paint as broad a picture as possible of the pathways at work. That being said, we do a couple things to account for the nuances you mention: 1) In Section 6.2 "Robustness Checks and Extensions" we show and discuss heterogeneous results by household type (elderly vs non-elderly with child and non-elderly without child); 2) When looking at changes in non-employment around program receipt, we define "non-employment" as being either unemployed while remaining in labor force or out of the labor force between the ages of 25-59 (non-elderly, prime earnings age).

3. The paper could include more comparison of its results to similar results from the literature, including discussion of relative magnitudes. When discussing declines in income, how do the magnitudes of the income losses experienced at the household level for UI entrants compare to income losses at the individual level as measured by Jacobson, LaLonde, and Sullivan? When discussing changes in household structure, how do the results here compare to Bane and Ellwood's estimates of change in household structure before entry into poverty? I know that there are important differences in the samples and exact variables being measured. But I'd like to see some use of other estimates from the literature to convey whether the changes documented in this paper should be thought of as large or small.

We agree that our paper is missing some comparisons of magnitudes. We have now inserted a few footnotes in Section 6 "Results" where we directly compare our results to comparable estimates from the literature:

In footnote 18 (comparing estimates of market income) on p. 18:

"Hastings and Shapiro (2018), using 2004-16 data from Rhode Island, find decreases in combined earnings and UI amounts of approximately a third over the 4 quarters preceding SNAP adoption. We find a smaller fractional decline in market income (14%) leading up to SNAP receipt, but we are also using a broader income concept and data for the entire U.S. Meyer and Mok (2019), using PSID data for the entire U.S., find declines in earnings and pre-public transfer income of 12.7% and 7.1%, respectively, in the year prior to disability onset (relative to five years before onset). We find declines of comparable magnitudes (8.2% and 10.1%) for market income in the quarter of DI and SSI receipt, relative to four quarters before receipt."

In footnote 19 (comparing estimates of employment) on p. 19:

"Our result for Medicaid is qualitatively consistent with Dague et al. (2017), who find significant employment reductions of approximately 10 percentage points (off of a baseline of 55%) in the quarters leading up to public health insurance enrollment in Wisconsin."

In footnote 21 (comparing estimates of having young children) on p. 20:

> *"Hastings and Shapiro (2018) find quantitatively similar patterns, showing that the number of children under the age of 5 increases by approximately 20% over the 4 quarters leading up to SNAP receipt."*

In Appendix B's footnote 2 (comparing estimates of employment for VA DC recipients) on p. 63:

> *"Autor et al. (2016) study a VA policy change which granted VA DC to Vietnam veterans exposed to Agent Orange and find that VA DC receipt reduced labor force participation by 18pp."*

Jacobson, Lalonde, and Sullivan (1993) find earnings declines of approximately 50% immediately associated with job displacement (and a 20% decline in the long run), while we find a 25% decline in market income leading up to UI enrollment. While these results are qualitatively consistent with each other, we did not explicitly compare our results to JLS in the paper given that there are non-comparabilities with the event examined (UI receipt vs mass layoff), timing, sample, etc. Similarly, our results aren't directly comparable to Bane and Ellwood because they decompose those in poverty by their beginning events, which is a different approach from what we take in our paper. Just to ensure apples-to-apples comparisons, we therefore compare our results (in the above footnotes) to those studies that use similar event study-style approaches.

4. Connected to the previous comment, I'd like to see the paper be more explicit about its unique contribution. I understand that the simultaneous look at many aspects of the safety net is an innovation. Is the paper also extending the time period over which we have estimates of how households react to certain elements of the safety net? Is the paper providing household-level information on the responses to some programs about which we previously have mostly individual-level information?

Thank you for this comment. In response to your comment as well as those by Referee 1 and the Editor, we have made major changes to the introduction and added a new Section 2 on "Related Literature and Contribution". Below we copy our response to the Editor's comment about this same issue.

This comment—related to Referee 1's comment 1 and Referee 2's comment 3 and 4—is one that we took the most seriously in this revision. We have made some major changes including a concise overview of the literature and our contribution in the introduction, followed by a new Section 2 entirely dedicated to the prior literature as well as our contribution. This new section is warranted given the plethora of important work that came before our paper. In Section 2, we also cite the relevant papers brought to our attention by Referee 1. Details of our changes are described above in Section A. To briefly describe our main contributions:

> *i) We study the safety net in the modern era. Much of the previous literature focuses on the context of AFDC, which targets only low-income families with children (often single mothers) and is very different from the diverse array of programs we have today. While some of these older studies focus directly on the events triggering AFDC*

receipt, there are also emphases on other issues—such as comparisons of static characteristics between participants and non-participants and what is correlated with getting off welfare or out of poverty. We are making comparisons between time periods within participants and looking at entry into programs.

ii) Within this contemporary context, we use a unified framework to compare and contrast patterns around multiple programs for the entire United States. MThe more recent studies tend to focus on a single program, a single state, and/or earnings/employment outcomes solely. We move beyond labor supply and can assess outcomes that are largely ignored by the studies using data afterin the 2000s.

iii) We also study granular changes in outcomes at the quarterly level to paint a richer picture of the pace and nature of events compared to analyzing "any event" over some time period (as many of the prior studies do). This is in part driven by trends in social science to depict panel data with "event study"-type figures, which are more conducive for studying dynamic patterns.

5. The paper covers a broad array of social insurance programs, over a fairly long time period. It would be impractical to discuss every change in each of these programs over the whole time period analyzed. But it would be useful to include some general information about whether the scope of these programs has changed over time. Perhaps the authors could add a time series figure showing total expenditure on each of the major safety net programs considered, spanning 1996 to 2020.

This is a very useful suggestion. We have now included a new appendix figure (Figure A1) that plots inflation-adjusted trends in total benefit amounts for seven of our eight programs (we omit Medicaid because the amounts are so high that they are off-the-charts). We plot the trends over 1990-2019 so that we have a few years of pre-welfare reform data and stop before the COVID-19 pandemic.

We also discuss this figure at the beginning of Section 3 "Programs and Institutional Details" (p. 7-8):

> *"Figure A1 illustrates the trends over time in benefit amounts corresponding to the programs we analyze. Over the last three decades, total benefits have grown in inflation-adjusted terms for many transfers, including DI, SNAP, SSI, and Medicaid. The two programs that experience inflation-adjusted declines between 1990 and 2019 are workers' compensation (which fell by 21%) and AFDC/TANF (which fell by 80%). While much of the existing literature centers on AFDC/TANF (which today is approximately the size of WIC), our paper provides a different perspective by focusing on the programs that are most prominent today."*

6. Page 9 acknowledges the potential for benefit receipt to be undercounted in the SIPP. Consider adding some more concrete details to this section. Offer a more quantitative description of the Meyer and Wu (2018) findings—what percentage of SNAP, UI, and Medicaid recipients and benefit payments seem to be accurately reported in the SIPP?

Thank you for this suggestion, we agree that including the quantitative percentages strengthens the argument for using the SIPP. We now do this in new footnote 10 (p. 11):

> *"In comparing weighted survey totals of recipients against administrative totals, Meyer et al. (2015) find that the SIPP in 2012 captured 108% of SSI recipients (versus 61% in the CPS), 100% of DI recipients (versus 75% in the CPS), 85% of WIC recipients (versus 68% in the CPS), 81% of SNAP recipients (versus 53% in the CPS), and 77% of TANF recipients (versus 55% in the CPS). Note that the CPS totals are derived from the CPS Annual Social and Economic Supplement, which is the official source of income and poverty statistics in the U.S. Furthermore, Meyer and Wu (2018) find that the 2008 SIPP Panel yields largely accurate estimates of the static poverty reduction effects of six government transfers (compared to linked administrative data). Among the programs overlapping with those examined in the present paper, the poverty reduction effects of Social Security, SSI, and SNAP estimated using SIPP data are within 5% of the corresponding estimates from administrative microdata."*

Minor Comments

7. On page 12, in the description of the covariate X_st, the phrase "year-month t" seems to indicate that the control varies monthly, but the next phrase says it "varies at the annual level." In this case, replace "year-month t" with "year t."

Thank you for catching this - this was an error on our end and the unemployment rate covariate does indeed vary at the monthly (not annual) level. We have fixed this (now on p.15)

8. On the top of page 17, there is a reference to "declines in enrollment" and I wonder if this is supposed to be "declines in employment."

Thank you for pointing out this typo and we have fixed it (now on p. 19).

9. I think it is a useful exercise for the authors to do some analysis separately by race or by education level of the head. This should be mentioned earlier in the paper, and described in a bit more detail. Perhaps I missed it, but I didn't see any mention of this before the conclusion.

We have now included a discussion of these analyses (by race/ethnicity, education level of head, as well as family type) in Section 6.2 "Robustness Checks and Extensions". The results can be found in Figure A16. For Medicaid and SNAP, we find that income declines are more pronounced, albeit noisy, for Hispanic households. We do not find strong patterns by the education level of the household head. For all three major programs, we find

somewhat larger income declines for non-elderly households without children, a group that has been the subject of substantial expansions to safety net programs in recent decades.

10. In Figure 2, what's reported in parentheses next to the name of each transfer program on the vertical axis? I thought it would be the mean for the reference period, as in Figure 1, but the values for SNAP, Medicaid, and UI don't match up across Figure 1 and Figure 2. So it must be something else. Add this info to the note below the figure.

You are correct, it is the mean of the reference period 4 quarters (averaged over 10-12 months) prior to initial receipt, just like in Figure 1. The means are also the same across the two figures: compare Figure 1a to 2a and 1b to 2b. We have added the following sentence to the notes in Figure 2: "The reference period is four quarters prior to receipt; the mean value four quarters prior to initial receipt can be found in the parentheses on the y-axis."

Note, however, that these means are different from those in Table 1. The Table 1 means are calculated during the first month of program receipt (to give a static snapshot of recipients at a point in time) rather than four quarters before receipt.