Using Timer Data to Evaluate the Respondent Burden of the 2013 CPS ASEC Content Test

C. Adam Bee and Aaron Cantu*

US Census Bureau
4600 Silver Hill Road, Washington, DC 20233

Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference

Abstract

This paper estimates the difference in interview duration between the usual production version of the 2013 CPS ASEC and the redesigned questionnaire fielded in the 2013 CPS ASEC Content Test. We parse 5,519 production and 7,650 redesign audit trails generated by the instrument software from complete interviews. We find that the unconditional mean duration of the redesign was 5.1 minutes longer than the production duration. The two analysis samples were not assigned randomly, however, and there are sizable demographic differences between the two samples. After adjusting for these differences, we find that that the redesign took 4.6 minutes longer. We do not include incomplete interviews in our analysis, however, and we present simulation results that suggest this exclusion may slightly attenuate estimates of differences in interview duration.

I. Introduction

For the last few decades, response rates have been declining in US household surveys. The Current Population Survey Annual Social and Economic Supplement (CPS ASEC), the source of the nation's household income and poverty estimates, has not been immune to this trend, as its response rates have decreased from 84.3 percent in 1997 to 79.6 percent in 2013.

At the same time, some researchers have raised concerns that over the years the CPS ASEC questionnaire has become less appropriate for capturing the relevant facets of households' income streams and health insurance coverage, as the institutional frameworks around those realms have slowly drifted. For example, the question about whether the responding household receives alimony is much less relevant now than it was in 1994, when the questionnaire underwent its most recent significant revision.

In response to these challenges, the Census Bureau contracted with Westat, Inc., in 2011 to evaluate the CPS ASEC instrument and recommend improvements to its income questions. Westat's recommendations for capturing income included tailored skip patterns, a dual-pass approach to catalogue all income sources before asking about amounts, and asking for income ranges as follow-up questions to "don't know" or "refused" amount responses. The Census Bureau had also been researching new ways to capture health insurance coverage, as described by Pascale, et al. (2013) and Fried, et al. (2013).

In March 2013 the Census Bureau conducted a content test that implemented these recommended changes. This paper is one among a series of reports that describe the setup and results of that March 2013 CPS ASEC Content Test. Hornick (2013) describes the sample selection process for the Content Test. Semega and Welniak (2013) describe the changes to the income sections of the ASEC and the resulting income estimates from the Content Test.

* C. Adam Bee is the corresponding author (charles.adam.bee@census.gov) and an economist in the Income Statistics Branch; Social, Economic and Housing Statistics Division; US Census Bureau, 4600 Silver Hill Road, Washington, DC 20233-8500. E-mail: charles.a.bee@census.gov; Phone: 301-763-9001. Aaron Cantu is a survey statistician in the Current Population Surveys and American Time Use Team; Associate Directorate for Demographic Programs; US Census Bureau. The views expressed in this research, including those related to statistical, methodological, technical, or operational issues, are solely those of the authors and do not necessarily reflect the official positions or policies of the Census Bureau. The authors accept responsibility for any errors. This paper reports the research and analysis undertaken by Census Bureau staff. It has undergone more limited review than official publications.

Medalia, et al. (2013) describe the changes to the health insurance coverage section of the ASEC and summarize the health insurance coverage estimates. Brault (2013) evaluates potential non-response bias from the March 2013 CPS ASEC Content Test. This paper evaluates the effect of the changed questionnaire on respondent burden, insofar as it is reflected by interview duration.

Some of the changes in the Content Test might reduce respondent burden even if they add to the length, often because the new questions may be easier for respondents to understand and answer. That type of improvement in the cognitive demands of the interview is tough to measure, however, and it falls outside the scope of this paper.

What we want to know is the amount of time taken by the redesigned instrument, in comparison to the amount of time required by a non-redesign instrument. This is important for at least three reasons. First, a long interview may reduce respondent cooperation, leading to lower response rates, higher dropout rates as respondents decline to finish interviews, lower response rates in subsequent waves of the CPS survey, and less accurate responses as respondents try to hurry through the interview. Second, long interviews are more difficult and tiring for Census field representatives, making it more difficult to retain talented interviewers. Interviewers may also feel pressured to speed up the pace of the interviewer by talking more quickly, which may make the questions more difficult to understand and worsen data quality. Third, the Office of Management and Budget (OMB) explicitly requires an estimate of respondent burden in any clearance package submitted for new surveys or changes to existing surveys.

Only a handful of previous papers have exploited the timer audit trails generated by computerized instruments in order to assess respondent burden. Walsh and Fields (2012) evaluate the durations of the 2010 and 2011 content tests of the redesigned Survey of Income and Program Participation Event History Calendar (SIPP-EHC). Perez-Lopez and Walsh (2012) find that first-person interviews that take more time are associated with subsequent person non-response. Many other studies, like Dillman, et al. (1993) and Fricker, et al. (2012), address issues of interview length, but they are focused on the number of questions rather than duration per se.

None of the previous work on interview duration per se has dealt with the issue of incomplete interviews. If longer surveys cause a higher dropout rate, then excluding incomplete interviews will attenuate estimates of difference in interview duration. Although we also do not directly address this issue in our analysis, we present simulation results that clarify the mechanics of the problem.

The remainder of this paper proceeds as follows. Section II describes the data used in the analysis. Section III presents the methods and their results. Section IV explains the implications of excluding incomplete interviews from analyses of interview duration. Section V concludes.

II. Data

We only use complete interviews for this analysis because if a respondent only gets through a fraction of the interview, the duration would appear artificially shorter, even if the pace of the interview was slow up to that point. We designate interviews as "complete" those audit trails that have a positive time for the *SUNITS* section. *SUNITS* asks "How many housing units are in your building?" We chose this section because it is chronologically the last section that all completed interviews are guaranteed to encounter on their skip pattern through the instrument. For a counter-example, suppose we chose *Q95*, which comes after *SUNITS*, and which asks about expenses for children in the household. If the household has no children, this question will not be asked, so even if a childless household completed the interview, the program would not recognize it as complete since it never got to the *Q95* screen. Likewise, it is also possible that an interview gets to *SUNITS* and still becomes an "incomplete" interview for the purposes of CPS ASEC estimation. We accept this mismatch between definitions of "complete" interview because our purpose is to accurately gauge interview duration, not to identify candidates for whole-supplement imputation.

We also restrict the production sample to only include computer-assisted telephone interviews (CATI) in order to ensure comparability, since the redesign cases were only CATI. We exclude the (in-person) computer-assisted personal interviews (CAPI) by removing from the sample all audit trails that contain the *bCAPI_front* section.

There are 5,521 complete Production audit trails, and 7,650 complete Content Test audit trails.[1] These audit trails constitute the full universe of cases meeting the above criteria, and thus the analysis sample is identical to the population of interest. This paper should be considered a historical summary of results and not an estimation of the performance of the two instruments in a repeated sampling context. Accordingly, the measures reported in this paper do not include measures of dispersion, so as to avoid conflation with the usual measures of sampling error.

In analyses that employ demographic covariates like age and educational attainment, we use pre-edit variables because the Content Test data were not edited at the time we conducted the analyses. For example, for age we use *peage*, for relationship to reference person *perrp*, and for education *peeduca*.

Semega and Welniak (2013) describe the unweighted differences between the Production and the Content Test samples, which are reprinted in Table 1. As one might expect from the differences in the ways these samples were selected, the discrepancies in demographic characteristics are large. The proportion of elderly respondents at least 62 years old, for example, is more than twice as large in the Content Test sample as in the Production sample. This age discrepancy is reflected in the differences in median age, as the Content Test sample has a median age 12 years older.

These observed differences suggest that unobservable characteristics will also vary between the two samples, which raises concerns that the non-random Content Test sample selection invalidates any assessment of the performance of the redesigned instrument relative to the current version. Differences in response distributions could be due to the changed instrument, or they could be due to differences in the sample. One seeming point of optimism in Table 1 is that the proportion of workers was much more similar between the samples, but that similarity is likely coincidental given the large differences in age distributions.

III. Methods and Results

As a first pass at describing the differences in interview duration between the production and redesigned instruments, in Table 2 we take simple means over more than 5,500 complete production telephone interviews and 7,700 complete redesign interviews. We see that the redesigned instrument takes a little over five minutes longer on average, or about 20 percent.

The audit trails generated by the interview software let us break out the durations by section. The structure of the interview is that it starts off with the CPS Basic questions about employment, which are asked every month, and then the ASEC supplement proceeds sequentially through a section on income, then health insurance coverage, followed by migration, and ending with a few questions that are inputs to the Census Bureau's new Supplemental Poverty Measure (SPM).

One useful aspect of the redesign is that not every section was changed. Some of the sections are exactly identical between the two instruments. Specifically, the CPS Basic, Migration, and SPM sections are identical. We see in Table 2 that the durations for those unchanged sections are quite close,[2] which is what one would hope to see if the two instruments had actually been randomly assigned to households. The very small differences in the durations of unchanged sections could be interpreted as potentially reassuring evidence that the demographic differences in the two samples, both observed and unobserved, make little difference when it comes to interview duration.

If that is indeed the case, then the observed differences in the changed sections would in fact reflect real differences in the instruments themselves rather than differences in the samples. These observed differences are small but meaningful. The Income section took 3.7 minutes longer on average in the Content Test, representing a 32 percent

[1] Two of the Production cases are exact duplicates—they have the same values for every field as two other cases. After deleting these two cases, we are left with 5,519 complete Production interviews.

[2] Specifically, the CPS Basic time is 0.3 minutes longer in the Content Test, while the Migration and SPM sections are 0.1 minutes shorter. The decrease in Migration time does represent a 22 percent decrease in the average time since the Migration section is so short, but this could be explained by the fact that in order for households to be eligible to be interviewed for the Content Test, they could not have moved since the last time they participated in a production CPS interview, and thus by construction the Content Test respondents could not have any migration to report.

increase from the mean production duration of 15.3 minutes. The Health section took 1.5 minutes longer, 58 percent more than the 2.7 minutes taken by the production version.

Given that the Content Test sample was so much older than the production ASEC sample, the lack of a difference in durations of unchanged portions of the instrument is somewhat of a puzzle. Anecdotes from interviewers suggest that older respondents seem to differ systematically in interview duration from younger respondents. We investigate this hypothesis in Table 3.

The top half of Table 3 presents differences in mean section durations by the age of the householder. We find that the differences between the instruments swamps the difference between age groups, which indeed is negligible.[3] The hypothesis that older respondents take substantially longer to interview is thus rejected.

This unexpected finding may raise concern that the timer data was improperly extracted from audit trails or improperly matched to the files of demographic characteristics. To check this, we compare mean durations by household size. Household size should be highly correlated with interview time because the interview has to collect information on each family member. In the field, when a respondent asks how long the interview is expected to take, interviewers often ask for the household size in order to give an estimated duration. It would be extremely surprising if duration were not monotonically increasing in household size.

Reassuringly, the bottom half of Table 3 indicates this is indeed the case: each additional person adds time to the interview, though at a decreasing rate. The proportional increase in duration from the production instrument to the redesign is roughly constant across groups, however. At each level, the difference between instruments is about the same, about 22 to 26 percent. This suggests that even if the two samples vary in their distributions of household sizes, that variation would not explain the difference between instruments.

Rather than investigate a variety of demographic characteristics individually, we include several householder characteristics at once as controls in an ordinary least squares (OLS) regression of ASEC section durations in Table 4. Controlling for the age, sex, and education of the householder, as well as for household size, we find that the measured difference between instruments is 4.4 minutes, down from the unconditional difference of 5.1 minutes. This decrease, about 40 seconds, suggests that the Content Test sample's characteristics correlate slightly with longer interview durations. Whether this 40-second difference is substantively significant is a matter of judgment. To the extent that this difference is surprisingly small given the higher age of the Content Test sample, however, suggests that the demographic differences between the two samples made surprisingly little difference when it comes to interview duration, compared to the additional duration associated with the redesign of the instrument. The top half of Table 3 supports this interpretation, as it suggests that age has relatively little effect on interview duration.

We repeat this regression exercise for each of the sections of the questionnaire. Table 5 contains the results from regression of the Income section of the questionnaire. The coefficient on the Content Test indicator indicates that the redesigned instrument took 3.3 minutes longer, slightly less than the 3.7 minutes indicated by the unconditional comparison reported in Table 2. Table 6 shows that the Health section took 1.4 minutes longer, conditional on observables.

Table 7 compares coefficients on the Content Test indicators across all of the instrument sections. Again we see that for the unchanged sections there is little difference in durations between the two instruments, which is what one would expect.

These results represent differences in means, though, which implies that they may be sensitive to outliers. One might also wonder about the differences in median durations, 25th and 75th percentiles, and other characteristics of the distribution. This may be a lesser concern since very long interviews still represent burdensome interviews, and since interviewer costs are linear in interview time. These would imply that we would want to account for even very long interviews. On the other hand, particularly long interviews may instead represent cases in which the interviewer and respondent engage in conversation only tangentially related to the interview, which arguably do not add to

[3] Although some householders are younger than 25, they are excluded from the table because they are few in number and because they are a relatively disparate group.

respondent burden or interviewer fatigue as typically conceived. At any rate, we characterize the full distributions of interview durations in Figures 1 through 4.

Figures 1 through 4 are histograms of residuals from regressions of various sections' durations, run with both samples pooled (i.e., regressions are not run separately for each instrument). The regressions have the same set of controls as the previous regressions, except for the Content Test Indicator (which would be collinear). Histogram bins are one minute wide. One could think of these distributions as "controlled" for differences in demographic characteristics. If sample observables were identical between the two samples, as would have been the case if the instruments had been randomly assigned, then the difference between the means of the two distributions would be exactly the same as the coefficient on the Content Test indicator from the previous OLS specifications. Although the magnitude of the difference between the distributions is not easily interpreted here, it is still revealing to compare the positions of the distributions to each other.

Figure 1 shows the distributions of residual durations from regression of the ASEC supplement durations on householder age, sex, education, and household size. We see that the distribution of durations of the redesigned instrument in the Content Test (green bars) has first-order stochastic dominance over the production distribution (clear bars). It is immediately obvious that the redesign distribution is "to the right" of the production distribution. Although both distributions are positively skewed, the redesign right tail still dominates the production tail. The takeaway is that not only are the redesign durations longer on average, but also that the short redesign interviews take longer than the short production interviews, the long redesign interviews take longer than their production counterparts, etc.

Figure 2 and Figure 3 present the analogous distributions of residuals for the Income and Health sections, respectively. They illustrate the same first-order stochastic dominance of the redesign durations that we saw in Figure 1. The Health residuals have a much narrower distribution both because the underlying distribution of durations is narrower but also because the observables explain a higher proportion of the variance in Health durations.

Figure 4 looks very different from the preceding three figures. It presents the distributions of durations from the Basic CPS questions. Basic CPS is the portion of the interview that is conducted every month, and it is administered before the ASEC supplement. Crucially, the CPS Basic questionnaire is identical between the redesign and the production instruments. If the Content Test had been a carefully controlled trial, one would expect that, not only should the mean CPS Basic durations be nearly identical between the two surveys, but also that the distributions be very similar, differing only due to sample variability. This is what we in fact observe, even though the Content Test was in fact not randomly assigned: after controlling for some basic demographic characteristics, the two distributions turn out to be very similar. We interpret this as evidence that controlling for the demographic characteristics of the sample does a good job of pulling the instrument durations into comparability, and that the remaining unobserved covariates do not bias the results very much.

In Table 8 we further exploit the fact that the CPS Basic questionnaire is identical between the Content Test and production instruments, by adding the CPS Basic duration as an explanatory variable in the regressions of the other sections' durations. The idea is that the rate of progression through the CPS Basic portion of any given interview should be roughly the same speed as the rest of the interview. This implies that cross-sectional changes in the durations of the CPS Basic interview are directly proportional to changes in durations of the ASEC section, which allows construction of a valid counterfactual ASEC duration in the absence of the changes to the ASEC instrument. In essence, instead of having the usual control group employed in standard randomized control trial designs, we employ a "control section" that stays constant between groups.

The first column in Table 8 simply reprints the Content Test indicator coefficients from the original specification in Table 7, with controls for the age, sex, and education of the householder, as well as for household size. The second column displays the coefficients on the Content Test indicator when the CPS Basic section duration is added as an additional control.[4] This second column of coefficients does not materially differ from the first column. For example, the measured effect of the redesign on the Income duration is 3.2 minutes with the additional control,

[4] The second row of Table 8, in which the CPS Basic section duration is also the dependent variable, is included for completeness, as obviously the regression is collinear.

compared to 3.3 minutes without. The narrowness of these differences demonstrates that even the limited set of demographic controls already does a good job of absorbing the variation between the two samples' characteristics to the extent that they relate to interview duration. Comparison of R^2 values across specifications reveals the same story: adding the CPS Basic time as a control does not dramatically increase the explained variation.

Having addressed household characteristics, we turn our attention to interviewer characteristics as a potential confounding factor. Although production and Content Test cases were not assigned to interviewers randomly, anecdotal evidence suggests that assignment was at least fairly arbitrary. Most interviewers fielded a mix of Content Test and production cases. Still, if there was a systematic correlation between interviewers' speed and the number of Content Test cases they took on, that would bias our estimates. For example, if more experienced interviewers were disproportionately assigned to Content Test cases in an effort to bolster its low response rate, and experienced interviewers tended to get through interviews faster, that would imply that our estimate of the effect of the redesign on duration would understate the true effect.

To address these concerns, we employed interviewer fixed effects to obtain the results presented in Table 9. The format of Table 9 mirrors that of Table 8: the first column reprints the coefficients from the original OLS specification in Table 7, and the other four columns refer to the specification with interviewer effects included. The Content Test coefficient increases very slightly when variation in interviewer speeds is accounted for, which implies that faster interviewers did conduct more Content Test interviews. Still, the practical significance of this slight difference in coefficients is negligible. Difference in interviewer speeds does not explain the longer duration of the redesigned questionnaire.

Table 10 summarizes the results of all the above specifications, along with one in which both CPS Basic section durations are included at the same time as interviewer effects. The message of Table 10 is that the estimated duration of the redesigned questionnaire is about four or five minutes longer than that of the current version, and that estimate is robust across a wide variety of specifications.

IV. Caveat: The Implications of Excluding Incomplete Interviews

A potentially important caveat to the above analyses is that they are all based only on complete interviews. This is justified on the grounds that incomplete interviews will obviously appear artificially shorter. If one of the instruments has a higher dropout rate, including incompletes would bias estimates of the difference. For example, if Content Test interviews have a higher dropout rate, then including these artificially short interview durations will make the Content Test seem shorter than it would had every household completed the full interview. Further, if dropouts are caused by respondents' irritation or fatigue, the short duration of incomplete interviews would not reflect an easing of respondent burden but rather its opposite. It is clear that naively including the durations of incomplete interviews, without adjustment, and analyzing them as equivalent to complete interviews that took the same amount of time, would be a gross error.

Still, simply excluding incomplete interviews and considering only completed interviews creates a potential for sample selection. If one of the reasons that respondents quit an interview is that it is taking too long, then a lengthier instrument may lead to more dropouts from the high end of the duration distribution. This differential selection will cause comparisons of means (or conditional means) between the two instruments to be biased. Note that this is a separate problem from the sample selection associated with this particular Content Test, which yielded a comparison sample with large differences in age and other demographic characteristics. The sample selection created by excluding incomplete interviews would exist even if the samples had been drawn randomly as in a randomized control trial.

This process is concretely illustrated in a simulation depicted in Figures 5 and 6. Figure 5 illustrates the setup of the simulation. Suppose that each instrument has a true, latent, unobserved distribution of durations corresponding to how long each instrument would take if all respondents were to complete their interviews. These latent distributions are represented in Figure 5 by the blue and red lines. The latent distributions are parameterized as gamma distributions with a shape parameter 5 and a scale parameter of 1. We normalize the true difference in interview length to be 10 minutes. We then assume that each household has some time threshold at which point they terminate the interview. These thresholds are parameterized as $T \sim 2*\Gamma(5,1)$, or double the "production" latent distribution, and represented in Figure 5 by the green line.

We draw 100,000 simulated households, each household getting a draw from each of the three distributions. If a household's draw from the latent duration distribution is greater than their draw from the distribution of thresholds, we suppress that duration. Note that the same threshold is used for both the "production" and the "redesign" households, illustrating that this differential sample selection is a problem even when the underlying characteristics of the same are identical, as in the case of a randomized control trial.

We compare the distributions of the remaining, "complete" interviews in Figure 6. The blue and red lines are the latent duration distributions, reprinted from Figure 5. The green and orange lines represent the "observed" duration distributions, remaining after the thresholds are applied. We observe that the threshold tends to remove the top end from each distribution: households with longer interviews are more likely to hit their threshold and stop the interview. Thus the remaining, "observed" interviews are shorter on average than their true, latent distribution. Further, since the "redesign" distribution is higher than the "production" one, it is more highly selected. This attenuates the estimated difference in the interview duration by 7 percent: the observed difference is 9.3 minutes, compared to the latent difference that we normalized to be 10 minutes.

This sample selection phenomenon highlights the inherent problem that interview duration is simply not a sufficient statistic for respondent burden. Even if an estimate of interview duration perfectly accounted for incomplete interviews, it would not necessarily correspond to the frustration that may lead respondents to abort interviews, nor would it reflect the resources that go into fielding the interviews. Although a perfectly adjusted duration would tell what durations would be if respondents never stopped answering questions, the real-world relevance of that information is not obvious. Further, it is not clear to what extent households may correctly guess that the interview will take more time than they are willing to give, and thus not even begin an interview that they believe will take too long. If this mechanism is sufficiently strong, lengthier instruments could lead not just to higher dropout rates, but also higher non-response rates.

V. Conclusion

This paper has three main takeaway results. First, all the signs point toward the redesigned instrument taking somewhat longer to complete than the current CPS instrument. That does not necessarily reflect the entirety of changes to respondent burden, however: we've made other improvements that reduce cognitive burden and increase data quality, and those might well be worth adding four or five minutes to the length of the interview.

The second take-away is that even though the samples are very different and we do not have a clean, randomized trial, we can still adjust for the observable characteristics. Those adjustments slightly attenuate the change, which suggests that the redesign sample's observed characteristics were slightly correlated with being slower. But the fact that controlling for observables makes so little difference suggests that even if we could perfectly control for all the differences, the redesign would still come out slightly longer.

The last take-away is that evaluations of interview duration should account for incomplete interviews. That does turn out to be a technically complex challenge, but it is something we hope to incorporate in future evaluations, including an evaluation of the split-sample test fielded in 2014. Methods that may help address the problem posed by excluding incomplete interviews may include Heckman selection correction and imputation of missing section durations based on observable characteristics and completed section durations.

We have also identified a few other avenues for possible extension or future research. One could relatively easily add an analysis of interviewer experience. The concern is that many of these interviewers have been giving the current questions for years, and they have not had much practice yet with the new instrument. Although the redesigned element anecdotally does not appear to be more difficult to learn, even with custom skip patterns and bracket responses, it is certainly possible that interviewers got somewhat faster as they completed more and more of these redesign interviews, which may suggest that our measured time difference will decrease as interviewers get more comfortable with the new instrument.

An additional caveat to be addressed in future research is that production households are selected for CATI administration only if they are deemed to be relatively cooperative. Otherwise they continue to be administered in-person CAPI interviews, and thus our production sample is selected to contain only relatively "easy" interviews.

Future work may attempt to adjust for this by considering the difference between CATI and CAPI interview durations in the production sample.

Lastly, even though the sample construction of the 2013 Content Test did not allow for a clean experimental design, we can still credibly evaluate the new skip pattern for low-income households. The new skip patterns moves sources like public assistance and SNAP farther up in the instrument, based on a family income question at the end of the CPS Basic questionnaire. This low-income skip pattern is triggered when family income is below $75,000.[5] We can exploit that hard cut-off. If we compare households just on either side of that cutoff, they should be essentially the same and we would have a clean, credible experimental design for assessing the effect of this particular change.

REFERENCES

Brault, Matthew W. 2013. "Non-response Bias in the 2013 CPS ASEC Content Test." Paper presented at the 2013 meetings of the Federal Committee on Statistical Methodology.

Dillman, Don A., Michael D. Sinclair, and Jon R. Clark. 1993. "Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys." *Public Opinion Quarterly* 57(3): 289-304.

Fricker, Scott, Brett Creech, Jeanette Davis, Jeffrey Gonzalez, Lucilla Tan, and Nhien To. 2012. "Does Length Really Matter? Exploring the Effects of a Shorter Interview on Data Quality, Nonresponse, and Respondent Burden." Paper presented at the 2012 meetings of the Federal Committee on Statistical Methodology. March 16.

Fried, Brett, Joanne Pascale, and Michel Boudreaux. 2013. "Measuring Transitions, Spells of Uninsurance, and Churning Using the Redesigned CPS." Paper presented at the 2013 Joint Statistical Meetings.

Hornick, David. 2013. "The 2013 Annual Social Economic Supplement Health Insurance Questionnaire Test: The Sample Design." Paper presented at the 2013 meetings of the Federal Committee on Statistical Methodology.

Medalia, Carla, Amy Steinweg, Brett O'Hara, David Lee, Jessica Smith, Joanne Pascale, and Jonathan Rodean. 2013. "Health Insurance in the Current Population Survey: Redesign and Production." Paper presented at the 2013 meetings of the Federal Committee on Statistical Methodology.

Pascale, Joanne, Michel Boudreaux, and Amy Steinweg. 2013. "Health Insurance Measurement in the Current Population Survey: Test Results and Next Steps for a Redesign." Paper presented at the 2013 Joint Statistical Meetings.

Perez-Lopez, Dan, and Rachael Walsh. 2012. "Differing Person Nonresponse Rate by Interview Length and Respondent Characteristics: Results from the 2011 SIPP-EHC Field Test." Paper presented at the 2012 meetings of the Federal Committee on Statistical Methodology.

Semega, Jessica, and Edward J. Welniak, Jr. 2013. "Evaluating the 2013 CPS ASEC Income Redesign Content Test" Paper presented at the 2013 meetings of the Federal Committee on Statistical Methodology.

Walsh, Rachael M., and Jason M. Fields. 2012. "Reducing Respondent Burden: Evaluating the Progress of the SIPP-EHC" Paper presented at the 2012 meetings of the Federal Committee on Statistical Methodology.

[5] Previously, transfer program participation questions were not asked at all for households with income under $75,000.

Table 1: Descriptive statistics

	Production	Content Test
Less than age 15	23%	14%
Age 15-24	14%	9%
Age 25-61	49%	47%
Age 62+	14%	31%
Median age	43	55
Worker	56%	57%

Source: Semega and Welniak (2013), from the 2013 CPS ASEC and the 2013 CPS ASEC Content Test.

Table 2: Simple mean interview durations

	Mean durations			
Section	Production	Content Test	Difference (minutes)	Percent Difference
All sections	26.8	32.2	5.4	20
CPS Basic	7.5	7.8	0.3	4
ASEC	19.3	24.4	5.1	27
Income	11.5	15.3	3.7	32
Health	2.7	4.2	1.5	58
Migration	0.3	0.3	-0.1	-22
SPM	4.7	4.7	-0.1	-1
Sample size	5,519	7,650		

Note: Each column is rounded individually.
Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Table 3: Differences in mean instrument duration, by age of householder and household size

	Mean ASEC Section Durations		
Group	Production	Content Test	Diff. (%)
Age of householder			
25 to 61	19.5	24.7	27
62 and older	19.0	24.1	27
Household size			
One person	14.0	17.0	22
Two people	19.6	24.7	26
Three or four people	22.6	28.1	24
Five or more people	25.9	32.2	24

Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Table 4: Regression of ASEC section time on demographic characteristics

Independent variable	Coefficient
Content Test indicator	4.4
Age	0.1
Female	-0.3
Education (<HS omitted):	
High school	0.9
Some college	1.6
Associate	1.5
College	1.9
Household size	3.1
Sample size	13,169
R^2	0.252
Unconditional mean difference	5.1

Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Table 5: Regression of Income section time on demographic characteristics

Independent variable	Coefficient
Content Test indicator	3.3
Age	0.1
Female	-0.3
Education (<HS omitted):	
High school	0.8
Some college	1.5
Associate	1.4
College	1.9
Household size	1.7
Sample size	13,169
R^2	0.193
Unconditional mean difference	3.7

Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Table 6: Regression of Health section time on demographic characteristics

Independent variable	Coefficient
Content Test indicator	1.4
Age	0.0
Female	-0.3
Education (<HS omitted):	
High school	-0.1
Some college	-0.2
Associate	-0.2
College	-0.2
Household size	0.7
Sample size	13,169
R^2	0.304
Unconditional mean difference	1.5

Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Table 7: Regression coefficients on Content Test indicators across all instrument sections

Dependent variable	Coefficient on Content Test Indicator	R^2
All sections	4.6	0.27
CPS Basic	0.2	0.16
ASEC	4.4	0.25
Income	3.3	0.19
Health	1.4	0.30
Migration	-0.1	0.06
SPM	-0.3	0.16

Controls are head's age (linear), sex, and educational attainment; and household size. Sample consists of 13,169 household interviews. Section durations (the dependent variables) are measured in minutes. Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Table 8: Regressions controlling for CPS Basic section duration

	Coefficient on Content Test Indicator		
Dependent variable	Original OLS	Including CPS Basic control	R^2
All sections	4.6	4.3	0.61
CPS Basic	0.2	0.0	1.00
ASEC	4.4	4.3	0.32
Income	3.3	3.2	0.26
Health	1.4	1.4	0.33
Migration	-0.1	-0.1	0.09
SPM	-0.3	-0.3	0.18

Controls are head's age, sex, and educational attainment; household size; and CPS Basic section duration. Sample consists of 13,169 household interviews. Section durations (the dependent variables) are measured in minutes. Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Table 9: Regressions controlling for interviewer effects

	Coefficient on Content Test Indicator		
Dependent variable	Original OLS	Including Interviewer Effects	R^2
All sections	4.6	4.9	0.40
CPS Basic	0.2	0.2	0.23
ASEC	4.4	4.7	0.40
Income	3.3	3.5	0.31
Health	1.4	1.5	0.37
Migration	-0.1	-0.1	0.08
SPM	-0.3	-0.2	0.32

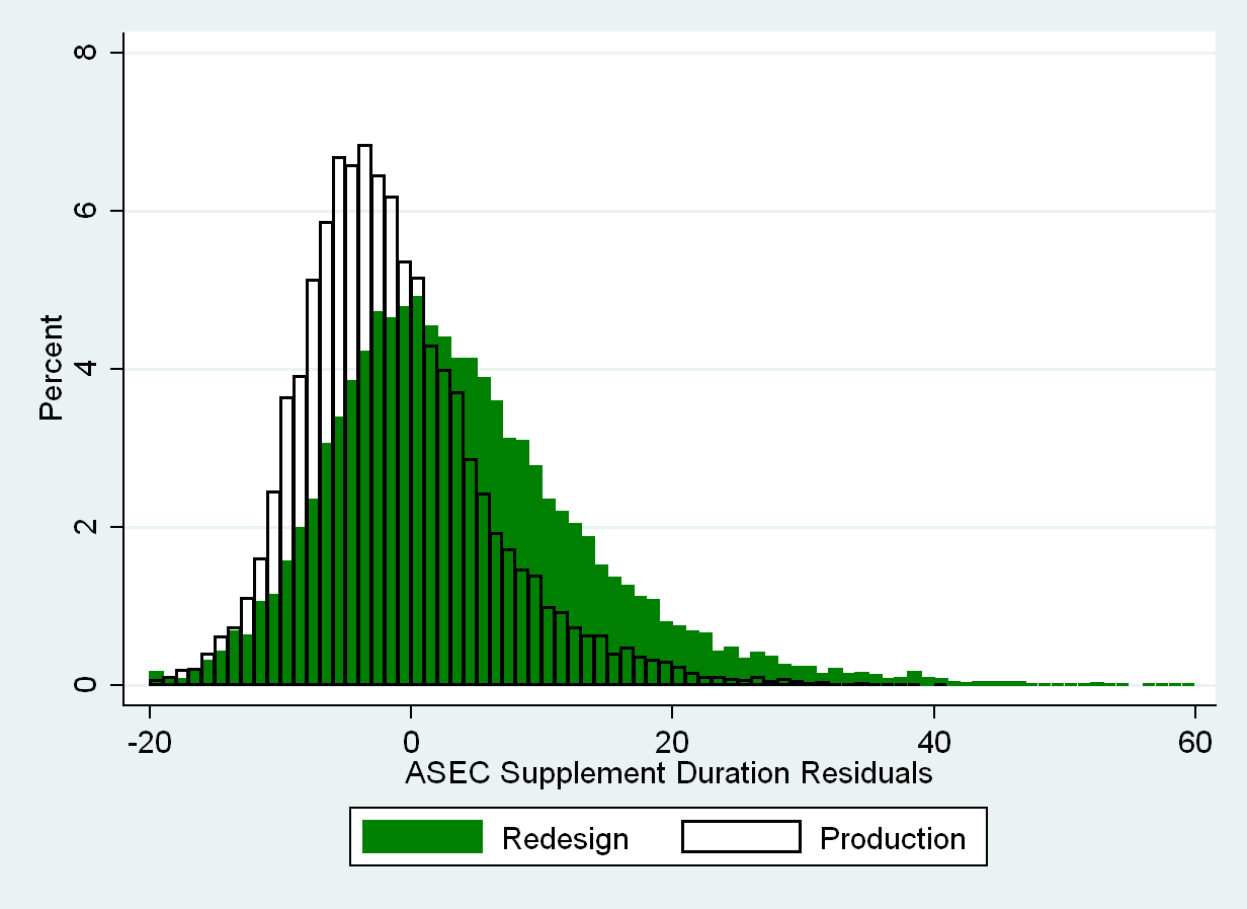
Controls are head's age, sex, and educational attainment; household size; and interviewer effects. Sample consists of 13,169 household interviews. Section durations (the dependent variables) are measured in minutes. Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Table 10: Comparison of regression coefficients across all specifications

Dependent variable	Unconditional means	OLS Coefficients on Content Test Indicator			
All sections	5.4	4.6	4.3	4.9	4.6
CPS Basic	0.3	0.2	0.0	0.2	0.0
ASEC	5.1	4.4	4.3	4.7	4.6
Income	3.7	3.3	3.2	3.5	3.4
Health	1.5	1.4	1.4	1.5	1.4
Migration	-0.1	-0.1	-0.1	-0.1	-0.1
SPM	-0.1	-0.3	-0.3	-0.2	-0.2
CPS Basic			X		X
Interviewer effects				X	X

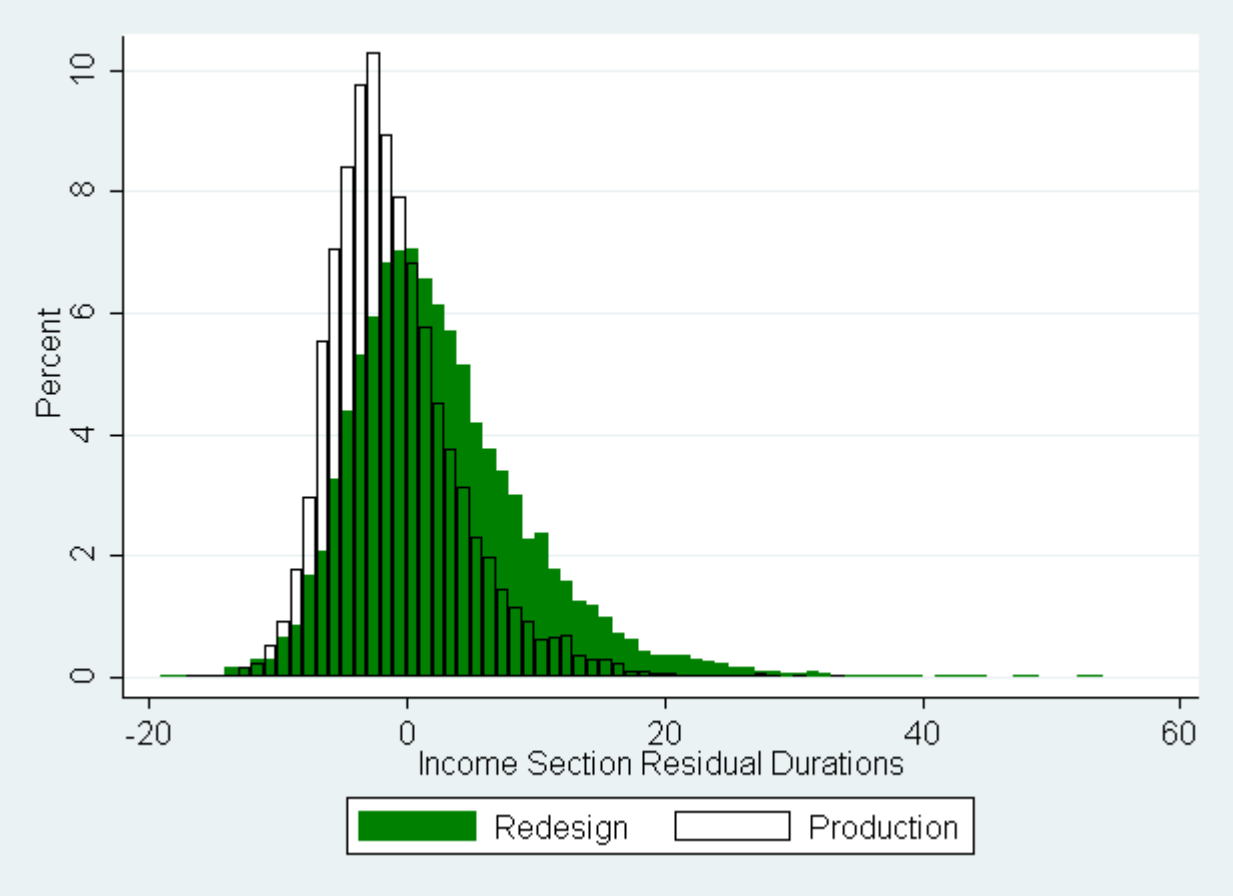
Controls are head's age, sex, and educational attainment; and household size. CPS Basic duration and Interviewer effects are included in some specifications as indicated. Sample consists of 13,169 household interviews. Section durations (the dependent variables) are measured in minutes. Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Figure 1: Distributions of ASEC Supplement residual durations, by instrument



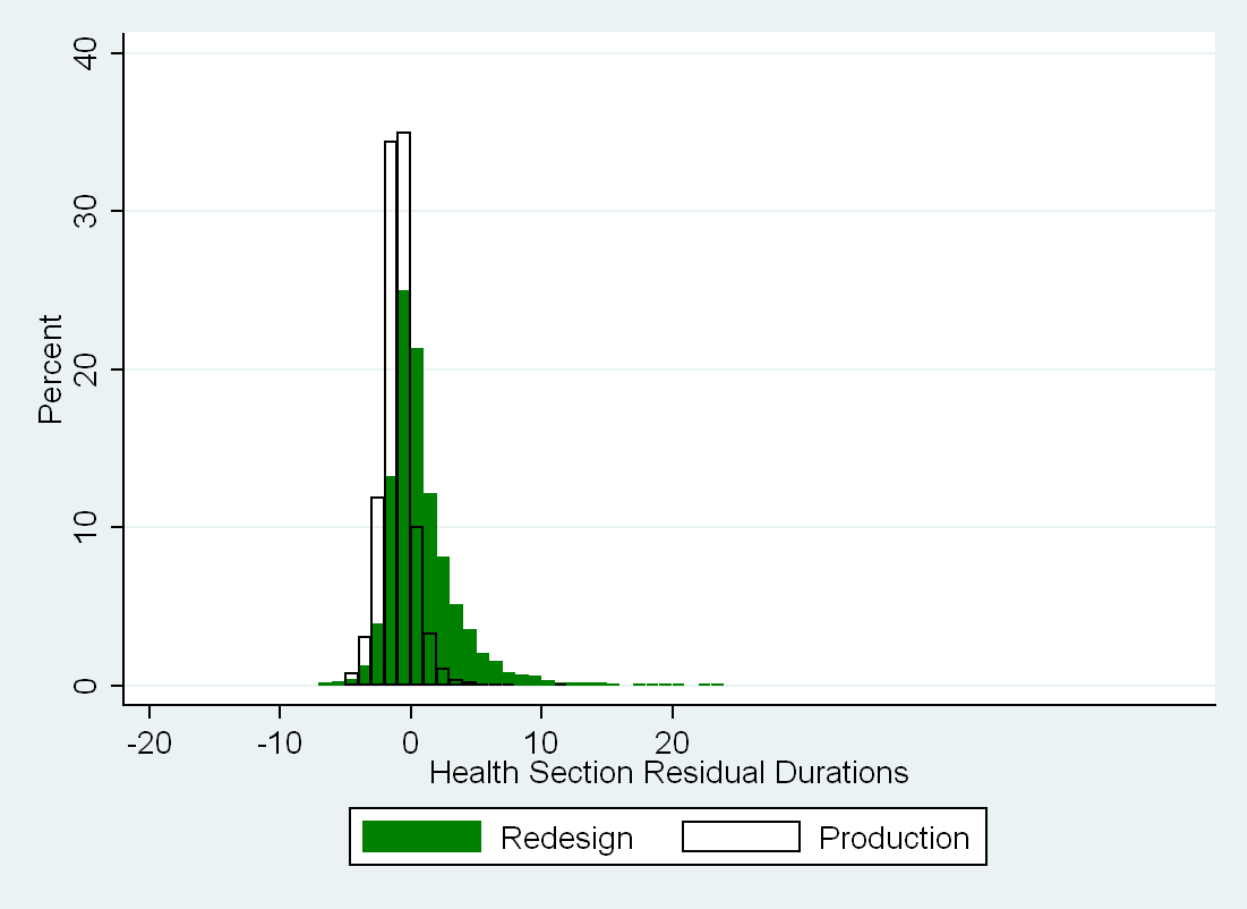
Controls are head's age, sex, and educational attainment; and household size. Sample consists of 13,169 household interviews. Section duration residuals are measured in minutes. Histogram bins are one minute wide. Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Figure 2: Distributions of Income section residual durations, by instrument



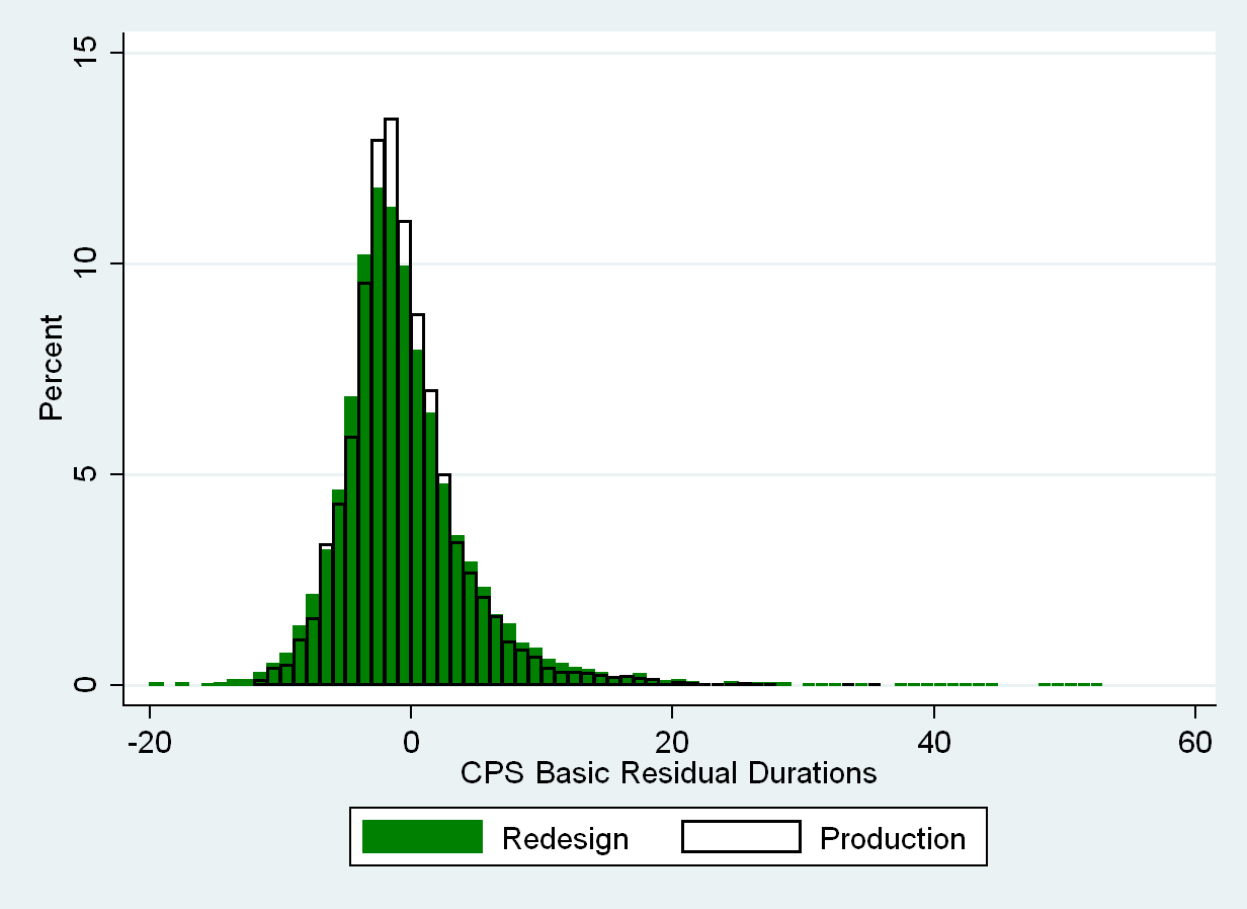
Controls are head's age, sex, and educational attainment; and household size. Sample consists of 13,169 household interviews. Section duration residuals are measured in minutes. Histogram bins are one minute wide. Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Figure 3: Distributions of Health section residual durations, by instrument



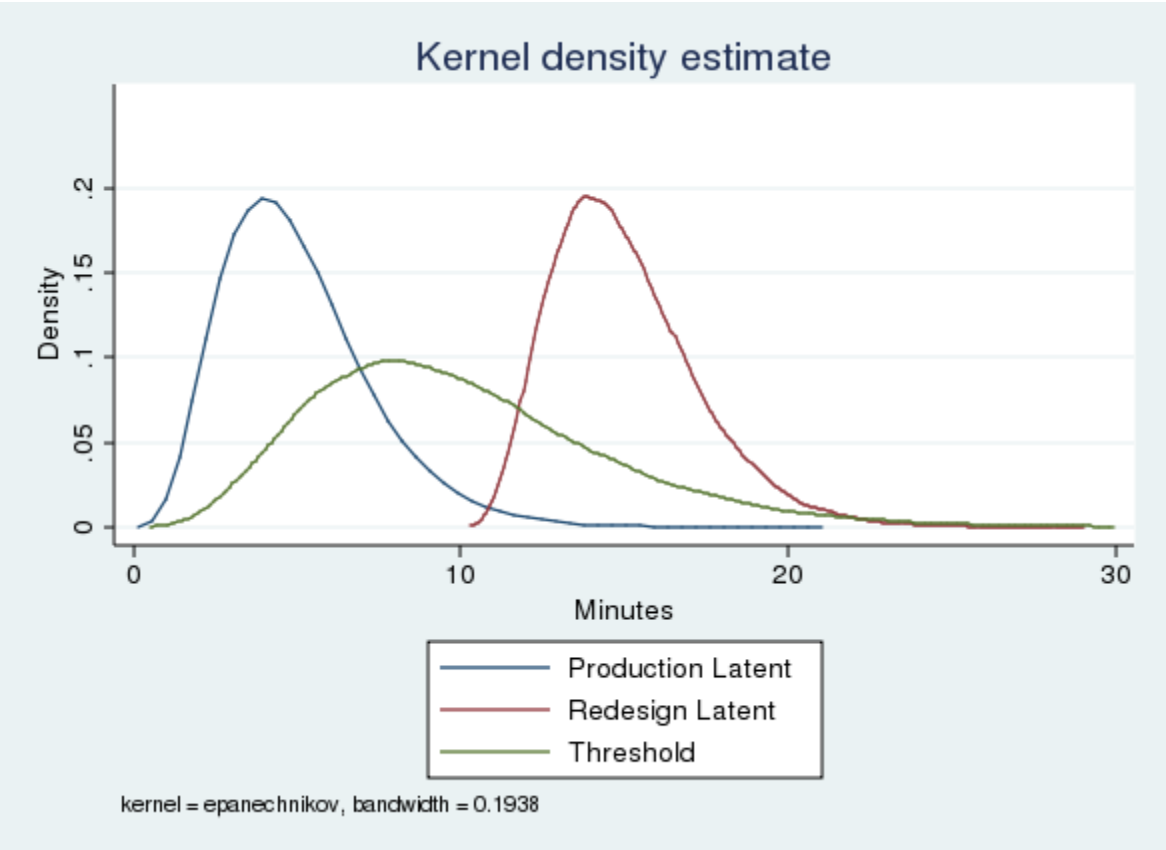
Controls are head's age, sex, and educational attainment; and household size. Sample consists of 13,169 household interviews. Section duration residuals are measured in minutes. Histogram bins are one minute wide. Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Figure 4: Distributions of Basic CPS residual durations, by instrument



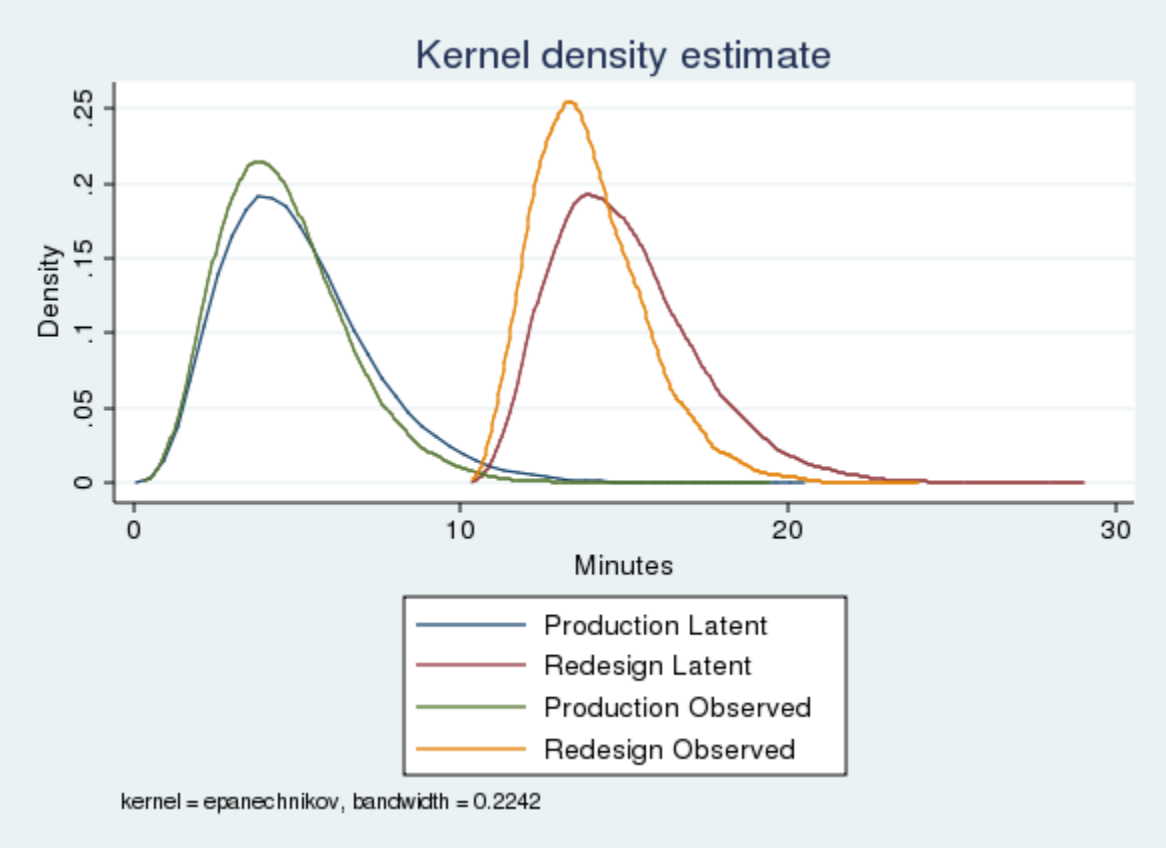
Controls are head's age, sex, and educational attainment; and household size. Sample consists of 13,169 household interviews. Section duration residuals are measured in minutes. Histogram bins are one minute wide. Source: 2013 CPS ASEC and 2013 CPS ASEC Content Test.

Figure 5: Simulated distributions of unobserved interview durations for Production and Redesigned instrument, with simulated distribution of respondent-level duration thresholds



Source: Simulated data.

Figure 6: The effect of applying simulated respondent-level duration thresholds on observed distributions of interview durations



Source: Simulated data.