Survey of Income and Program Participation

Jason Fields
Matthew Marlay
Patrick Campanello

US Census Bureau

December, 2015 – FCSM
Washington, DC

Outline

- Reengineering SIPP background and key changes
- Timeline and release plans
- Staffing and training
- Some innovations
  - Model based incentives
  - Model based imputation with administrative inputs
  - Monitoring
- New data access tool from Orlin Research

Survey of Income and Program Participation

- National panel survey – Since 1984 with sample size between about 11,000 and 45,000 interviewed households
  - The duration of each panel varies from 2½ yrs to 4 yrs
  - The SIPP sample is a multistage-stratified sample of the U.S. civilian non-institutionalized population
  - Longitudinal – following original sample household members (all 15 and over are followed)
- SIPP 'Classic' – 1984-2008
  - Uses a 4-month recall period – 3 interviews / year
  - The sample is divided into 4 rotation groups for monthly interviewing
  - Paper from 1984-1993 and DOS based CASES instrument from 1996-2008
- SIPP 2014
  - Annual survey with four month interviewing window – recall to beginning of prior (reference) year
  - Event History Calendar (EHC) component to facilitate recall
  - Paradata and ancillary data include (contact history and reluctance, training certification, keystroke files, cost and case management, prior wave data for waves 2+)
- Interviews are conducted by personal visit and by decentralized telephone if requested for follow-up

Key Design Changes and Benefits of Reengineering

- Annual interview
- 12-month reference period from 4-month
- Event History Calendar (EHC) methods - Facilitates respondent recall over longer reference period
- Reduced cost through annual administration

Scope
- Similar to SIPP
- Broader than core / includes key topical module content in each wave

Better integration of concepts
- EHC - integrates reporting across domains – incorporates dependent data
- Topics previously implemented as add-on modules now integrated

Increased efficiency in processing and producing data products
Flexibility in administration (dynamic interview month and reference period)

Release Plans

- Wave 1 (Collected Feb-Jun 2014)
  - Research file release – approximately the end of CY 2015 (limited content)
  - Full public use release – mid-2016
  - Available for RDC use later in 2016

- 2014 Social Security Administration Supplement on Supplement on Retirement, Pensions and Related Content
  - Dependent on Wave 1 edited input
  - Reviewing WebCATI outcomes to retain additional cases

- Wave 2 (Collected Feb-May 2015)
  - Full public use release – late 2016

- Wave 3 (Fieldwork begins ~~Feb~~ April 2016)

2014 SIPP: Content Overview

- **Coverage Questions**
- **Roster**
  - **Sex**
  - **Birthdate/Age**
- **Demographics**
  - **Hispanic origin**
  - **Race**
  - **Citizenship**
  - **Language**
  - **Marital status**
  - **Parent/child relationships**
  - **Educational attainment**
  - **Armed forces status**
  - **Type 2 people**
  - **Program/income screeners**
- **Event History Calendar**
  - **Residency**
  - **Marital history**
  - **Educational enrollment**
  - **Jobs/Time not working**
  - **Program receipt**
  - Health insurance
- Post-EHC Questions
  - Health insurance
  - **Dependent care**
  - ***Non-job income**
  - ***Program income**
  - Asset ownership
  - Household expenses
  - Health care utilization
  - Medical expenditures
  - Disability
  - ***Fertility history**
  - ***Biological parents' nativity and mortality**
  - Child care
  - Child well-being
  - Adult well-being
- Closing Screens (not on public-use file)
  - Respondent Identification Policy
  - Contact information
  - Moving intentions

Bold – in Research File Release

*** – *Tentatively included***

SIPP 2014 Interviewer Training

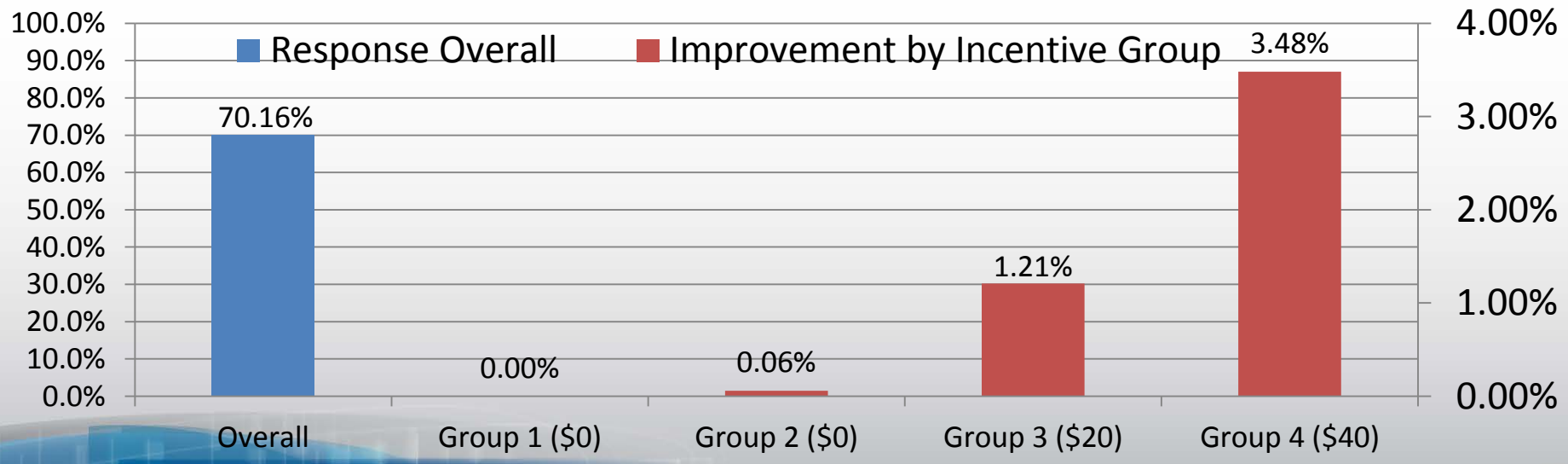
- Decentralized training after centralized 'Train-the-Trainer' at Census HQ
- Two-day generic Census training
  - New hires only - Covers cross-survey skills
  - Communicating with respondents - Administrative training
- Four-day classroom training
  - All SIPP Interviewers (FRs) - Content specific to SIPP
  - Decentralized verbatim training - Daily quizzes
  - Paired-practices - Computer based training sequences
- Pre- and post-classroom self-study modules
- Ends with certification test
  - Required before fieldwork can be started

Innovations

- Focused use of dependent data in an Event History Calendar
- Model based incentive assignment
- Type-Z model-based imputation
  - informed by administrative records
  - operationalizing methods discussed in the early 1990s - sequential regression multiple imputation
- Monitoring
  - Integration of paradata streams for management and evaluation
  - Intensive interviewer training – many aspects to monitor
  - CARI – Audio Recorded Interviews

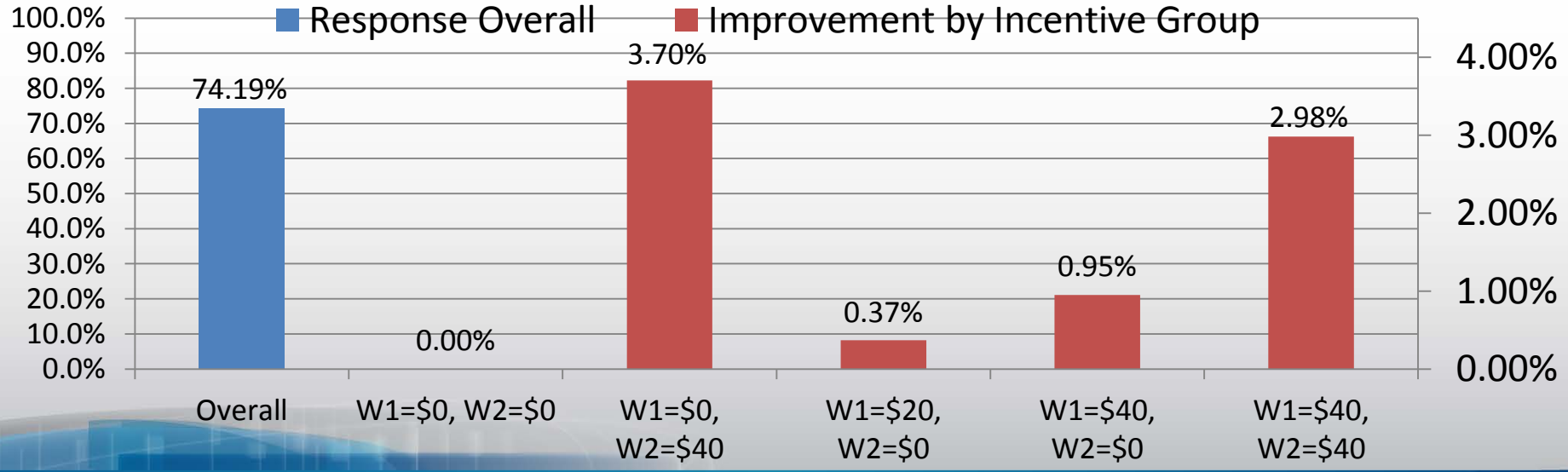
Completed Cases by Incentive Receipt Wave 1

	Total	Completed Cases	Type As	Type Bs	Type Cs
Overall	53,070	70.16%	29.84%	16.07%	3.93%
Group 1 ($0)	13,549	69.01%	30.99%	16.16%	4.15%
Group 2 ($0)	13,471	69.07%	30.93%	16.14%	3.88%
Group 3 ($20)	13,470	70.22%	29.78%	16.49%	4.10%
Group 4 ($40)	12,580	72.49%	27.51%	15.44%	3.51%



Completed Cases by Incentive Receipt Wave 2

	Total	Completed Cases	Type As	Type Bs	Type Cs	Type Ds
Overall	35,530	74.19%	21.94%	0.68%	11.87%	3.87%
W1=$0, W2=$0	8,805	72.68%	23.43%	0.77%	11.48%	3.90%
W1=$0, W2=$40	8,906	76.38%	19.86%	0.63%	12.09%	3.76%
W1=$20, W2=$0	8,911	73.05%	23.33%	0.67%	12.01%	3.62%
W1=$40, W2=$0	4,408	73.63%	21.78%	0.73%	11.80%	4.59%
W1=$40, W2=$40	4,500	75.66%	20.50%	0.60%	12.02%	3.84%



Wave 3 Incentive Plans

- Incentive assignment model:
  - Logistic regression model that predicts the probability of response using household characteristics such as:
    - Metropolitan status
    - Age
    - Sex
    - Household size
    - Tenure
    - Poverty strata
  - Assign incentives to households with the lowest likelihood of responding without an incentive and highest increase in response given an incentive

Topic Model Imputation

Problem:

- How to improve process for creating fully imputed data where whole people are missing from the household?
  - Previously relied on matching to donors and substituting prior to edits.
  - How to implement new imputation methods and still release data in a timely manner for a survey with 11,000 collected and 2,000 released variables?

Solution

- Replace item-level hot deck with parametric model-based approach
  - Helps handle small hot deck cell size problems
  - Allows inclusion of many more predictor variable
  - SIPP SSB provides the methodological foundation for modelling
  - Use administrative data to mitigate problems caused when survey data are not “missing at random”
- Use topic flags as alternative to whole-record donation for cases where respondent did not complete the whole sections of the survey.
- Indicator variables for all the major topics covered by SIPP (See Ref. Sect. 1)
- Implement new methods only for these 40+ variables

List of Topic Flags in 2014 SIPP

EHC topics:

- Education Enrollment
- Employment (job lines 1-7)
- General Assistance
- SNAP
- SSI
- TANF
- WIC
- Health insurance
  - Private
  - Medicaid
  - Medicare
  - Military
  - Other

Non-EHC topics:

Biological parent (fertility)
Dependent care
Disability- adult and child functional limitations (seeing, hearing, etc.)
Disability (difficulty finding or keeping a job because of disability)
Disability (not being able to work because of disability)
Disability payments
Energy assistance
Lump sum payments
Retirement
Retirement payments
Life insurance
School lunch
School breakfast
Social Security- Adults
Socials Security- Kids
Survivor payments
Unemployment compensation
Veterans affairs benefits
Worker's compensation

Results

Overall Percentages for cases where SIPP respondent answered the first question about jobs held (94.5% of in-universe respondents)			
Worked for pay in 2013?		W-2/Schedule C positive earnings in 2012?	
Yes	58.2	Yes	58.1
No	41.8	No	41.9

Overall Percentages for cases where SIPP respondent DID NOT answer the first question about jobs held and TF was imputed (5.5% of in-universe respondents)			
Worked for pay in 2013?		W-2/Schedule C positive earnings in 2012?	
Yes	61.5	Yes	60.4
No	38.5	No	39.6

Topic Model

Conclusions:

- Model-based imputation is feasible in a production environment for a large-scale survey
- Outside data sources (especially administrative data) are valuable:
  - Additional predictor variables in a model
  - Independent of survey non-response mechanism

Next steps:

- Model respondent-reported earnings
- Model beginning and end of spells
  - Help mitigate seam bias
- Model more topics
  - Defined benefit pension contributions
- How to best take account of spouse/parent/sibling relationships in the data when modeling

Paradata/Auxiliary Sources in Use

- Audit trail data from the Blaise/C# instrument
- Certification test for interviewer training
- Interviewer characteristics
  - Census experience
  - Prior SIPP experience
  - Supervisory status
  - Demographics
- Contact history instrument
- Mileage, case load, supervisor observation, hours billed
- Neighborhood observation
- Regional office progress management application data
- Interviewer debriefing
- Interview recordings

New data access tool from Orlin Research

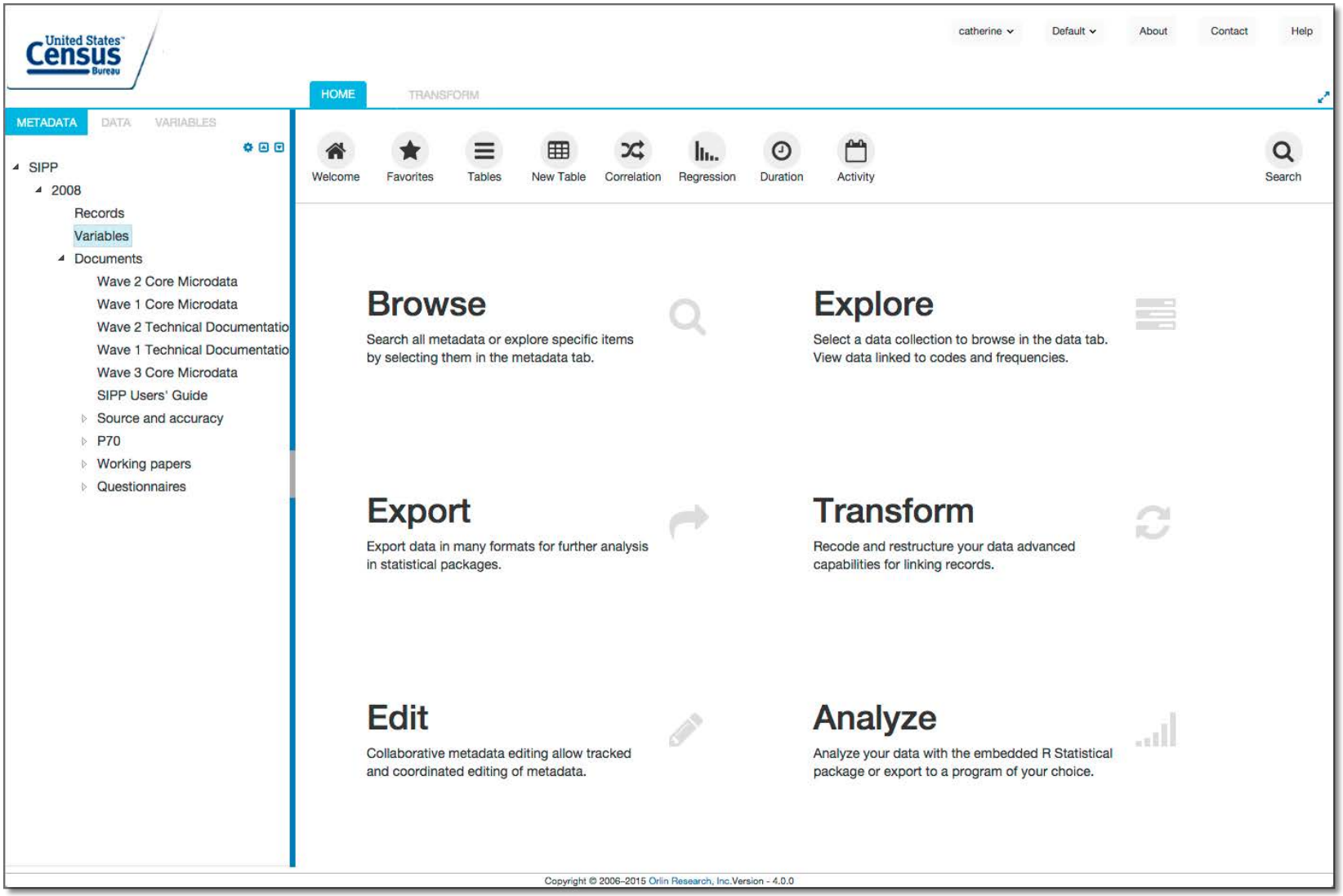
- The Orlin Longitudinal Data System (OLDS) is a tool that organizes SIPP data, creating the necessary linkages across persons and units and over time

- It allows the analyst to easily recode variables and manage data using these linkages and includes full linkages with SIPP metadata such as questionnaires and variable descriptions
  - Data and metadata search and exploration tools
  - Automatic tracking of relationships across records types and across time
  - Easy variable creation and modification

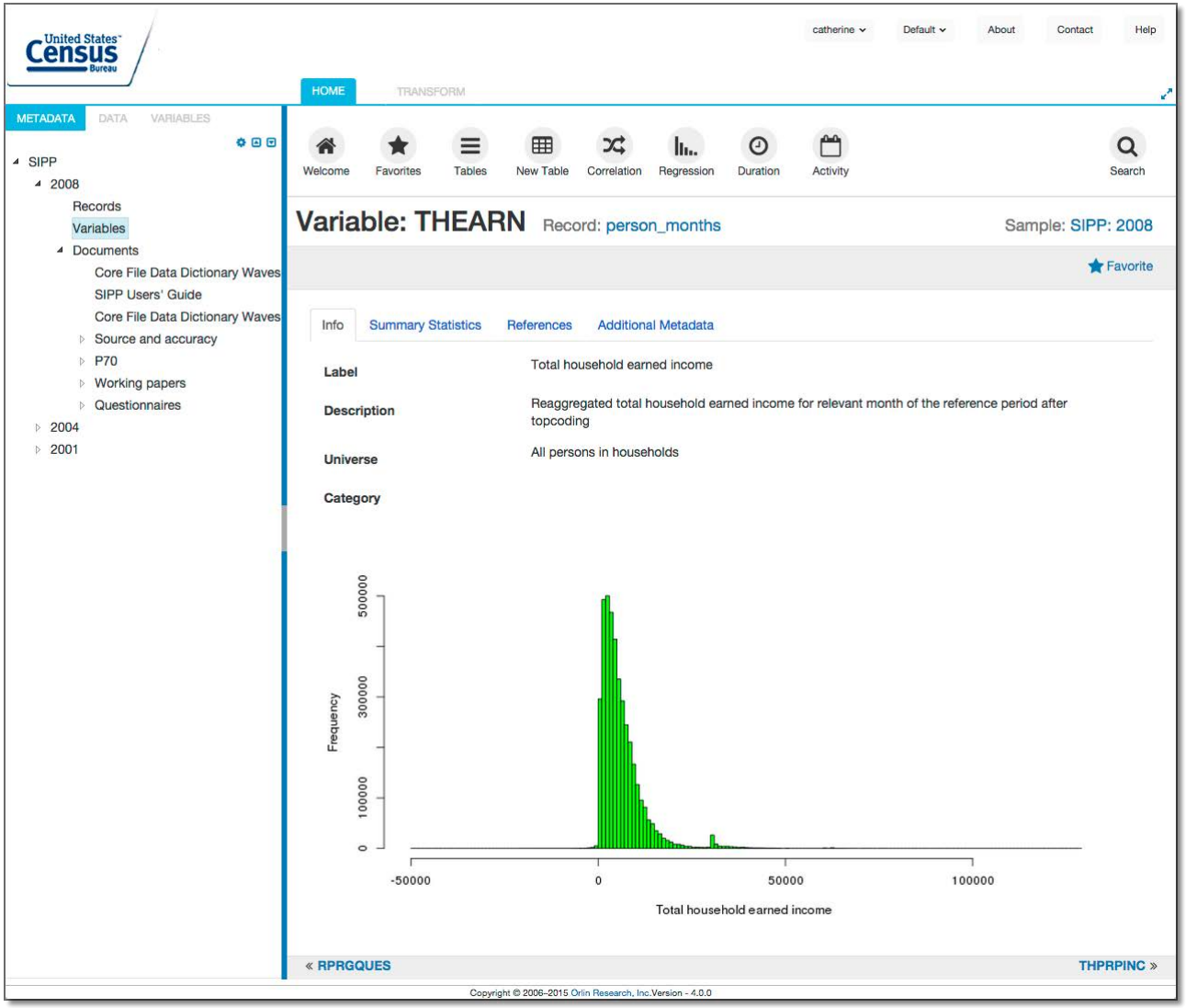
- Built-in analysis tools, using the R statistical language
  - Simple templates provided for each type of data manipulation or analysis
  - Complete audit trail, documenting all actions performed
  - Export of data in any format

- Orlin Tool has loaded 1996-2008 SIPP Panels
  - The 2008 SIPP panel: 60 months of information on 131,337 individuals
  - Core data and topical modules
  - Linked longitudinally at the person level

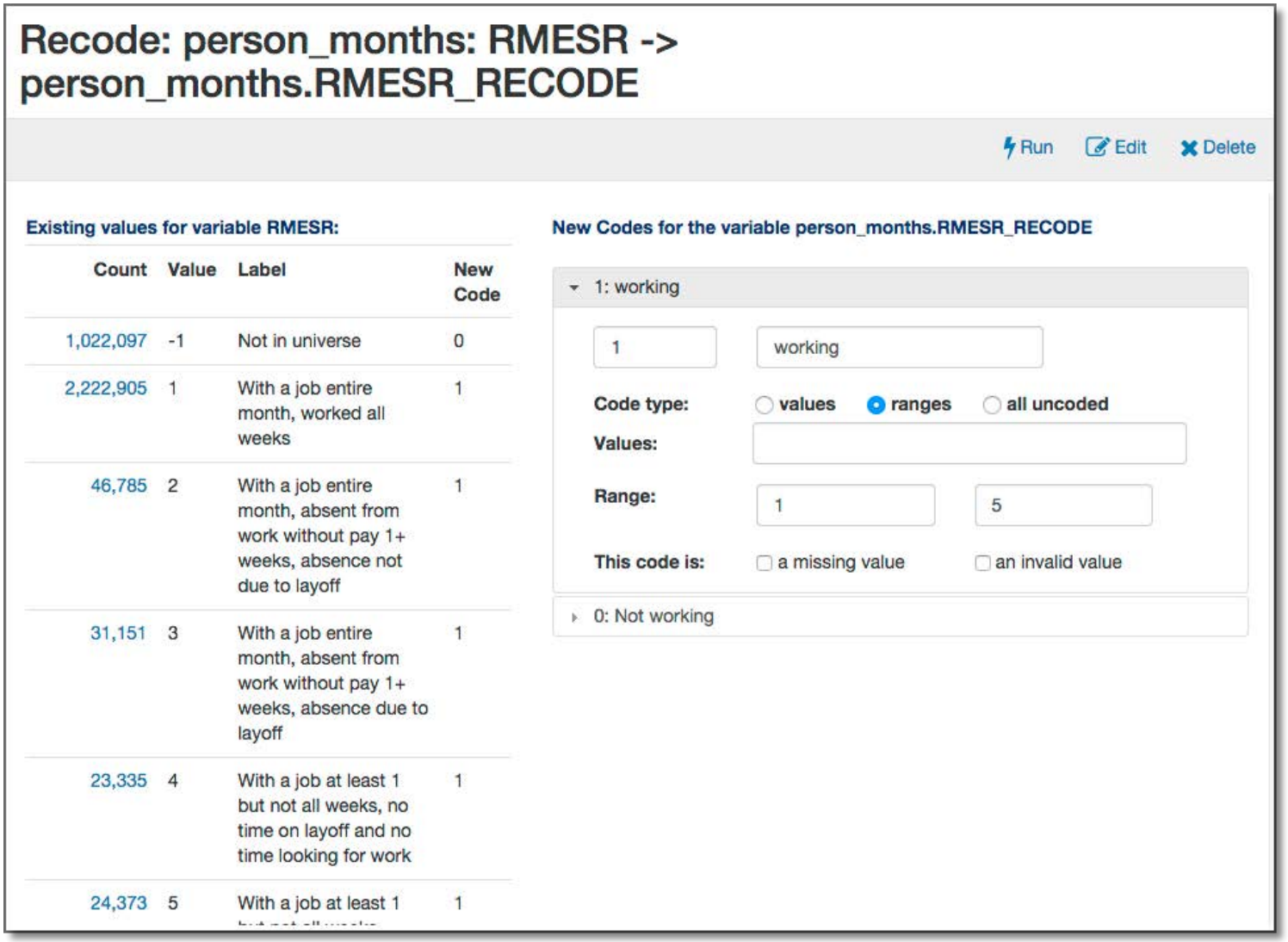
Introduction – The Interface



Continuous Variables

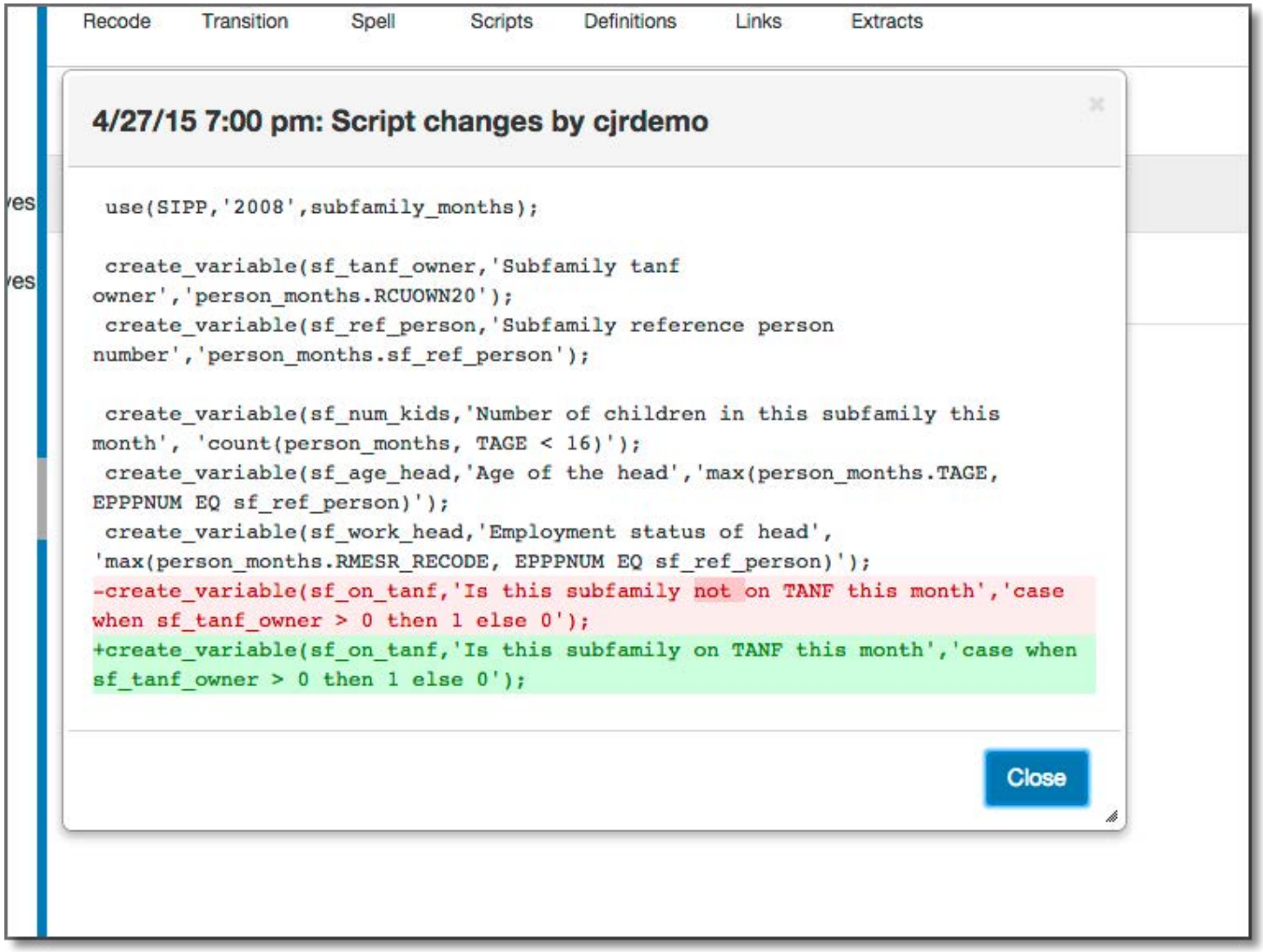


Let's Recode RMESR



And we get the stats for the new var

With the changes pointed out



Results

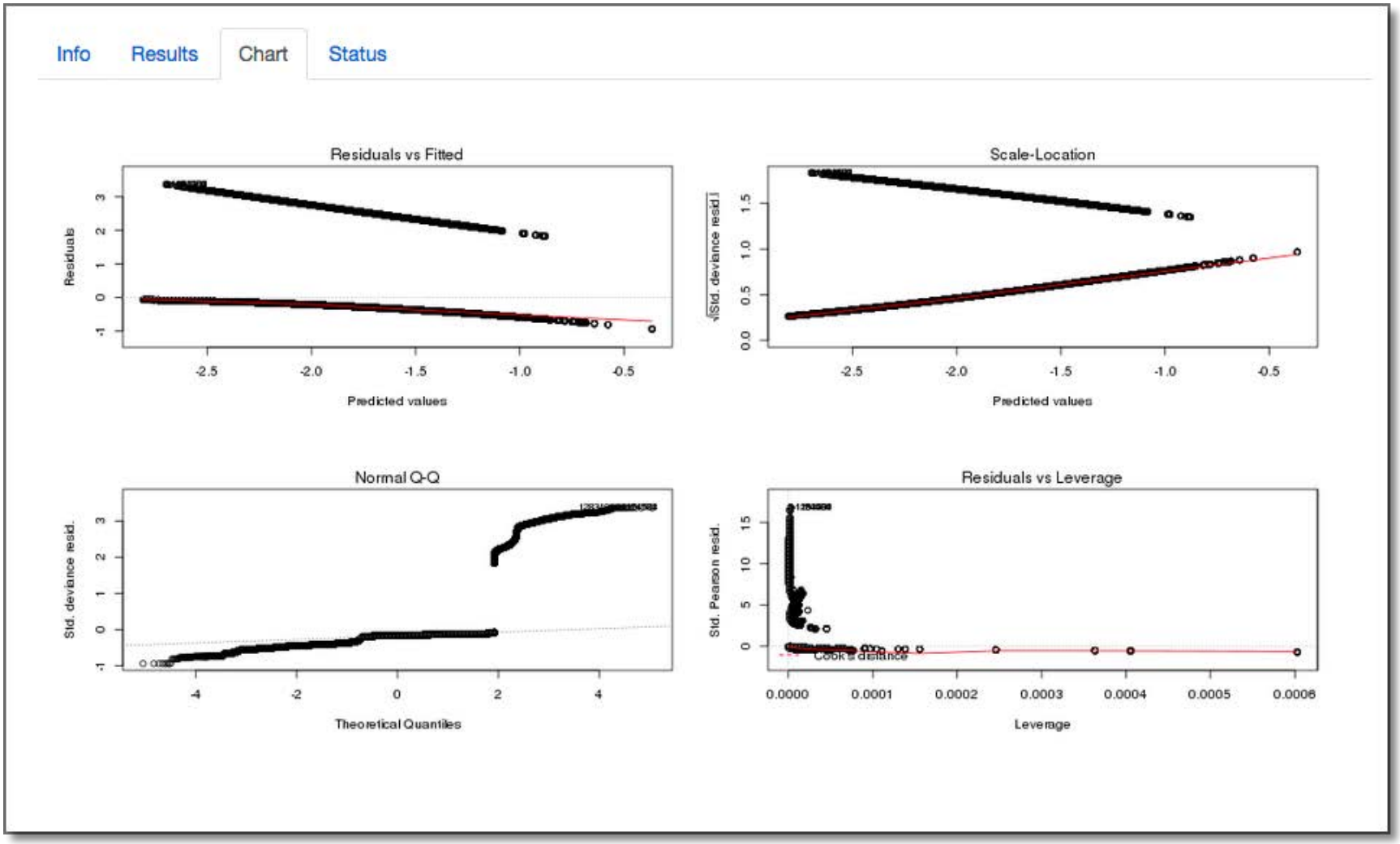
Create a regression - Regression for sf_on_tanf_u13

Run Edit Delete

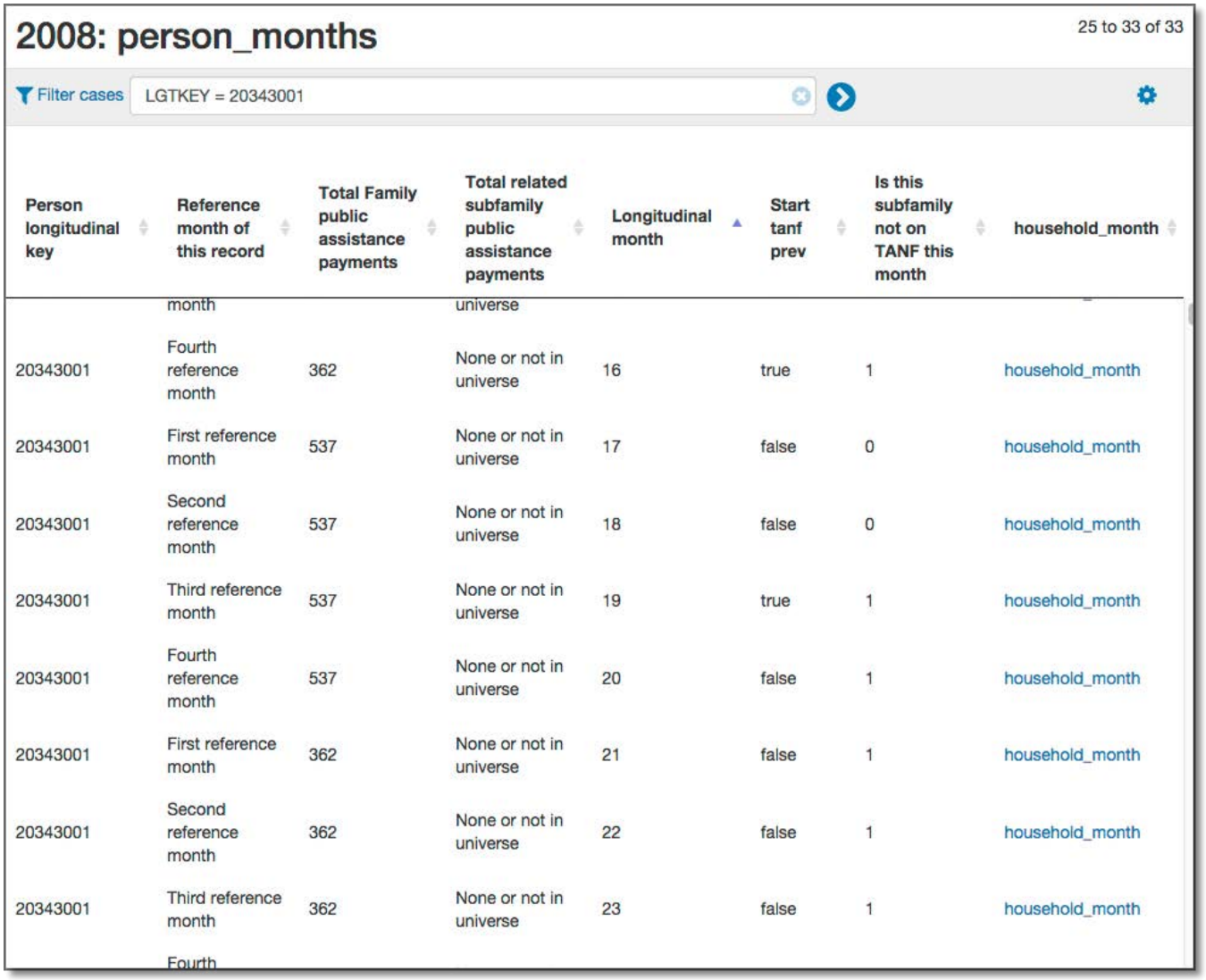
Info | Results | Chart | Status

```
Call:
glm(formula = sf_on_tanf_u13 ~ sf_num_kids_u13 + sf_age_head_u13 +
    sf_work_head_u13, family = binomial(link = "probit"), data = y)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9401  -0.2065  -0.1586  -0.1401   3.3617

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.3404158  0.0084722 -158.21   <2e-16 ***
sf_num_kids_u13   0.0668372  0.0010849   61.60   <2e-16 ***
sf_age_head_u13  -0.0090036  0.0002112  -42.63   <2e-16 ***
sf_work_head_u13 -0.7772969  0.0037101 -209.51   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 567999  on 2253572  degrees of freedom
Residual deviance: 512142  on 2253569  degrees of freedom
  (128479 observations deleted due to missingness)
AIC: 512150

Number of Fisher Scoring iterations: 7
```

And the chart



Transitions



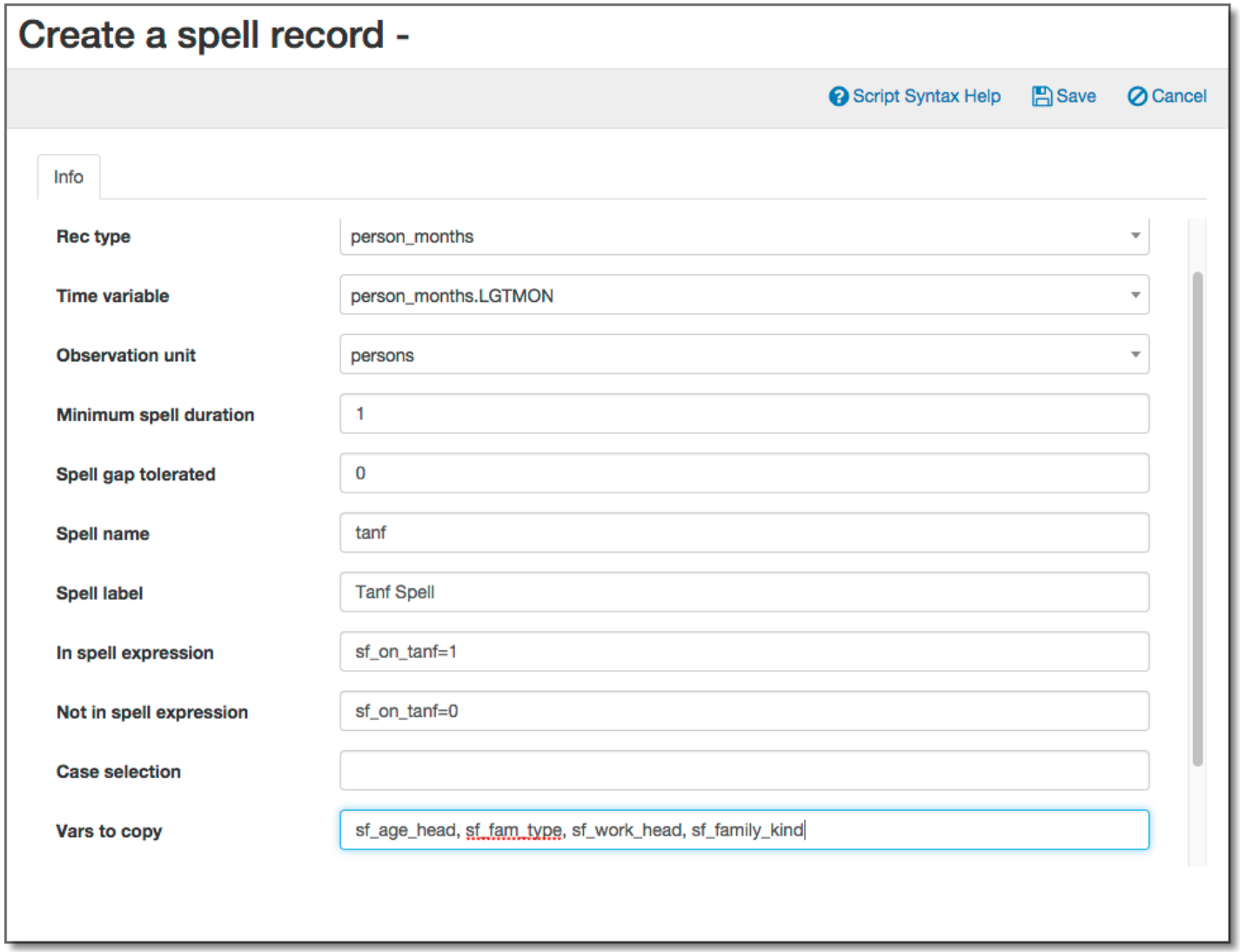
2008: person_months

25 to 33 of 33

Filter cases | LGTKEY = 20343001

Person longitudinal key	Reference month of this record	Total Family public assistance payments	Total related subfamily public assistance payments	Longitudinal month	Start tanf prev	Is this subfamily not on TANF this month	household_month
	month		universe				
20343001	Fourth reference month	362	None or not in universe	16	true	1	household_month
20343001	First reference month	537	None or not in universe	17	false	0	household_month
20343001	Second reference month	537	None or not in universe	18	false	0	household_month
20343001	Third reference month	537	None or not in universe	19	true	1	household_month
20343001	Fourth reference month	537	None or not in universe	20	false	1	household_month
20343001	First reference month	362	None or not in universe	21	false	1	household_month
20343001	Second reference month	362	None or not in universe	22	false	1	household_month
20343001	Third reference month	362	None or not in universe	23	false	1	household_month
	Fourth						

Creating a spell



THANK YOU!

Jason Fields
Survey Director
Survey of Income and Program Participation
National Survey of Children's Health
Jason.M.Fields@census.gov

www.census.gov
www.census.gov/sipp