Looking Beyond Level of Service: Using Behavioral Insights to Improve Taxpayer Experience

Presented by: Jan Millard
Jan Millard, RAAS; ASR Analytics LLC (Omar Faruqi, Jonah Flateman, Jamil Mirabito, Sarah Smolenski, Michael Stavrianos, Lauren Szczerbinski)
IRS TPC Joint Research Conference | June 22, 2023

The Customer Voice Portal (CVP) Message Redesign Pilot built upon IRS experience using Behavioral Insights to redesign Collection notices

Automated Collection System (ACS) Notice Redesign

IRS conducted a series of pilot tests to measure the benefit of redesigning Collection notices. Pilot test results showed using Behavioral Insights to design notices can achieve Collection's top three goals:

1. Improve taxpayer experience and understanding
2. Reduce IRS costs
3. Increase taxpayer compliance actions

Her Majesty's Revenue and Customs (HMRC) Announcement Redesign

The United Kingdom's tax authority increased taxpayer self-service by applying Behavioral Insights to revise recorded voice messages and encourage customers who can self-serve to go online while they are on a call waiting for an advisor.

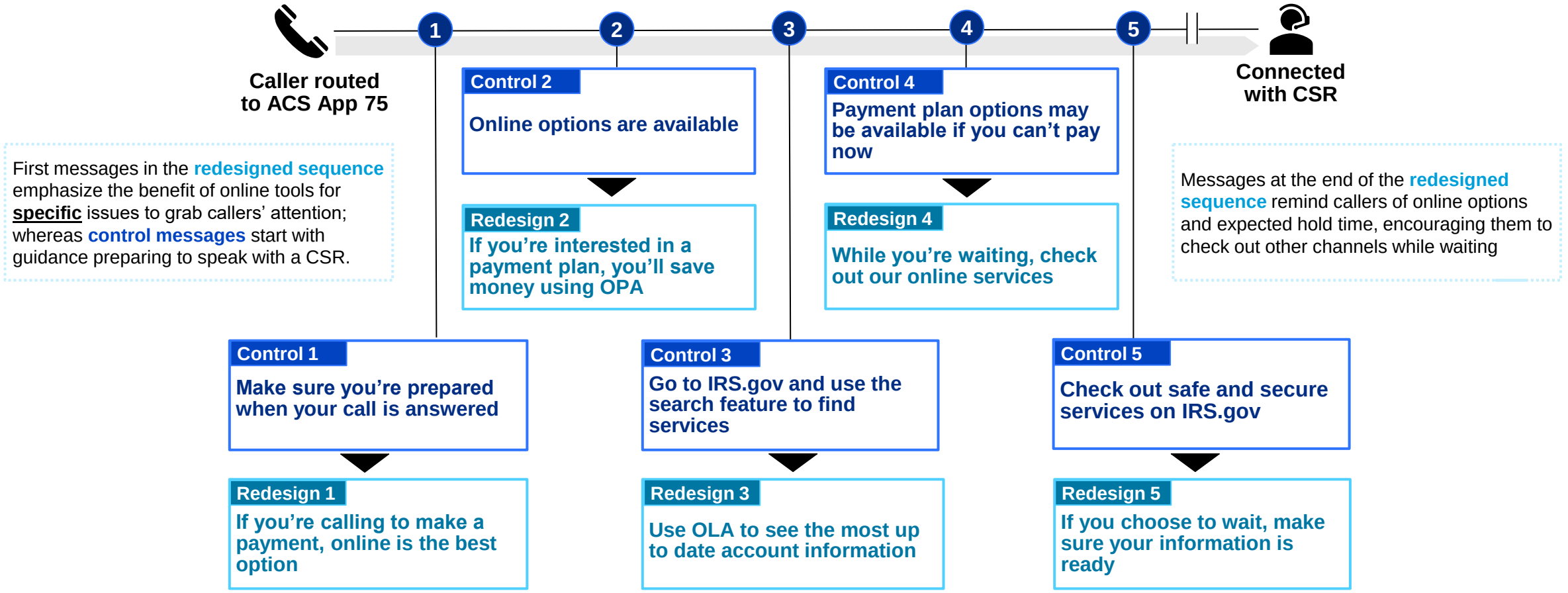
Customer Voice Portal Message Redesign

Callers routed to **ACS Applications 75 (IMF)** are played a **sequence of five message prompts** while waiting in queue to speak with a CSR. Some taxpayers call the IRS about issues which can be resolved using online self-service tools, saving them both time and money. IRS used Behavioral Insights to redesign CVP message sequences, informing callers of the benefit of online resources and freeing up phone resources for taxpayers with issues requiring CSR assistance.

IRS used Behavioral Insights to develop voice prompts that provide callers with information necessary to consider self-service channels to resolve issues

App 75 Queue Messages

Current Message Themes
Redesigned Prototype Message Themes



The pilot tested the effectiveness of redesigning voice prompts to nudge ACS callers to shift to IRS online services

CVP Message Redesign Pilot Goals

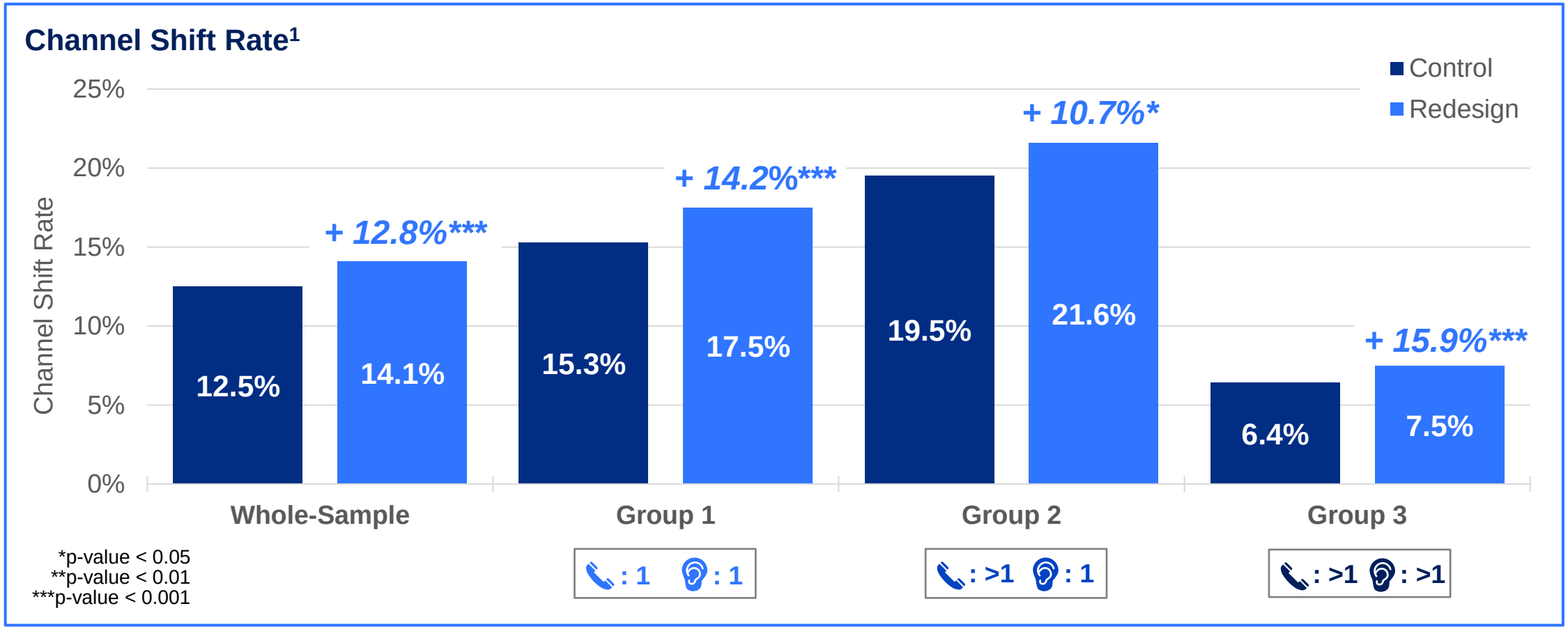
- **Increase Channel Shift**: Encourage taxpayers who can self-serve to hang up and use online resources than wait on hold for a CSR.
- **Increase Use of Online Services:** Enhance taxpayer experience by improving awareness and use of online resources relevant to their tax issue.
- **Improve Call Resource Allocation**: Reduce IRS costs by informing callers of online services and reducing CSR's phone time on issues that can be addressed online (e.g., obtain additional information) in favor of those requiring CSR support.

To analyze pilot results, callers were assigned to one of three groups based on the number of calls made during the test period

CVP Pilot Caller Group	Description	Pilot Callers
1 **Called once and heard at least one announcement in the message sequence** : 1 : 1	Group 1 callers were routed to App 75, remained on the line to hear at least the first announcement in the sequence, and did not call again during the pilot.	**Group 1** Redesign: 31,146 Control: 30,580
2 **Called multiple times and heard announcement(s) once** : >1 : 1	Group 2 called the IRS more than once during the pilot, however during only one call attempt were they on the line to hear at least one message in the sequence.	**Group 2** Redesign: 3,437 Control: 3,606
3 **Called multiple times and heard announcement(s) on more than one call** : >1 : >1	Group 3 callers heard at least one message in the sequence, called back at least once more and again heard at least one message in the sequence.	**Group 3** Redesign: 8,449 Control: 7,884

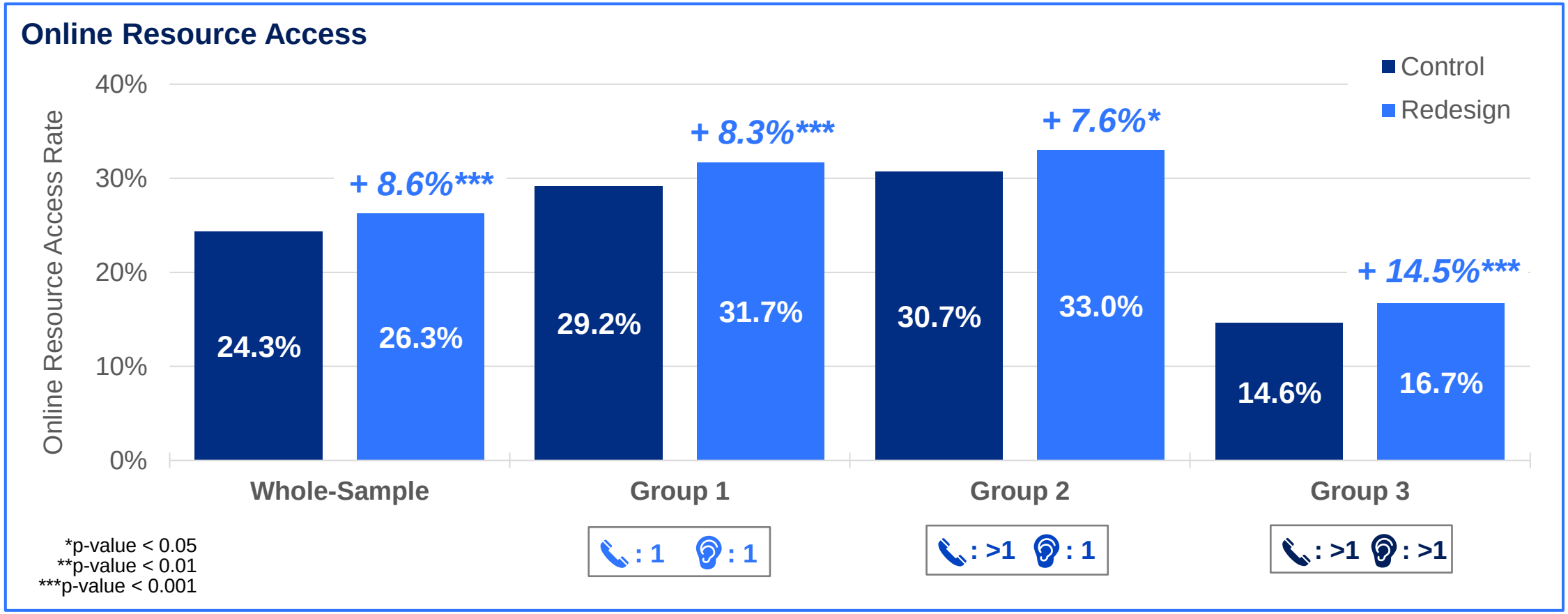
- **The total sample size consisted of 85,102 taxpayers and 103,512 calls**
- **Outcomes were compared between control and redesign groups and evaluated in the 30 days after the final pilot call**

Redesigned messages increased the channel shift rate relative to the existing message sequence for all caller groups

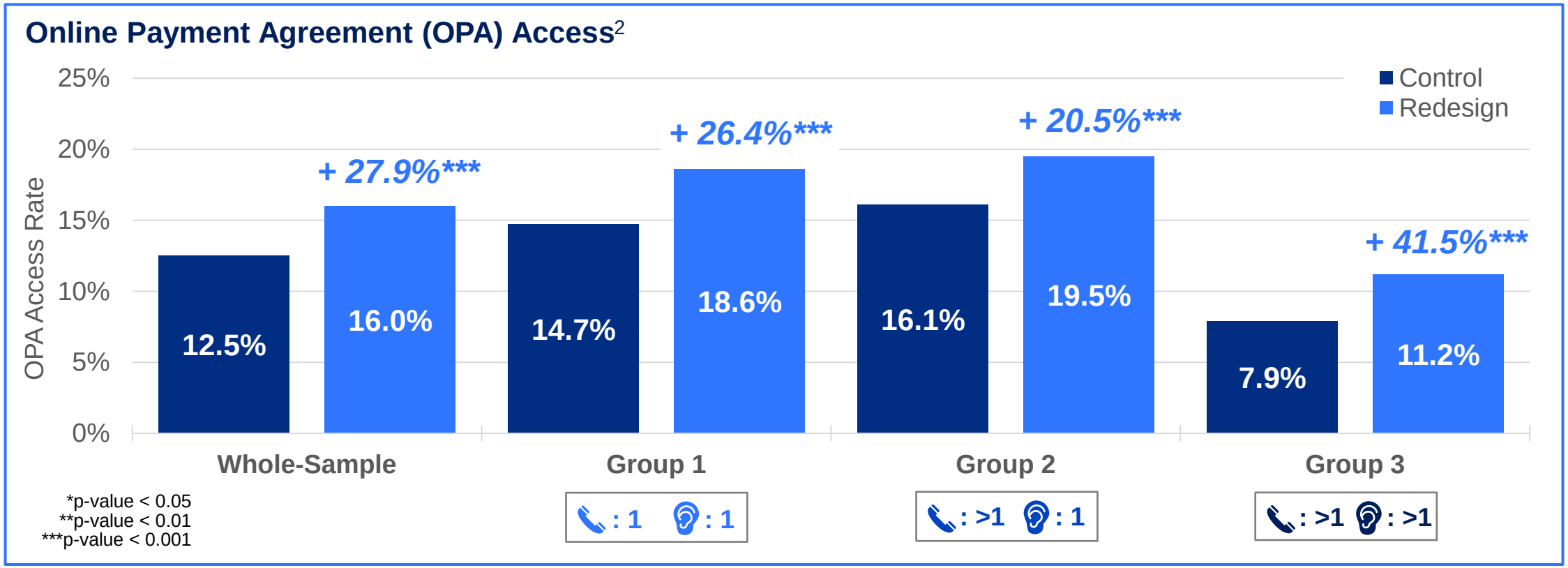


1. Among callers who channel shifted, roughly 59 – 70 percent channel shifted on the same day as their call

Redesigned messages increased online access among callers in all groups



The OPA access rate was significantly higher for callers who heard redesigned messages



If redesigned announcements were implemented at scale on App 75, monthly savings attributable to using OPA to set up a payment plan instead of over the phone would amount to $86,264 - $107,830.[1]

1. Taxpayers save between $76 - $95 by setting up or modifying a payment plan via OPA rather than over the phone.
2. OPA allows individuals and businesses with an outstanding balance in aggregate assessed tax, penalties, and interest, to request a payment plan. Eligibility for Short-Term Plan is balance less than $100K. Eligibility for Installment Agreement is balance less than $50K. 86,234 Calls from OPA-Eligible Taxpayers

Among callers abandoning in queue, a larger proportion of callers who heard redesigned messages abandoned after Message 2 than control callers

Proportion of Callers Who Abandon After Each Message in the Sequence
Pilot Callers who Abandoned in Queue

Group	Prototype	Last Message Heard				
		1	2	3	4	5
Group 1 (: 1 : 1)	Control	4.29%	9.32%	5.08%	2.81%	78.50%
	Redesign	6.36%	13.94%	4.19%	5.14%	70.37%
Group 2 (: >1 : 1)	Control	5.25%	10.23%	5.93%	2.57%	76.02%
	Redesign	6.45%	14.37%	4.81%	5.26%	69.10%
Group 3 (: >1 : >1)	Control	3.66%	8.15%	4.58%	2.77%	80.84%
	Redesign	5.70%	12.09%	4.26%	4.97%	72.99%

Most callers who abandoned before the final message in the sequence do so after the second message. However, a larger proportion of redesign callers abandoned after the second message compared to control callers.

The 2nd message in the redesign sequence informs taxpayers of **cost savings associated with establishing or modifying a payment plan online** rather than over the phone.
The second message in the control sequence informs taxpayers of **general** online payment options available at IRS.gov/payments

Callers who heard redesigned messages were more likely to abandon their call and spent less time waiting to connect with a CSR

Measures of Call Resource Allocation

Group	Prototype	Abandon Rate	Average Speed to Answer (ASA)
Group 1 : 1 : 1	Control	40.5%	88 mins
	Redesign	42.4%	85 mins 56 secs
	Relative Uplift	***+ 4.57%****	***- 2 mins 4 secs****
Group 2 : >1 : 1	Control	49.6%	89 mins 56 secs
	Redesign	51.4%	84 mins 31 secs
	Relative Uplift	***+ 3.63%***	***- 5 mins 27 secs****
Group 3 : >1 : >1	Control	50.9%	85 mins 34 secs
	Redesign	52.8%	85 mins 48 secs
	Relative Uplift	***+ 3.59%****	***+ 14 secs***

Abandon Rate: Callers in Groups 1 and 3 who heard the redesigned messages abandoned at a higher rate than those who heard control messages.

Average Speed to Answer: Callers in Groups 1 and 2 who heard redesigned messages waited, on average, 2 – 5.5 fewer minutes to connect with a CSR than callers who heard the control messages.

Understanding taxpayers' reasons for calling the IRS can inform further improvements to voice messages

- Understanding taxpayer motivations to speak with a CSR can allow for **tailoring of announcements to provide specific guidance for self-service resolution**
- Events occurring on taxpayer accounts (e.g., notice issued, phone call, online authentication, etc.) were evaluated in the 30 days leading up to a pilot call. This string of events is called a **taxpayer journey**.
- Taxpayer journeys were analyzed and segmented to **identify common events or combinations of events leading up to a phone call.**

Taxpayer Journey
Events occurring in 30 days leading up to Pilot call

Online Authentication + **Notice Issued** + **Payment**

Phone Call

Evaluating notice types issued to taxpayers suggests the type of notice issued could influence taxpayers' willingness to channel shift

- Over 60% of pilot taxpayers were sent at least one notice in the 30 days prior to their pilot call. CP504 was the most issued, followed by the CP14, CP90, LT11 and CP49.
- CP14 channel shift rates were highest among the five notices for both redesigned and control messages. CP49 channel shift was lowest for both redesigned and control messages.
- Taxpayers issued CP49 may prefer to connect with a CSR if the call queue messages did not reference the issue specific to the notice

Channel Shift Rate
Most Issued Notices Prior to Pilot Call

Notice Type	Prototype	Channel Shift Rate
CP504 *Final/3rd Balance Due*	Control	**14.8%**
	Redesign	**15.6%**
CP14 *Balance Due*	Control	**16.5%**
	Redesign	**20.1%**
CP90 *Final Notice – Levy, Right to CDP Hearing*	Control	**13.8%**
	Redesign	**15.5%**
LT11 *Final Notice – Notice of Intent to Levy*	Control	**13.1%**
	Redesign	**16.7%**
CP49 *Refund Applied to Other Tax Liability*	Control	**12.5%**
	Redesign	**14.1%**

Taxpayers sent multiple notices may face more complex issues and prefer to wait to speak with a CSR

- More than 50% of taxpayers issued multiple notices prior to calling remained in the queue to connect with a CSR
- Among taxpayers who received two notices, the most common were CP14 and CP504 balance due notices sent in the same 30-day window
- Notices containing conflicting information (e.g., different amount due or different due dates) may cause confusion or stress for taxpayers. Voice prompts could address issues of this nature by direct taxpayers to confirm how much they owe using Online Account.

Call Outcomes
Pilot Callers Issued More than 1 Notice 30 Days Prior to Call

# Notices Issued	Prototype	Call Outcome	
		Connected	**Abandoned**
Two Notices	Control	51.8%	47.0%
	Redesign	49.4%	48.7%
Three Notices	Control	55.6%	43.6%
	Redesign	54.2%	44.2%
Four or More Notices	Control	58.0%	40.2%
	Redesign	52.8%	45.1%

IRS uses LOS to evaluate its ability to answer taxpayer questions and assist taxpayers in meeting their tax obligations over the phone

$$Level\ of\ Service\ (LOS) = \frac{(CSR\ Answered + Automated\ Answered)}{(CSR\ Answered + Automated\ Answered + \mathbf{Abandoned} + Busy + Disconnected)}$$

Limitations

- LOS does not consider the quality of service provided. Aspects of caller experience, such as time to connect, utility of the call, or overall effort exerted in issue resolution, are not represented
- Redesigned CVP messages sought to improve awareness of available self-service tools for specific issues, empowering callers to decide whether to remain on hold to speak with a CSR or shift online. An increase in the number of callers who shift to self-service, increases the number of ***Abandoned calls***, which may reduce LOS.
- Increased rates of channel shift will negatively impact the LOS metric but will improve the taxpayer experience.

1. Congress requires IRS use LOS to evaluate call center performance – the metric is tied to IRS budget. TIGTA (June 12, 2019). Telephone Performance Measures Do Not Provide an Accurate Assessment of Service to Taxpayers. Page 29.

Incorporating additional measures of service may provide a more holistic view of the taxpayer experience

Level of Access (LOA) measures the proportion of calls received within business hours connected with a CSR.	$$LOA = \frac{(CSR\ Answered + Automated\ Answered)}{(CSR\ Answered + Automated\ Answered + Abandoned + Busy + Disconnected) - x}$$ *x = # calls received outside of business hours*
Average Speed to Answer (ASA) quantifies the amount of time spent waiting to connect with a CSR.	$$ASA = \frac{\sum(Time\ Spent\ in\ Queue\ for\ Connected\ Calls_i)}{Total\ Connected\ Calls}$$
First Contact Resolution (FCR) measures the proportion of taxpayer engagements resulting in resolution without high-touch follow-up events (e.g., phone calls, TAC visits).	$$FCR = \frac{\#\ Taxpayers\ with\ Issue\ Resolved\ on\ First\ Call}{Total\ Connected\ Calls}$$
Taxpayer Effort (TE) estimates effort required to resolve issues; considers all possible channels to engage IRS and assigns weights associated with perceived effort required	$$TE = (1 * OLS) + (2 * Mail) + (3 * Connected\ Calls) + (4 * TAC) + (4 * TAS)$$
Effort to Serve (ETS) evaluates IRS effort required to assist taxpayers with resolving issues	$$ETS = (17 * Mail) + (41 * Connected\ Calls) + (67 * TAC)$$

The effect of redesigned CVP messages is positively captured by measures like Taxpayer Effort and Effort to Serve

Estimated Taxpayer Effort (TE)
30 Days Following 1st Pilot Call

Group	Prototype	Average TE
Group 1 : 1 : 1	Control	3.35
	Redesign	3.27
	Relative Uplift	***- 2.34%****
Group 2 : >1 : 1	Control	2.19
	Redesign	2.19
	Relative Uplift	***- 0.10%***
Group 3 : >1 : >1	Control	3.85
	Redesign	3.67
	Relative Uplift	***- 4.62%****

Redesigned messages significantly decreased the estimated TE for Group 1 and 3 callers. Callers in the redesign group were more likely to abandon and self-serve online, requiring less effort than waiting to connect with a CSR.

Estimated Effort to Serve (ETS)
30 Days Following 1st Pilot Call

Group	Prototype	Average ETS
Group 1 : 1 : 1	Control	30.8
	Redesign	29.7
	Relative Uplift	***- 3.42%****
Group 2 : >1 : 1	Control	17.9
	Redesign	17.3
	Relative Uplift	***- 3.44%***
Group 3 : >1 : >1	Control	38.3
	Redesign	36.3
	Relative Uplift	***- 5.26%****

Redesigned messages significantly decreased IRS ETS for Group 1 and 3 callers. Callers in the redesign group were more likely to abandon and self-serve online, reducing CSR effort to serve these taxpayers

Conclusions

Applying Behavioral Insights principles to enhance IRS call queue voice prompts can benefit both taxpayers and the Service.

Continuing to develop understanding of taxpayer motivations for calling the IRS can inform further improvements to voice messages.

Using a combination of metrics can offer a more complete picture of the impact of IRS efforts to serve taxpayers and insight into the level of effort require to resolve certain issues.

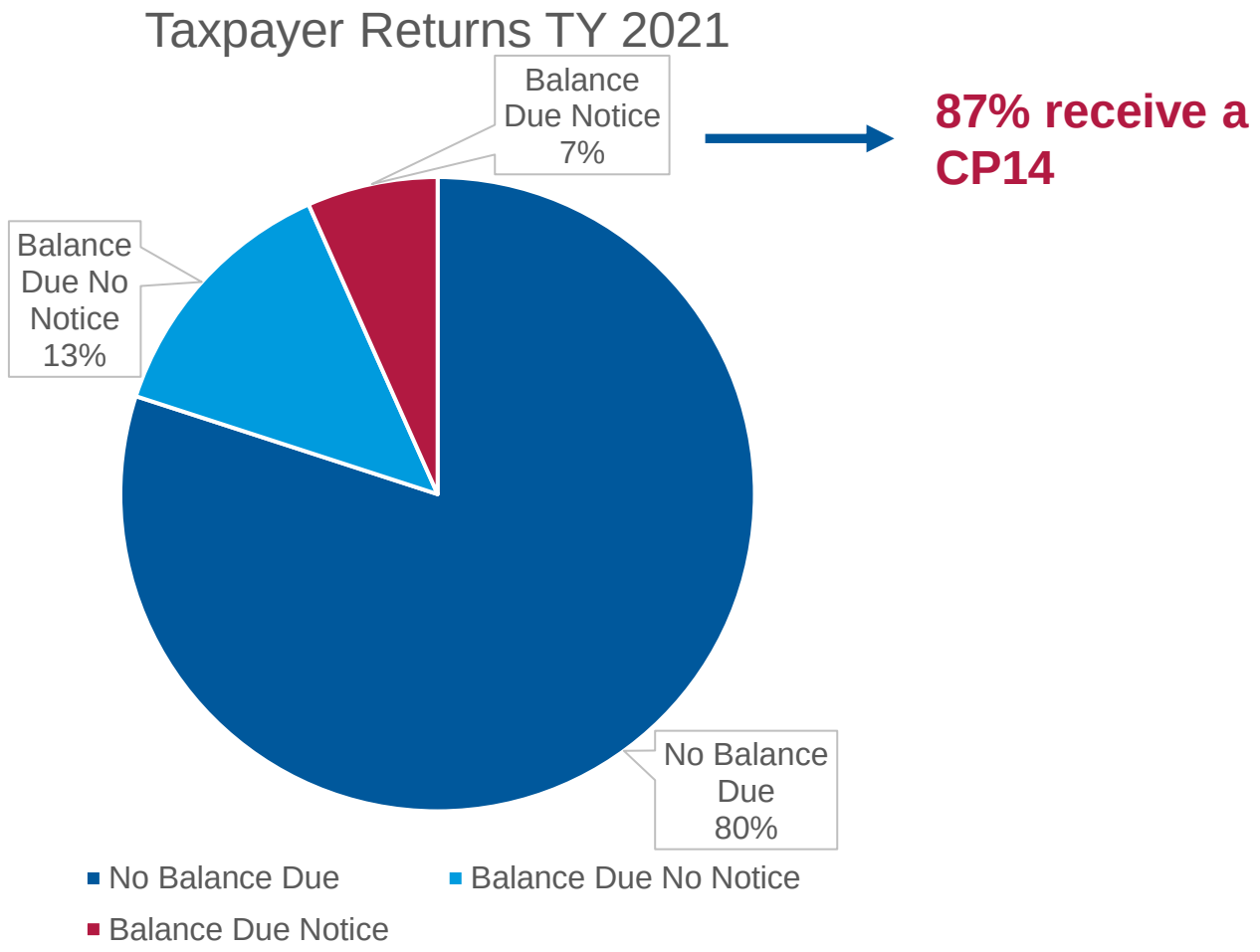
IRS-TPC Research Conference

The Balance Due Taxpayer: Reducing IRS Cost and Taxpayer Burden

June 22, 2023

Balance Due Returns and Notices

- **20% of all returns for TY2021 had a balance due**
- **7% received balance due notice OF SOME TYPE**
- **87% of those received a CP14 Balance Due Notice**
- **CP14 is sent to taxpayers who do not fully pay the amount due or set up an installment agreement by the filing deadline**



87% receive a CP14

Cost to Resolve CP14

The IRS issues approximately 7.5 million CP14 notices per year resulting in multiple downstream costs to resolve.

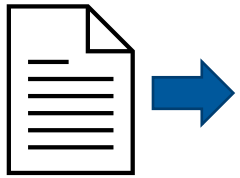
CP14 Issuance	Count	Cost (Per)	Total
CP14	7.5 M	$.51	$3,825,000

Outcome	Count	Cost (Per)	Total
Full Pay	1.2 M	-	-
Installment Agreement	3.3 M	$6.12	$20,196,000
Ignore (Receive CP501)	2.6 M	$0.51	$1,326,000
Call	900,000	$72.73	$65,457,000

Other Outcomes	Cost (Per)
Taxpayer Assistance Center	$251.38
Written Response	$95.47

Where do we want to intervene?

Current Process



Change the notice

- Change format
- Change fonts
- Change language
- Change tables
- Include web links

Improve outcomes

- Increase full pay rate
- Increase IA rate
- Decrease subsequent notices
- Decrease call volume
- Increase web-based tool use

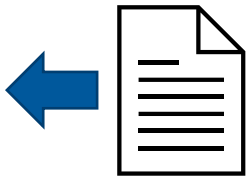
Common theme: The balance due has already occurred and we are remediating the issue.

Change where we focus

- Identify the causes of balance due returns
- Develop strategies to mitigate the causes

Improve outcomes

- Decrease balance due notices
- Decrease downstream cost
  - Calls
  - Notices
  - Other enforcement activity
- Decrease amount owed



Proposed Process

Common theme: Prevent the balance due from occurring through early intervention.

Identifying Balance Due Populations

Three balance due categories

Balance Due Change TY16 to TY17	Count	% of Total
No Change	**110,688,000**	**82.6%**
Refund/even to refund/even	97,512,000	72.8%
Balance due without a CP14 to balance due without a CP14	10,904,000	8.1%
Balance due with a CP14 to balance due with a CP14	2,272,000	1.7%
Favorable Shift	**10,718,000**	**8.0%**
Balance due without a CP14 to refund/even	7,619,000	5.7%
Balance due with a CP14 to refund/even	1,922,000	1.4%
Balance due with a CP14 to balance due without a CP14	1,177,000	0.9%
Unfavorable Shift	**12,539,000**	**9.4%**
Refund/even to balance due without a CP14	9,132,000	6.8%
Refund/even to balance due with a CP14	2,221,000	1.7%
Balance due without a CP14 to balance due with a CP14	1,186,000	0.9%
Total (excludes unknown who filed in TY16 but not in TY17)	**133,945,000**	

Percentages reported may not equal 100% due to rounding

Balance Due Category Changes

Using Chi Square tests of independence and Cramer's V to calculate effect sizes, key triggers for unfavorable balance due changes emerged.

1040 Characteristics	Balance Due Category Change
Filing Status	Marriage -> favorable balance due change
	Divorce -> unfavorable balance due change
Schedule A	Adding Schedule A -> favorable balance due change
	Removing Schedule A -> unfavorable balance due change
Schedule C	Removing Schedule C -> favorable balance due change
	Adding Schedule C -> unfavorable balance due change
Schedules B/D/H	No consistent relationships discovered
Age	Both unfavorable and favorable change increased as age increased
Total Positive Income	Both unfavorable and favorable change increased as total positive income increased

All Chi Square tests of independence were statistically significant, and we followed them with effect size calculations using Cramer's V. Using standard effect size classifications from social science, all effect sizes are categorized as small but noteworthy.

Schedule A : Itemized Deductions, **Schedule B**: Interest and Dividends, **Schedule C**: Profit or Loss from a Business, **Schedule D**: Capital Gains and Losses, **Schedule H**: Household Employees

Risk of Unfavorable Balance Due Changes

Holding age constant and adjusting for total positive income, risk of an unfavorable change was calculated using logistic regression.

Divorce	Removing Schedule A	Adding Schedule C
Dramatic and consistent	Not as dramatic as divorce but impacts more taxpayers	Not as dramatic as divorce but impacts many more taxpayers
Risk of unfavorable change: • Slightly more than **triple the risk of other taxpayers**	Risk of unfavorable change: • Slightly less than **double the risk of other taxpayers**	Risk of unfavorable change: • **Double the risk for taxpayers who had Schedule C in both years or added it in TY17**
Between TY16 and TY17: • 578,000 taxpayers divorced • 7.8% (45,000) had an unfavorable change and issued CP14	Between TY16 and TY17: • 5 million taxpayers removed Schedule A • 4.1% (205,000) had an unfavorable change and issued CP14	Between TY16 and TY17: • 23 million taxpayers added Schedule C in TY17 or had it in both TY16 and TY17 • 5.2% (1.2 million) had an unfavorable change and issued CP14

Debt Ratio

Median Debt Ratio and Median Debt Ratio Difference	TY 2016 (%)	TY 2017 (%)	Difference
No Change			
Refund/even to refund/even	-5.17%	-4.30%	+0.87
Balance due without a CP14 to balance due without a CP14	2.53%	2.63%	+0.10
Balance due with a CP14 to balance due with a CP14	4.91%	4.57%	-0.34
Favorable Shift			
Balance due without a CP14 to refund/even	1.75%	-2.11%	-3.85
Balance due with a CP14 to refund/even	2.97%	-2.62%	-5.59
Balance due with a CP14 to balance due without a CP14	4.72%	3.86%	-0.86
Unfavorable Shift			
Refund/even to balance due without a CP14	-2.29%	1.63%	+3.92
Refund/even to balance due with a CP14	-2.95%	2.72%	+5.67
Balance due without a CP14 to balance due with a CP14	3.93%	4.40%	+0.47

Note: Negative debt ratios indicate refunds and positive debt ratios indicate balance due

Preventing Unfavorable Balance Due Change

How do we prevent the balance due from occurring? We focus on side effects and activity.

Divorce	Removing Schedule A	Adding Schedule C
• Loss of dependents • 401K withdrawal • Loss of Sch A (mortgage deduction) • Add Sch C (started a side hustle)	• Mortgage deduction • Medical expenses	• Gig economy • Side hustle • Taxpayer fails to account for no withholding

Research, Applied Analytics and Statistics (RAAS) 2021 Comprehensive Taxpayer Attitude Survey finds

- 66% mostly or completely agree with the statement: “I trust the IRS to help me understand my tax obligation”
- 86% mostly or completely agree that “the more information and guidance the IRS provides, the more likely people are to correctly file their tax returns.”

Gap Analysis

Online and other searches on divorce, starting a business, and working in the gig economy provide limited or no guidance to taxpayers naïve of the tax implications of significant life events.

IRS.gov	Google	In-person Support
• Divorce • Starting a small business • Gig Economy	• Divorce • Getting divorced • Starting a new business • How to start a new business • Driving for Uber • Driving for Lyft • Independent contractor	• Divorce attorney organizations • Divorce support groups • Tax preparation organizations, CPAs, and accountants • Tax workshops
Several publications • Technical nature caters primarily to tax professionals and those with knowledge of filing taxes • IRS.gov front page does not specifically address the issue of avoiding a balance due.	• Minimal or no guidance to naïve taxpayers • Adding “and taxes” generates somewhat more helpful information • No “early intervention” guidance • Must know specific keywords to generate useful results	• Provide links for local support groups • Provide links for tax preparation workshops

Intervention

The Taxpayer Experience Office (TXO) is using the results of this study to develop data-driven interventions to help taxpayers avoid a shift to an unplanned balance due.

Develop IRS.gov/divorce

- A landing page for divorced taxpayers and one-stop shop where taxpayers and tax professionals access divorce related tax material.

Develop new material

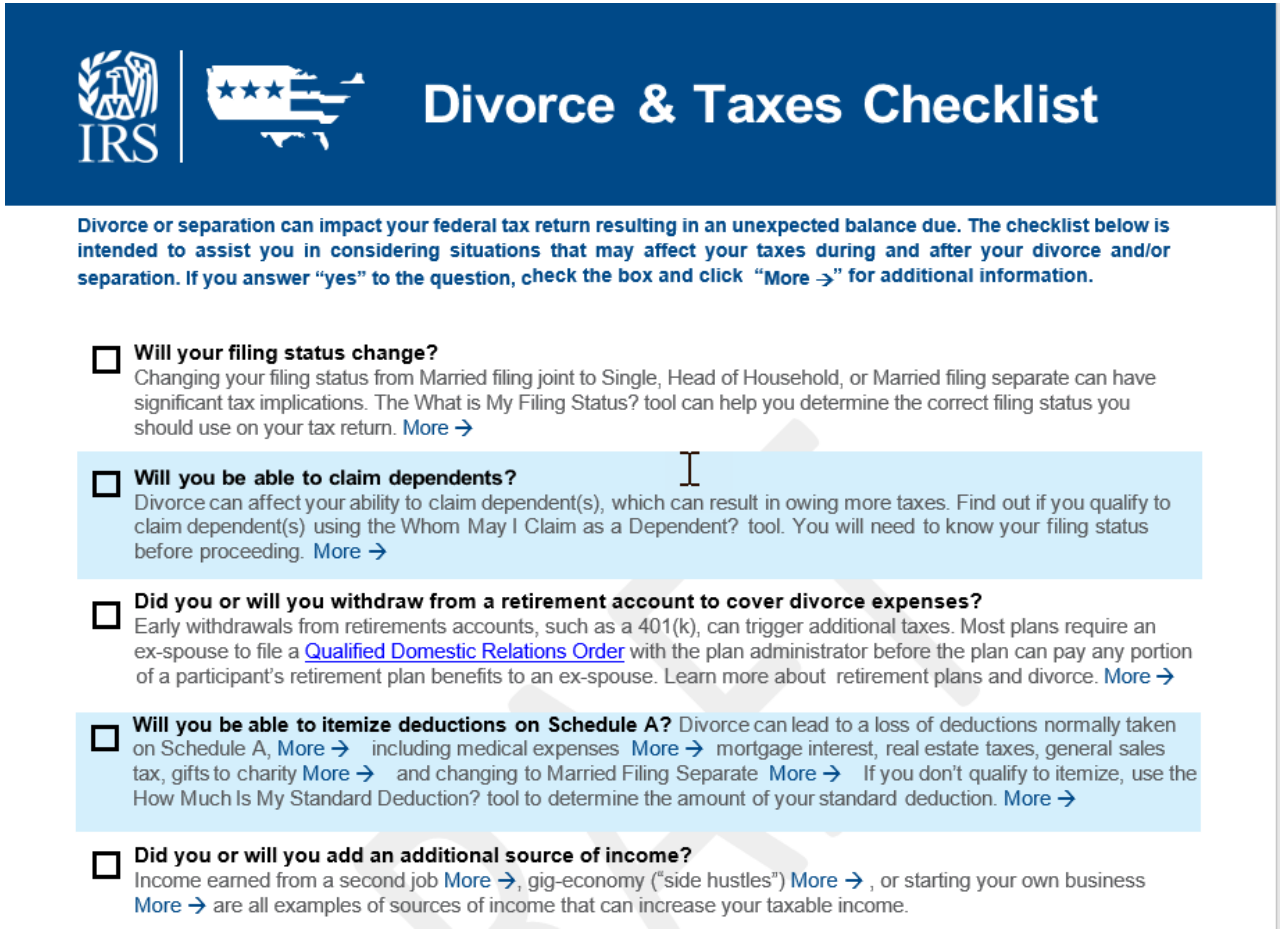
- One-page flyers such as “How to not owe taxes after a divorce” or “5 things to know about divorce and taxes” can grab attention avoid a balance due prior to filing.

Develop an external communication campaign

- Share content through social media, online (IRS.gov) and/or directly with partners.
- Leverage external networks and technology to develop an outreach campaign to drive traffic to IRS.gov/divorce.

Intervention

Divorce and taxes checklist



IRS

Divorce & Taxes Checklist

Divorce or separation can impact your federal tax return resulting in an unexpected balance due. The checklist below is intended to assist you in considering situations that may affect your taxes during and after your divorce and/or separation. If you answer "yes" to the question, check the box and click "More →" for additional information.

☐ **Will your filing status change?**
Changing your filing status from Married filing joint to Single, Head of Household, or Married filing separate can have significant tax implications. The What is My Filing Status? tool can help you determine the correct filing status you should use on your tax return. More →

☐ **Will you be able to claim dependents?**
Divorce can affect your ability to claim dependent(s), which can result in owing more taxes. Find out if you qualify to claim dependent(s) using the Whom May I Claim as a Dependent? tool. You will need to know your filing status before proceeding. More →

☐ **Did you or will you withdraw from a retirement account to cover divorce expenses?**
Early withdrawals from retirements accounts, such as a 401(k), can trigger additional taxes. Most plans require an ex-spouse to file a Qualified Domestic Relations Order with the plan administrator before the plan can pay any portion of a participant's retirement plan benefits to an ex-spouse. Learn more about retirement plans and divorce. More →

☐ **Will you be able to itemize deductions on Schedule A?** Divorce can lead to a loss of deductions normally taken on Schedule A, More → including medical expenses More → mortgage interest, real estate taxes, general sales tax, gifts to charity More → and changing to Married Filing Separate More → If you don't qualify to itemize, use the How Much Is My Standard Deduction? tool to determine the amount of your standard deduction. More →

☐ **Did you or will you add an additional source of income?**
Income earned from a second job More →, gig-economy ("side hustles") More → , or starting your own business More → are all examples of sources of income that can increase your taxable income.

Conclusion

Evidence-based intervention

- Identify area of concern
- Discover at-risk populations
- Partner with stakeholders to create and implement targeted solutions

Contact Information

- Howard Rasey

  Howard.W.Rasey@IRS.gov

- Shannon Murphy

  Shannon.Murphy@IRS.gov

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

Understanding Yearly Changes in Family Structure and Income and Their Impact on Tax Credits

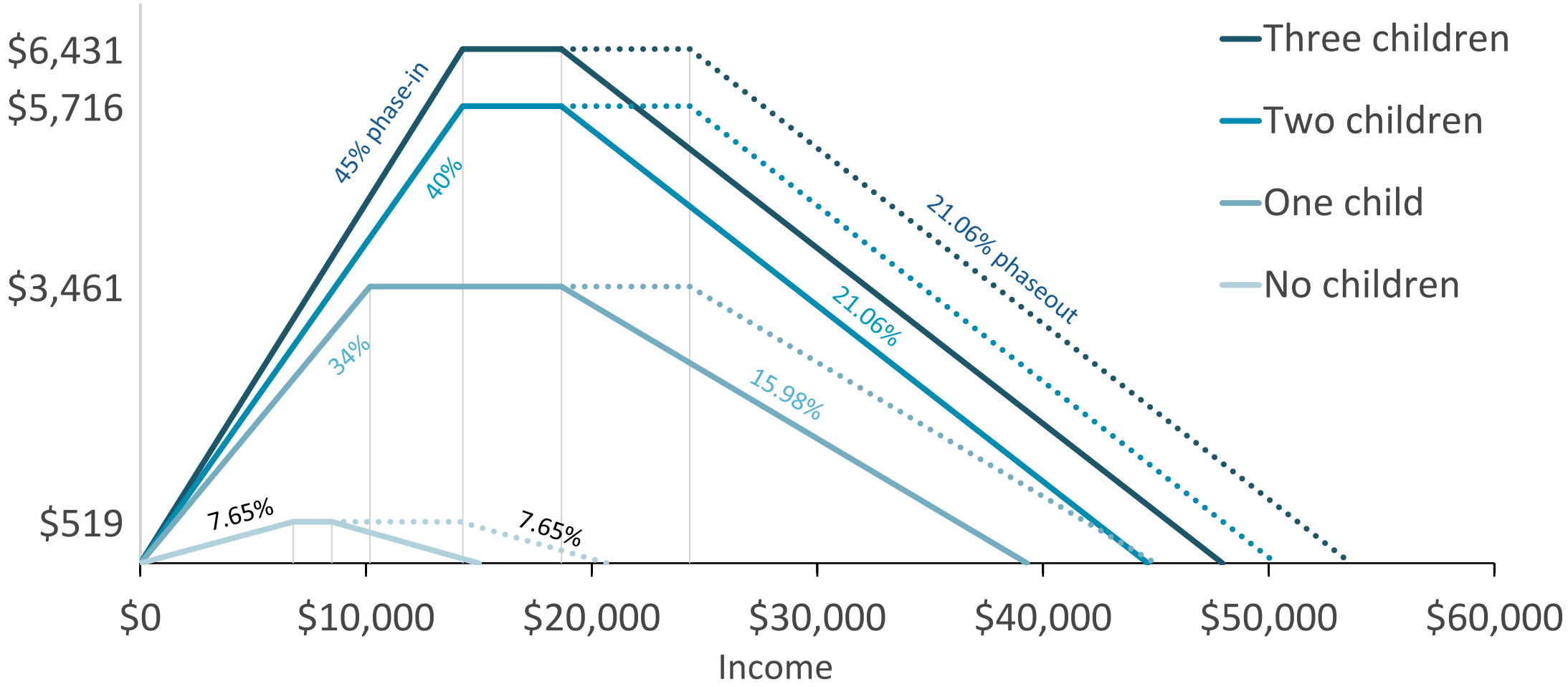
Considerations for Advancing Tax Credits

June 2023 | IRS-TPC Joint Research Conference on Tax Administration | Nikhita Airi | nairi@urban.org | @NikhitaAiri

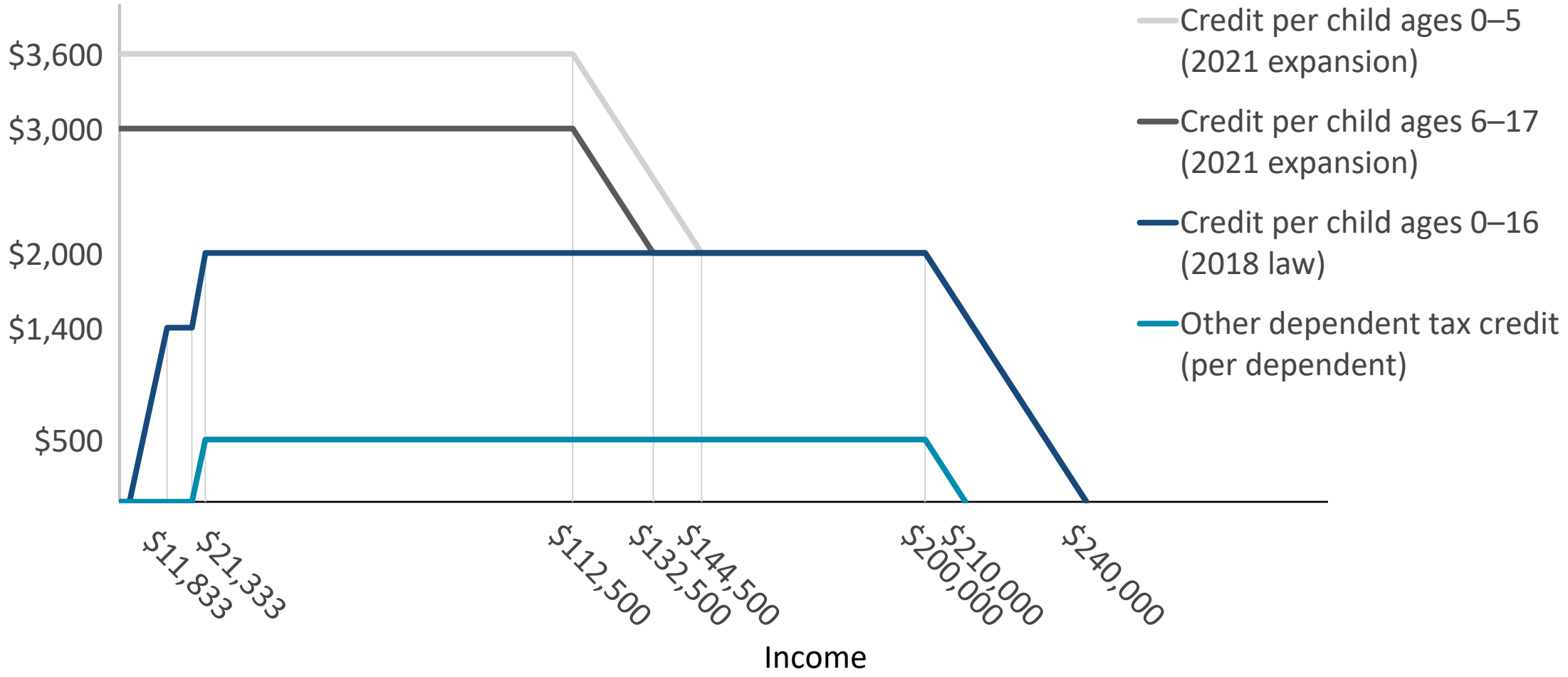
Refundable credits a larger source of income for low-income families

Source: Analysis of 2018 SIPP, Wave 1

Earned Income Tax Credit varies by number of kids, filing status, and income



Child Tax Credit varies by number of kids, filing status, and income



Refundable credits: a large, but unpredictable source of income

- Not all families understand tax system; feel **surprised by tax refund amounts**

- Low-income families have **more volatile incomes,** and increasingly, **complicated tax filing situations**
  - 64% of low-income adults' income spikes above or dips below their average at least one month a year (Maag et al 2017)
  - 60% of low-income families have tax filing ambiguities compared to 40% overall (Michelmore and Pilkauskas 2022)

Lump sum or advanced payments?

- Lump sum and advanced payments help with different forms of hardship (Parolin et al 2022)
- Receiving tax refunds associated with increased doctors' visits and college enrollment (Manoli 2018, Hamad 2019)
- Advanced payments can smooth income, cover day-to-day expenses
- **How accurately can we advance payments and how should overpayments of advance tax credits be resolved?**

Data

- **CPS-ASEC** collects income data from certain households in two consecutive years

- Subset: households with kids during at least one of two years from 2015-2016, 2016-2017, and 2017-2018 waves

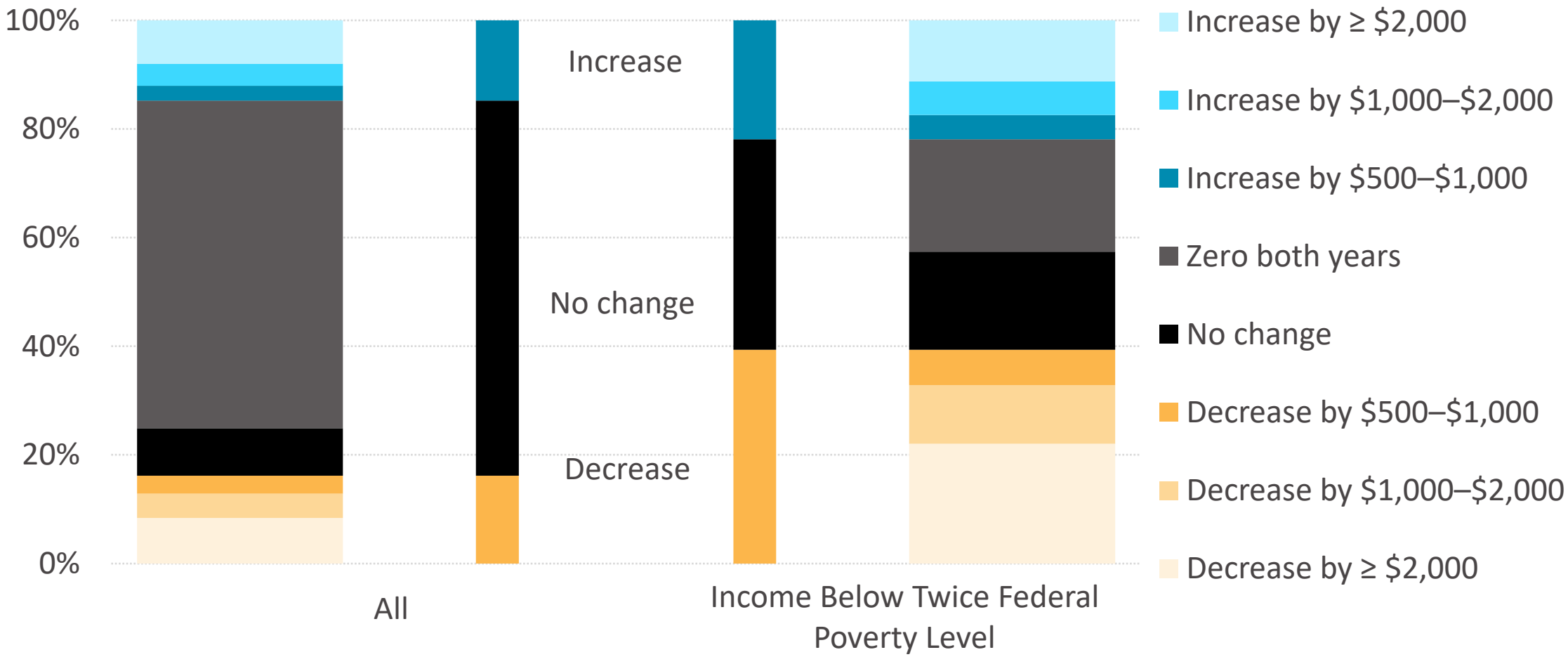
Methods

- Estimated EITC and CTC for households using TRIM3 microsimulation model
- Applied 2018 tax law to all years

- Counted changes in tax credits of at least $500
- Low-income = double federal poverty line

Results: Earned Income Tax Credit

Magnitude of EITC Changes by Income

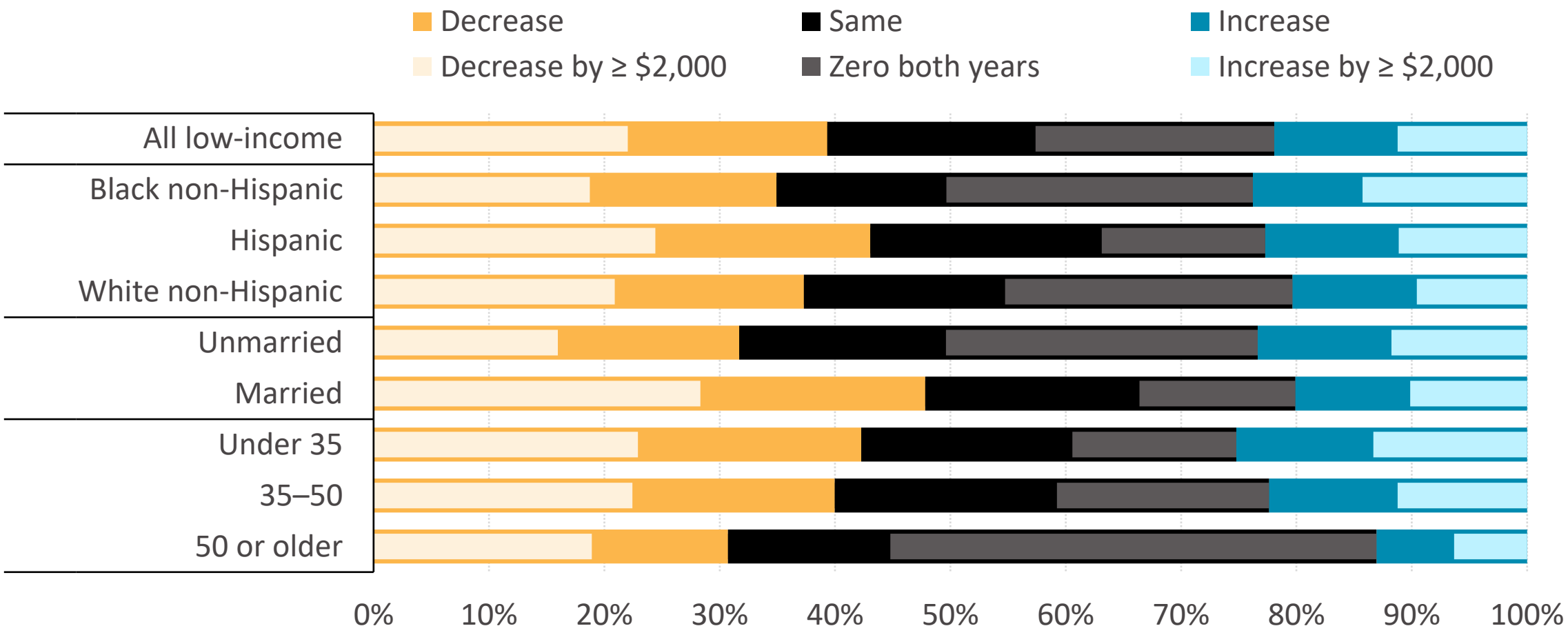


Source: Urban Institute TRIM3 model using data from Current Population Survey Outgoing Rotation Groups 2015–18.
Note: Sample includes households with one dependent child under age 18 in either year. No change is defined as a change of less than $500.

Income drives EITC changes for low-income families

- **Decreases (39%)**
  - **Increase in earnings (28%)**
  - Number of children decreased (5%)
  - Decrease in earnings (6%)
- Stays the same (39%)
- Increases (22%)
  - Income decreased (9%)
  - Income increased (8%)
  - Number of children increased (5%)

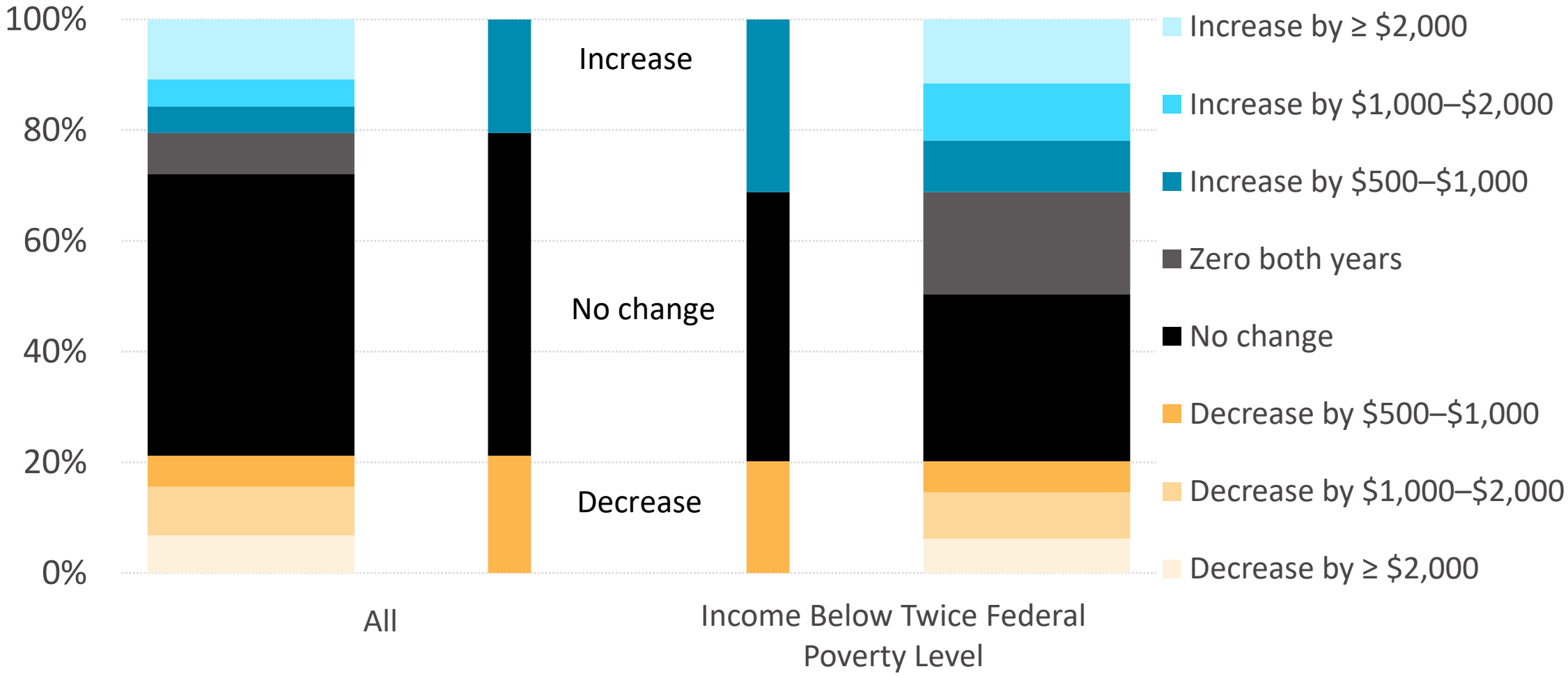
EITC Changes by Household Head Characteristics



Source: Urban Institute TRIM3 model using data from Current Population Survey Outgoing Rotation Groups 2015–18.
Note: Sample includes households with one dependent child under age 18 in either year with incomes below twice federal poverty level in the first year observed. No change is defined as a change of less than $500. Marital status only shown for those with same marital status in both years. Families with marital status changes excluded due to small sample size.

Results: Child Tax Credit

Magnitude of CTC Changes by Income

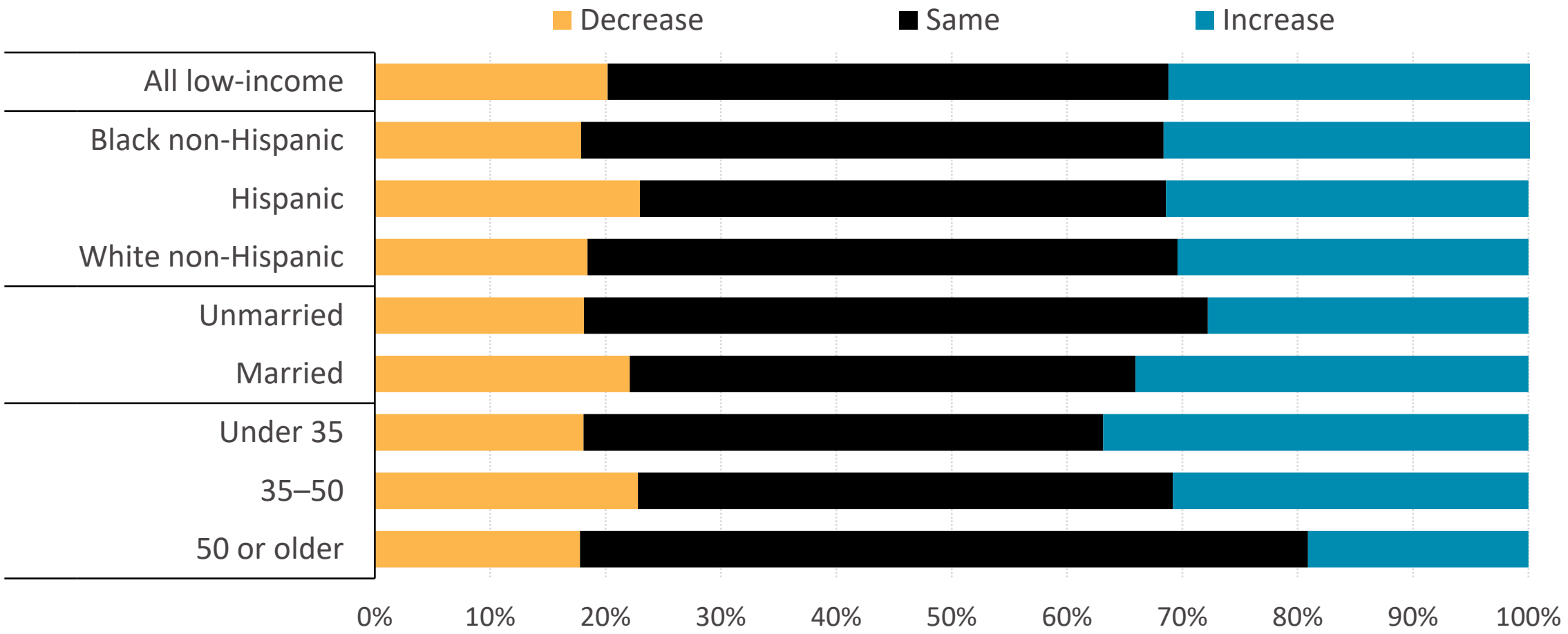


Source: Urban Institute TRIM3 model using data from Current Population Survey Outgoing Rotation Groups 2015–18.
Note: Sample includes households with one dependent child under age 18 in either year. No change is defined as a change of less than $500.

Income bigger driver of CTC changes than kids for low-income families

- Decreases (20%)
  - Income decreased (10%)
  - Number of children decreased (8%)
- Stays the same (49%)
- **Increases (31%)**
  - **Income increased (22%)**
  - Number of children increased (6%)

CTC Changes by Household Head Characteristics



Source: Urban Institute TRIM3 model using data from Current Population Survey Outgoing Rotation Groups 2015–18.
Note: Sample includes households with one dependent child under age 18 in either year with incomes below twice federal poverty level in the first year observed. No change is defined as a change of less than $500. Marital status only shown for those with same marital status in both years. Families with marital status changes excluded due to small sample size.

Implications for policymakers

- **Most but not all** families could accurately predict credits within $500 → A $500 safe harbor level can protect against income and family changes
- 39% of low-income families have EITC decreases and 20% have CTC decreases greater than $500 → Further safeguards like partial advances or higher thresholds might be needed
  - Outreach efforts can help families plan for foreseeable eligibility changes around kids entering/exiting tax unit

This document and accompanying presentation is intended to summarize and extend research in the paper by Elzayn et al. (2023). It does not represent IRS or Treasury Department policy.

Differences in Audit Rates by Race

Tom Hertz (RAAS), Brian Sartain (RICS), Kara Leibel (RAAS), Mark Payne (RAAS)

Presented to 13th Annual IRS/TPC Joint Research Conference on Tax Administration

June 22, 2023

Please note:

This document reflects the views of the authors, one of whom (Hertz) is also an author of the paper by Elzayn *et al.* (2023). This work is preliminary and pre-decisional and is being shared in the interest of eliciting constructive feedback to improve our understanding of the issues. The perspectives and findings expressed herein should not be taken to represent IRS or Treasury Department policy.

The IRS does not collect data on taxpayer race. Instead, race was imputed using Bayesian Improved First Name Surname Geocoding (BIFSG), which assigns each taxpayer a probability of belonging to each race/ethnicity category by matching names and addresses to published race/ethnicity distributions. These estimated race data are used for research purposes only; the IRS does not and will not consider race as part of its case selection and audit processes.

This document and accompanying presentation is intended to summarize and extend research in the paper by Elzayn et al. (2023). It does not represent IRS or Treasury Department policy.

Outline

1. Key findings reported in Elzayn et al. (2023, Stanford U., SIEPR Working Paper)
2. Appraisal of key findings from Elzayn *et al*. (the "Stanford paper")
3. Potential sources of audit rate differences by race
4. Enforcement objectives matter
5. Evidence for algorithmic bias in EITC audit selection
6. Evidence that unscrupulous paid preparers contribute to audit rate gap
7. Conclusions and caveats

This document and accompanying presentation is intended to summarize and extend research in the paper by Elzayn et al. (2023). It does not represent IRS or Treasury Department policy.

Key findings from Elzayn *et al.*

1. Using imputed race data, Elzayn *et al.* find that Black taxpayers were audited at between 2.9 and 4.7 times the rate of non-Black taxpayers in TY2014.
2. They find that the bulk of this gap reflects differences in audit rates by race among EITC claimants.
3. Looking at EITC claimants, they built alternative audit-selection models using representative audit data from the National Research Program (NRP) to try to infer what might be creating a race gap in the outcomes of the operational audit-selection models. They address the following questions:

   a) Which goal? Models that tried to find claimants with the highest *total tax understatements* picked non-Black taxpayers at higher rates; models that tried to find claimants with the *highest overclaimed refundable credits* picked Black taxpayers at higher rates.

   b) Which model? Selecting taxpayers with the highest *expected value* of tax understatements picked non-Black taxpayers at higher rates; selecting those with the highest *probability* of any understatement picked Black taxpayers at higher rates.

   c) Large Schedule Cs? Models that were constrained to audit limited numbers of EITC-claiming returns with large Schedule C enterprises audited Black taxpayers at higher rates than did unconstrained models.

This document and accompanying presentation is intended to summarize and extend research in the paper by Elzayn et al. (2023). It does not represent IRS or Treasury Department policy.

Appraisal of key findings from Elzayn *et al.*

1. **Subsequent research replicates the headline finding and documents that it is relatively stable over time**
   - This work also extends the analysis to cover Hispanic, Asian & Pacific Islander, White, and All Other/Multiple Race taxpayers, who were grouped together as "non-Black" in Elzayn *et al.*

2. **How much of the audit rate gap is due to…**
   (a) differences in audit rates by race among EITC claimants
   (b) differences in audit rates by race among non-EITC returns
   (c) differences in overall EITC versus non-EITC audit rates

   - Elzayn *et al.*: (a) 78% (b) 8% (c) 14%, which would seem to imply that to reduce the overall race gap we should focus on reducing the race gap in audit rates among EITC returns.

     ✓Other standard decomposition methods assign a larger share of the total to (c), thus placing more emphasis on the EITC/non-EITC audit rate differential.

     ✓Holding all else equal, equating the overall EITC and non-EITC audit rates would have about the same effect on the overall race gap as would equating Black and non-Black EITC audit rates: both would be expected to reduce the gap by about 60%.

This document and accompanying presentation is intended to summarize and extend research in the paper by Elzayn et al. (2023). It does not represent IRS or Treasury Department policy.

3. How informative are the Stanford modeling exercises?

Among EITC returns, the choice of goal, or audit objective, does indeed matter.

Among EITC claimants, the demographics of the distribution of *total tax understatements* are different from the demographics of the distribution of *overclaimed refundable credits*. As a result, models pursuing these two different objectives will produce different audit rate gaps by race.

- Historically, the Refundable Credit audit program has focused on incorrect claims of refundable credits not on total tax understatements.
- The Stanford paper seems to suggest that a change in objective would reduce the race gap and raise total revenue. However, it is important to be clear that the Stanford revenue estimates only hold if all audits are NRP-style audits (full scope, average duration 18 hours), whereas the vast majority of EITC audits have traditionally been conducted as correspondence audits (limited scope, average duration 1.5 hours).

This document and accompanying presentation is intended to summarize and extend research in the paper by Elzayn et al. (2023). It does not represent IRS or Treasury Department policy.

Expected values versus probabilities: Does not appear to be the problem

- Operational EITC audit selection models generally have the primary objective of minimizing the probability of selecting compliant taxpayers, with secondary consideration given to revenue
  - *Note*: Avoiding selecting compliant taxpayers is particularly important for pre-refund audits of low-income taxpayers, where the refund is frozen until the audit is completed.
- Subsequent analysis by the Stanford team has confirmed that when the objective is refundable credit overclaims, models of the probability of noncompliance select *fewer* Black taxpayers than do models of the expected value of noncompliance. (The opposite is true if the objective is total tax understatements.)
- This suggests that the use of probability-based models is not driving up the audit rate gap.

Under-representation of larger Schedule Cs

- Stanford found that EITC returns with larger Schedule C businesses are audited at lower rates than their models suggest is economically optimal, and that selecting more such returns would reduce the Black/non-Black audit rate gap.

- However, this conclusion applies to models that predict *total tax understatements* and *audit the full return*.

This document and accompanying presentation is intended to summarize and extend research in the paper by Elzayn et al. (2023). It does not represent IRS or Treasury Department policy.

Potential sources of audit rate differences by race

1. ***Eligibility criteria*: *Example:*** Married couples can claim stepchildren, but unmarried taxpayers cannot claim their partner's children, even if co-resident, and run the risk of being audited if they do. This could potentially have a disparate impact on Black taxpayers, who have lower marriage rates.

2. ***Unscrupulous preparers*:** Preparers who submit lots of false claims for EITC (and are spotted by IRS's Preparer Strategy program) draw clients disproportionately from minority communities.

3. ***Exam objectives:*** Minimize no-changes? Find largest credit overclaims? Find largest *total* tax changes? More single-issue audits or fewer, longer, multi-issue audits? These policy choices have demographic implications.

4. ***Actual algorithmic bias:*** Algorithmic bias occurs when a model for a particular audit selection workstream generates demographic differences in audit rates that cannot be explained by underlying differences in noncompliance (as defined in relation to the existing tax code, and in pursuit of chosen enforcement objectives).

This document and accompanying presentation is intended to summarize and extend research in the paper by Elzayn et al. (2023). It does not represent IRS or Treasury Department policy.

Enforcement objectives matter

- There are many plausible ways to measure and identify noncompliance with the tax code, which serve different enforcement objectives. Each will result in a different mix of returns selected for audit that may also be distributed differently among subgroups in the population. Enforcement objectives may include:

  ✓ Maximizing net enforcement revenue for a given enforcement budget: Audit according to marginal revenue/cost, subject to current staffing levels and skills constraints.

  ✓ "Proportionality:" The principle that audit risk should rise sharply with value of total tax understatement, even if that does not maximize enforcement revenue. (Note: not the same as vertical equity.)

  ✓ Minimize audits of compliant taxpayers (but sacrifice some revenue).

  ✓ Minimize improper payments of refundable credits: Leads to higher audit rates for Black taxpayers.

  ✓ Maintain minimum coverage across all types of returns and all types of noncompliance: Recognize variety of audit workstreams – there is no single model.

  ✓ Maximize total Federal revenue: Emphasize deterrent properties of enforcement.

- For any given set of objectives, it is possible to estimate the expected demographics of taxpayers who meet the corresponding audit criteria. Deviations from those targets are then diagnostic of algorithmic bias.

This document and accompanying presentation is intended to summarize and extend research in the paper by Elzayn et al. (2023). It does not represent IRS or Treasury Department policy.

Evidence for algorithmic bias in EITC audit selection

- Research suggests that the current Black/non-Black EITC audit rate ratio is higher than the Black/non-Black noncompliance rate ratio, no matter how noncompliance is defined.
- This suggests that algorithmic biases do contribute to the EITC audit rate gap.
- To date three mechanisms that contribute to algorithmic bias have been found:
  - ✓ The residency and relationship status of dependents must be imputed from incomplete information, and this process is not error-free. Imputation errors appear to raise the audit risk for Black EITC claimants relative to others. Modernizing models and supplementing existing data sources may be able to mitigate this problem.
  - ✓ Aging models: Updating the existing EITC scoring model could potentially increase exam revenue *and* reduce racial bias.
  - ✓ In the past, weekly audit selection targets led to over-selection in some weeks; this drove down audit quality and appears to have created racial bias. This problem has largely been resolved.

This document and accompanying presentation is intended to summarize and extend research in the paper by Elzayn et al. (2023). It does not represent IRS or Treasury Department policy.

Evidence that unscrupulous paid preparers contribute to audit rate gap

- Since 2005 the IRS has monitored paid return preparers who submit large numbers of high-risk returns claiming refundable credits on behalf of their clients. Preparers are subject to civil and criminal penalties, which do have some effect on subsequent behaviors. Treasury has proposed expanded and increased penalties for unscrupulous preparers.

- In TY2019, 17 million returns (of which 5.8 million claimed EITC) were submitted by preparers known to this program on behalf of clients drawn disproportionately from minority communities.

- Calculating audit rates after *excluding* all returns from identified unscrupulous preparers, the overall Black/non-Black gap in audit rates falls by 21% (for TY2019).

This document and accompanying presentation is intended to summarize and extend research in the paper by Elzayn et al. (2023). It does not represent IRS or Treasury Department policy.

Conclusions and caveats

- The emphasis on preventing overclaims of EITC credits is reflected in both the *number* of EITC audits conducted and in the way EITC claimants are selected for audit. In both cases, this emphasis serves to raise audit rates for Black taxpayers relative to others.
  - ✓Ongoing research is evaluating the hypothesis that a change in audit objectives, to focus on top-dollar tax understatements among claimants of refundable credits, is feasible in a pre-refund correspondence audit environment.
  - ✓This includes estimating outcomes in terms of differences in audit rates by race, burden on compliant taxpayers, and enforcement revenue.
- There is evidence of algorithmic bias. Preliminary research has identified potential updates to algorithms that may be able to lower the Black/non-Black audit rate gap while improving audit outcomes.
- Improvements to audit selection algorithms will take time to test and implement and are critically dependent on the funding made available through the IRA.

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

Comments on Papers by Millard et al and Hertz et al

IRS-TPC Conference on Research in Tax Administration

June 22, 2023

Janet Holtzblatt, Tax Policy Center

Millard et al
Our Common Experience

- Very relatable paper!
  - What makes you hang up and switch to business or government's website?
    - *How long the wait?*
    - *The messaging?*
    - *The number of messages?*
    - *The music?*
    - *Your tolerance for pain?*

Millard et alt
Sample and behavior

1. Taxpayer receives a collection notice from the IRS
2. Taxpayer calls the IRS with questions and get into queue
3. Taxpayer hears prerecorded message followed by music (5 different messages, followed by music)
   - Does caller hang up? (‘’abandon” rate)
   - Does caller go to web site? (“channel shift” rate)
     - *And when does caller acr? # calls, # days after call*
   - Does caller use web site application? (“access” rate)

Measuring impact

- To measure impact, authors look at the percent difference in those three measures between control group and test group

- But does focus on percent change overstate impact? For example:
  - 13 percent increase in channel shifting
  - But that's associated with just a 1.6 percentage point difference in channel shifting—12.5% of control compared to 14.1% test group

Millard et alt
Changing the message

- Begin each message with a question—and then suggest action
- Refer caller to an IRS website in each message
  - Repeat address at least twice in each message
  - Give link to the website that matches the suggested action in message
- Tell caller that setting up payment plan online is cheaper than on phone
- Order messages by most frequently asked
- Don’t tell caller that call volume is high

What is the impact of a change in messaging?

- Focus is effectively on how often taxpayer calls and hears messages before embarking on next step

- What do we learn from that focus?

- Should more focus been placed on what changes in messaging, on margin, work best?

What message had biggest impact on next step?

- Among those who hung up: **Message 2** (about 4% points)
  - Told cheaper to set up payment plan online using **irs,gov/opa**

- Among those who hung up and then channel shift: **Message 1** (6 – 14 % points)
  - Gave caller a link to **irs.gov/payments**

- Among those who channel shift: Largest difference was increase in share who set up payment plan online at **irs.gov/opa** (3.5% points)

From percentage point to numbers

- Size of sample
  - Control: 42,070
  - Test: 43,032
- The number of abandoners who set up OPA is small by any measure
  - Control: 684
  - Test: 1,069
- Is there a problem with the OPA web site?

Final thoughts on Millard et al

- Informative & fun, because who hasn't suffered thru customer service calls?
- Observation that IRS performance measure of "level of service" may understate experience by not following callers who hang up and go online
- Conclusion that one performance measure is not enough
- Potential power of being able to use TIN of caller to track results:
  - *Is it possible to track TINs to measure ROI for taxpayer services?*

Hertz et al
Race and the IRS

- The Stanford study confirms fears of racial disparities in audit rates without being able to explain why
  - Couldn’t reveal selection criteria to Stanford researchers
  - Objective matters
- Other possible reasons:
  - Disparities arise when factors affecting tax liability associated with race
  - Cracks in third-party information
  - “Actual algorithmic bias”
- Will Hertz et al (all IRS employees) be able to answer why?

Hertz et al
What is the objective? Enforcement revenues in 1990s?

- **Congress**
  - For some: Roll back EITC expansion, even repeal
- **Administration**
  - Keep EITC from being cut
  - Obtain some refundability for child tax credit
- **The deal**
  - No cuts to EITC, partial refundability of CTC
  - Bump up IRS appropriations for EITC administration, with commitment to **increase enforcement revenues**
  - Statutory changes with **associated revenue estimates**

Hertz et al
What is the objective? Improper Payments Act of 2002

- While understanding that the goal was to improve EITC compliance, some reluctance to focus on targeted **error rate**
- Did Improper Payments Act of 2002 change that view?
- Act required agencies to
  - *With OMB guidance*, to identify programs with significant improper payments
  - **Provide estimates of improper payments--dollars**
  - Report on actions to reduce improper payment
- Treasury does not provide target for future—but does annual list of improper payments focus attention

Hertz et al
Treasury and IRS perspectives on achieving objectives in 1990s

- Keep IRS cost per audit low so as to expand audit coverage
- Minimize audit burden for taxpayers through correspondence audits
- Supplement in 1996 and 1997 legislation:
  - Expanded math error authority when feasible (SSN, primarily)
  - Simplification
  - Due diligence requirements for paid preparers
  - Third-party information
- Informed by findings of compliance studies

Hertz et al
Cracks in the third party-info

- SSN: Require parents to supply own SSN when applying for child's SSN
  - Always viewed as long-term strategy
  - Crack opened after enactment when SSA said it could not guarantee SSN for more than one parent

- National registry of child support orders
  - Always viewed as having crack because only contained info when government involved in enforcement

Hertz et al
Still, hope was that data analysis would fill in hole

- Results suggest that data analytics incomplete

- Cracks had unanticipated consequences

- One troublesome question:
  - Marriage rates among Black adults are about half that of white couples
  - Compliance data pointed to problems with HOH filing status
  - But without marriage data, did the wrong HOHs get targeted?

Hertz et al
Still to be explored

- What happens after return selected for audit?
  - Why is nonresponse rate high?
  - If selection criteria is based, in part, on historical data of audit results, to what extent are post-selection biases perpetuated? "*Actual algorithmic bias*"?

- Would machine learning lead to improvements?

The Balance Due Taxpayer: How Do We Reduce IRS Cost and Taxpayer Burden for Resolving Balance Due Accounts?

Understanding Yearly Changes in Family Structure and Income and Their Impact on Tax Credits

Discussant:

Emily Lin

U.S. Department of the Treasury

Balance Due Accounts

- Costly to the IRS; burdensome to taxpayers; emphasis on preventive measures
- More situations in general
  - income withholding, departure of a child, etc.
- Restrict the analysis to CP14 issuance
  - Majority of those with an unfavorable shift in balance due status did not involve a CP14; Effect of income
- Behavioral insights to create additional opportunities for communication, education during the tax year
  - Not only when a life event occurs
- Prevent costly downstream activities

Yearly Changes in Credits

- Large year-to-year swings in EITC and CTC; effects vary with demographics; important implications for the design of advance credits
- Child may be claimed by another taxpayer
  - Child well-being vs. credit a taxpayer is entitled to
- Marital status change: income and child residency changes
- Long-term trend in living situations: Connect to the paper's focus on yearly shifts; implications
- What would happen if advance credits were based solely on prior-year income and family structure?
  - Conditions: changes are predictable, reportable; credit designs matter
- Minimize the risk of unexpected and unfavorable yearly changes in the credit amount: communications and education