The Brookings Institution

Metropolitan Policy Program
Alan Berube, Fellow

The Potential Value of IRS Data for State/Local Analysis

KIDS COUNT regional workshop
San Diego, CA
February 2, 2006

Overview

- Why might data from the IRS help organizations track the well-being of families and children?
- What do IRS data look like, and how are they made available?
- What are some examples of how organizations could make use of IRS income data?

Data availability might be one bright spot at the IRS….

But in this world nothing can be said to be certain, except death and taxes.

Benjamin Franklin

The hardest thing in the world to understand is the income tax.

Albert Einstein

Taxation **with** *representation ain't so hot, either.*

Gerald Barzan

What economic measures do KIDS COUNT organizations track?

- Median income (for households and families)
- Income distribution
- Child poverty; “working” poverty (< 200% poverty)
- Participation in income support programs (e.g., TANF, FSP)
- Unemployment
- Labor force participation (families with children)

Most programs track one or more of these indicators at the state and county levels on an annual/biennial basis

The bottom line: where and how IRS data might be useful

1. Examining topics at sub-county level (cities & towns, large neighborhoods)
2. Providing a view of “working poverty” (receipt of the EITC)
3. Tracking additional income support programs (EITC, Additional CTC in future years)
4. Analyzing income changes over time
5. Looking at structure of low-income working families

Where do state/local analysts get income information currently?

Source	Latest available	Description
American Community Survey	2004	Comprehensive ongoing survey of 800,000 households, increasing to 3 million in 2005
Decennial Census	2000	Comprehensive survey of one in six U.S. households
Current Population Survey, Annual Social and Economic Supplement	2004	March addendum to monthly labor force survey of 100,000 households
Census Bureau Small Income and Poverty Estimates	2003	Model-based income and poverty estimates for states and counties based on survey and administrative data

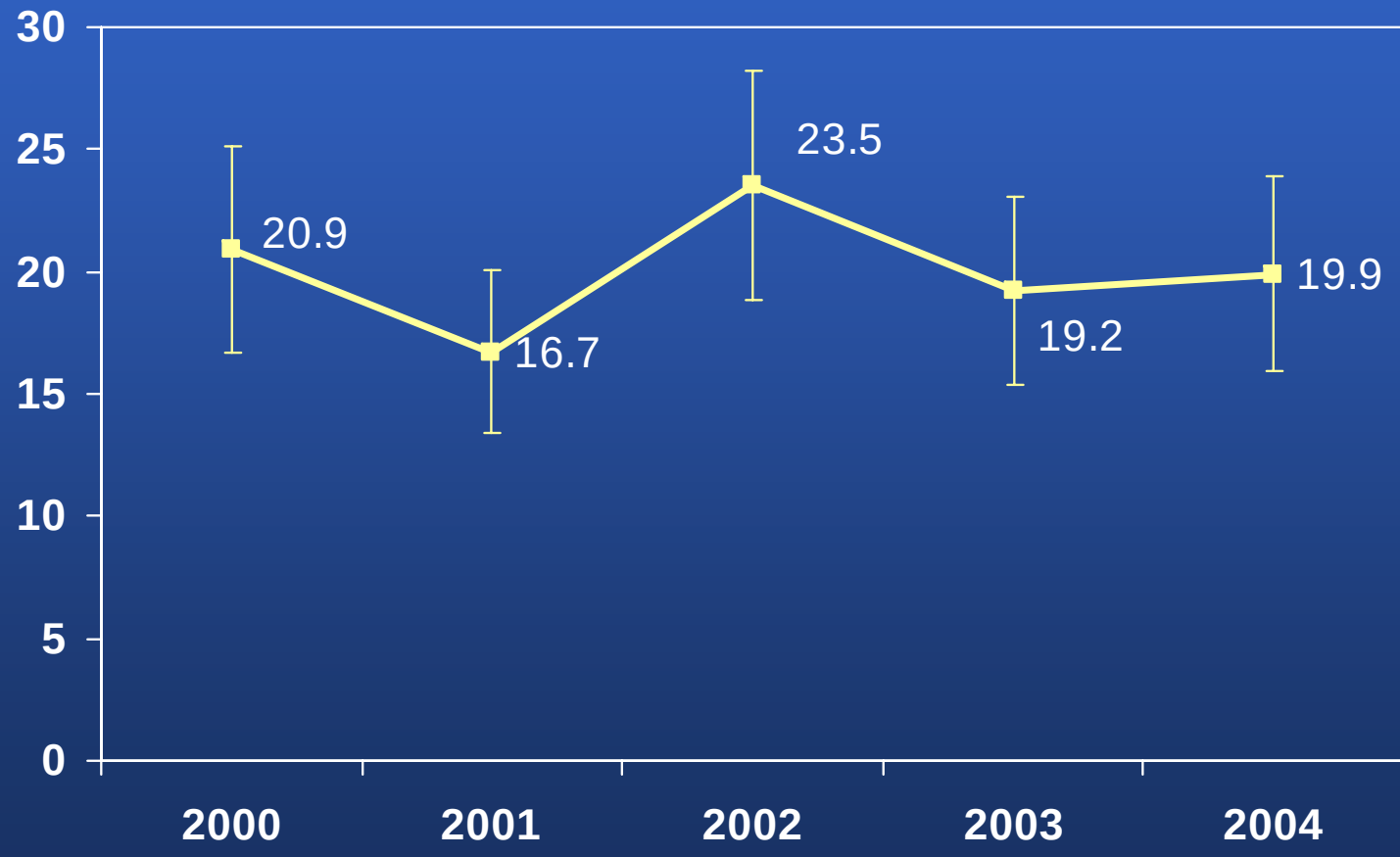
Why might IRS data be helpful?

- 2000 was a *long* time ago
  - Using post-census information is critical for staying relevant
  - The IRS collects data annually nationwide
- Taxes are increasingly important for low-income families
  - Far more families with children benefit from the EITC than from TANF or Food Stamps
- No sub-county information is available from these sources
  - IRS data provide subcounty detail
- Estimates based on surveys can get a little wacky
  - Sample sizes can be quite low, or change over time
  - IRS data are based on a 100% sample (of taxpayers)

For example, here's the trend in child poverty in San Diego according to ACS, with approximate 90% confidence interval

Child poverty rate, San Diego, 2000-2004

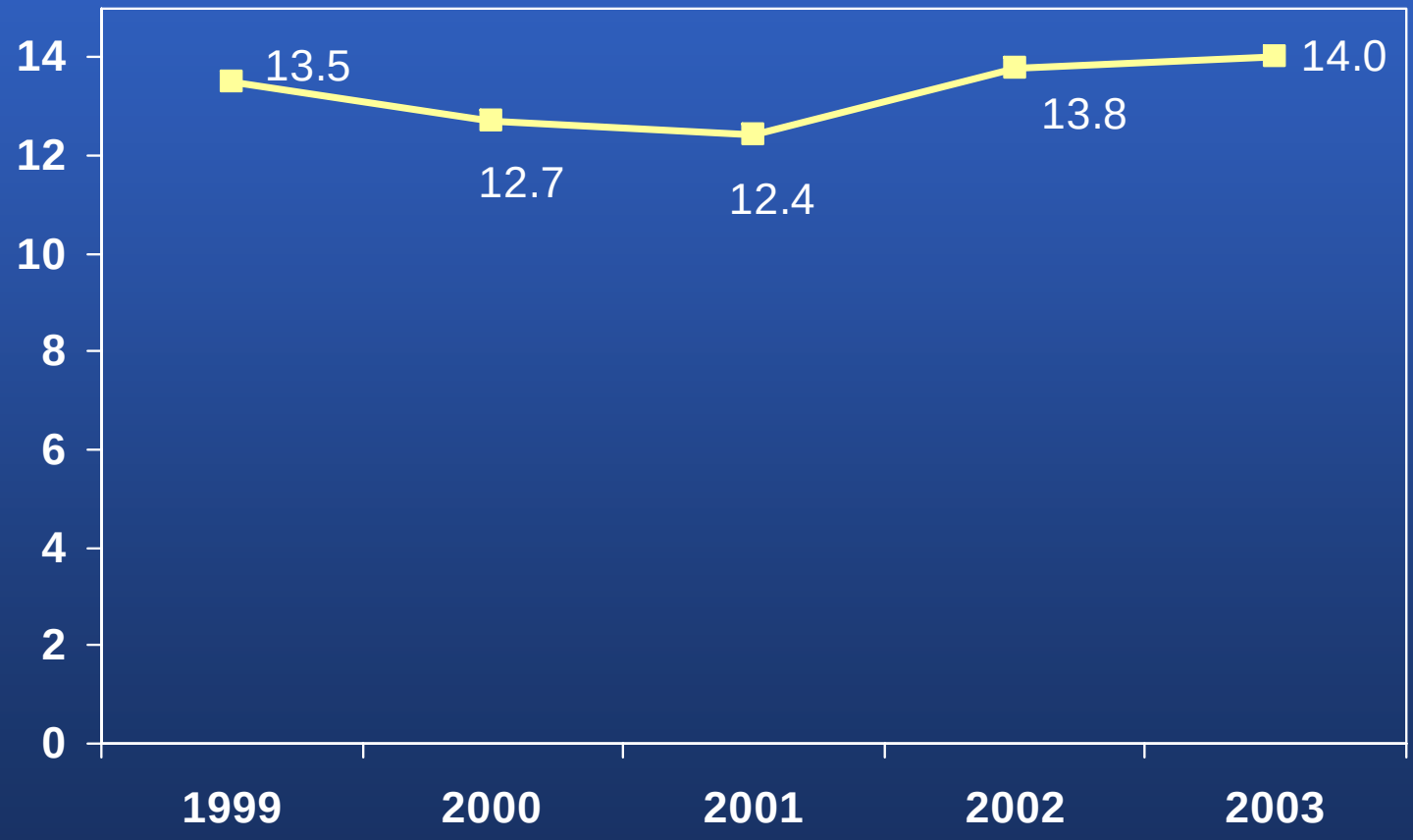
Source: American Community Survey



And here's the (much more stable, no error) trend in EITC receipt during roughly the same period

Percentage of taxpayers claiming EITC, San Diego, 1999-2003

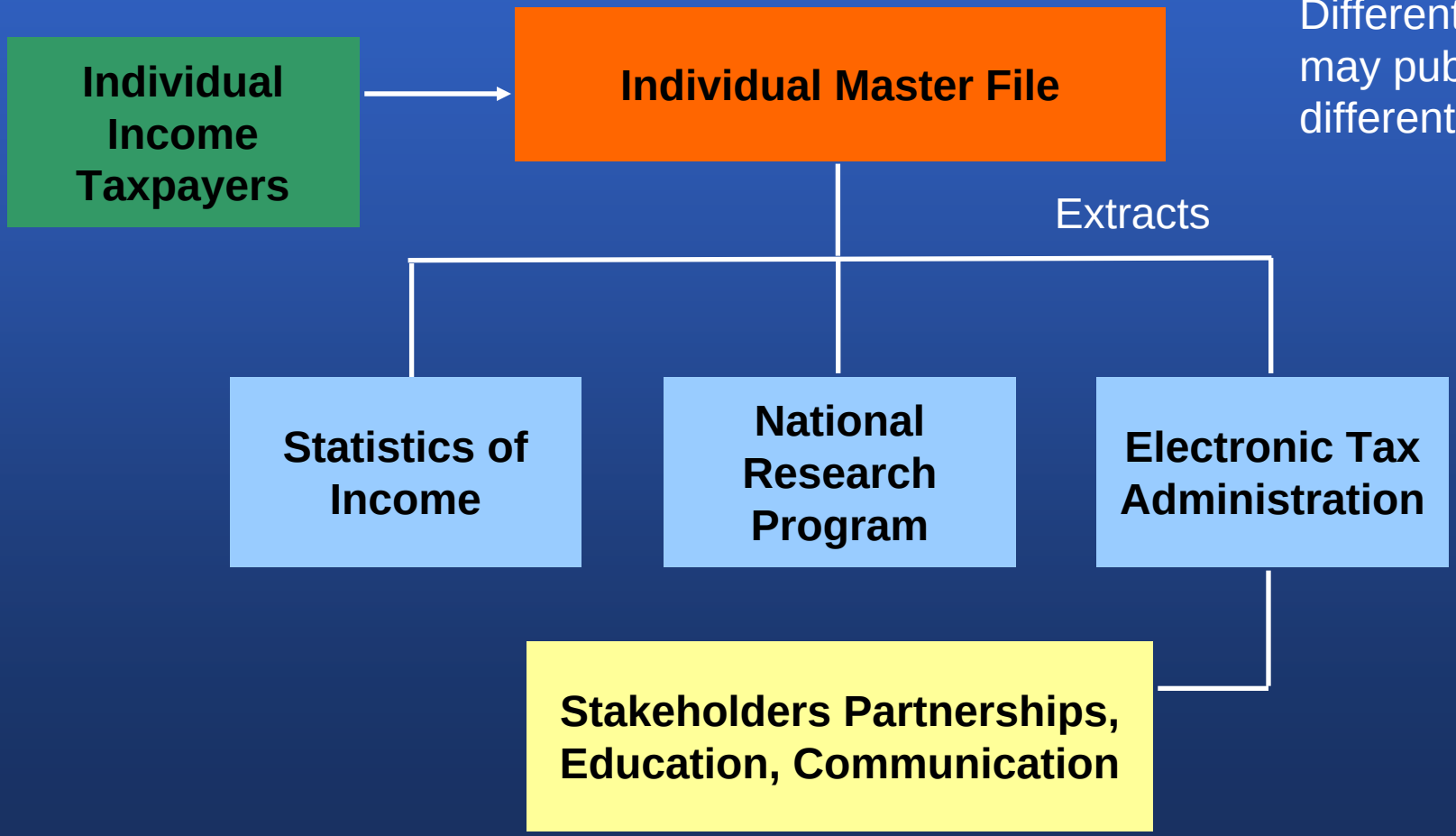
Source: Brookings analysis of Internal Revenue Service data



There are several things to understand about IRS data before diving in, however

- Who publishes these IRS data?
- Where (and how often) does one get them?
- How are they organized?
- What do they contain?
- Why do IRS income measures differ from others?

The WHO: several IRS offices publish tax data; those relevant for state/local work are published by the SPEC division



Different IRS offices may publish slightly different numbers

The WHERE: the data are made publicly available by the IRS (and by Brookings, in derivative format)

- Data are published annually, with approximately 18-month lag (e.g., tax year 2003 data—returns filed in 2004—arrive in fall 2005)
- IRS-SPEC publishes an Access database with a reporting interface (that doesn't always work very well)
- Every region of the U.S. has a SPEC Territory Manager that can provide requesters with data
- Meanwhile, Brookings will make many of these data available on its website in the next 3-5 months (some are already there)

The HOW: data are organized by taxpayer ZIP code, and higher-level geographies

IRS aggregates these data for cities, counties, and states

Each release contains most recent tax year, plus 2-3 prior years

Microsoft Access - [aggrZip03_R : Table]

File Edit View Insert Format Records Tools Window Help

Type a question for help

ZipCode	State	State_tx	TERR_CD	ttTOTRET	ltTOTRET	etTOTRET	ttNEW	ltNEW	etNEW
01532	MA	MASSACHUSE	11	6481	2499	272	378	323	20
01534	MA	MASSACHUSE	11	2376	1100	189	140	126	16
01535	MA	MASSACHUSE	11	2118	1078	213	138	123	19
01536	MA	MASSACHUSE	11	2959	1155	156	128	104	12
01537	MA	MASSACHUSE	11	1176	548	85	61	48	11
01538	MA	MASSACHUSE	11	351	215	44	17	16	0
01540	MA	MASSACHUSE	11	5222	2468	477	302	271	35
01541	MA	MASSACHUSE	11	1717	666	62	99	90	0
01542	MA	MASSACHUSE	11	1036	464	77	70	57	0
01543	MA	MASSACHUSE	11	3211	1283	179	182	163	0
01545	MA	MASSACHUSE	11	15339	5991	693	867	758	75
01546	MA	MASSACHUSE	11	0	0	0	0	0	0
01550	MA	MASSACHUSE	11	7816	4956	1544	572	533	149
01560	MA	MASSACHUSE	11	1519	719	156	88	80	15
01561	MA	MASSACHUSE	11	685	408	110	61	54	18
01562	MA	MASSACHUSE	11	5505	2829	536	323	293	40
01564	MA	MASSACHUSE	11	3622	1434	176	172	163	17
01566	MA	MASSACHUSE	11	3163	1370	181	131	112	0
01568	MA	MASSACHUSE	11	3018	1118	149	146	129	16
01569	MA	MASSACHUSE	11	5435	2243	349	281	250	28
01570	MA	MASSACHUSE	11	7910	4330	977	477	433	93
01571	MA	MASSACHUSE	11	4756	2252	404	268	242	32
01580	MA	MASSACHUSE	11	0	0	0	0	0	0
01581	MA	MASSACHUSE	11	8344	3348	463	628	546	73
01582	MA	MASSACHUSE	11	0	0	0	0	0	0
01583	MA	MASSACHUSE	11	3151	1429	177	174	150	17
01585	MA	MASSACHUSE	11	1908	914	179	117	105	11
01586	MA	MASSACHUSE	11	0	0	0	0	0	0
01588	MA	MASSACHUSE	11	4054	1932	372	226	200	21
01590	MA	MASSACHUSE	11	3825	1493	149	241	214	12
01601	MA	MASSACHUSE	11	238	185	76	20	18	0
01602	MA	MASSACHUSE	11	10424	5216	964	620	538	85
01603	MA	MASSACHUSE	11	8182	5230	1637	732	682	201
01604	MA	MASSACHUSE	11	15335	8996	2262	1235	1125	272

Record: 430 of 42290

Datasheet View NUM

The WHAT: the IRS database contains 75 data elements in each of several "market segments"

- Market segments aggregate filers that share a common characteristic relevant for research or IRS operations
  - e.g., all tax returns; returns receiving EITC; elderly returns, volunteer-prepared returns
- Several data elements could be valuable for socioeconomic analysis
  - receipt of certain tax benefits (EITC, Child/Dependent Care Credit)
  - detailed income (AGI) categories
  - filing status
- Others, not so much
  - type of return (1040, 1040EZ), type of preparer (paid, self), schedules (deductions, self-employment, capital gains)

The WHY: Important factors distinguish IRS income data from ACS/CPS/SAIPE income data (1)

TAXPAYERS

- Can be families (with or without children) or single individuals
- Don't represent households exactly because:
  - Not everyone is required to file a tax return (singles under $8,200; married couples under $16,400; many people age 65 and over)
  - Some people fail to file a return, because they owe (or think they owe) taxes
  - Some married couples, unrelated people living together, and those in group quarters file separate returns
- There were about 8% more returns in 2004 than households, but varied across US (5% fewer in AR; 25% more in HI)

The WHY: Important factors distinguish IRS income data from ACS/CPS/SAIPE income data (2)

ADJUSTED GROSS INCOME

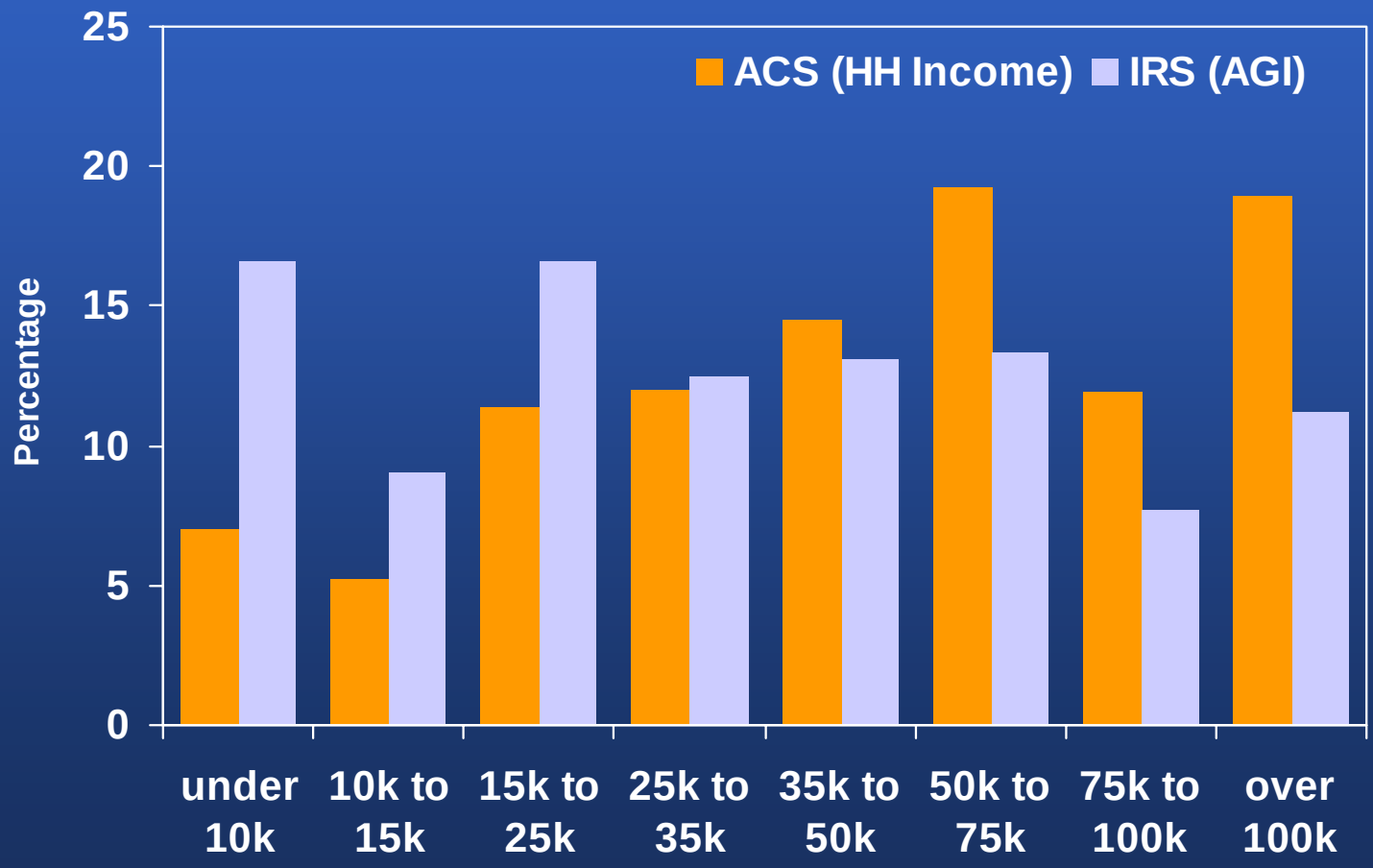
- Not the same as regular income reported on surveys—subtracts certain adjustments for tax purposes
  - Health savings/IRS deductions
  - Self-employment taxes
  - Alimony paid
  - Student loan interest
  - Tuition and fees
- Most of these are less relevant for lower-income taxpayers, but they alter the overall distribution of income
- Different reporting units also result in different distribution

About IRS data

Household incomes in San Diego County (per ACS) look much higher than taxpayer AGIs (per IRS)

Percentage of households/ taxpayers by income, San Diego County, 2003

Source: Brookings analysis of ACS and Internal Revenue Service data



The WHY: Important factors distinguish IRS income data from ACS/CPS/SAIPE income data (3)

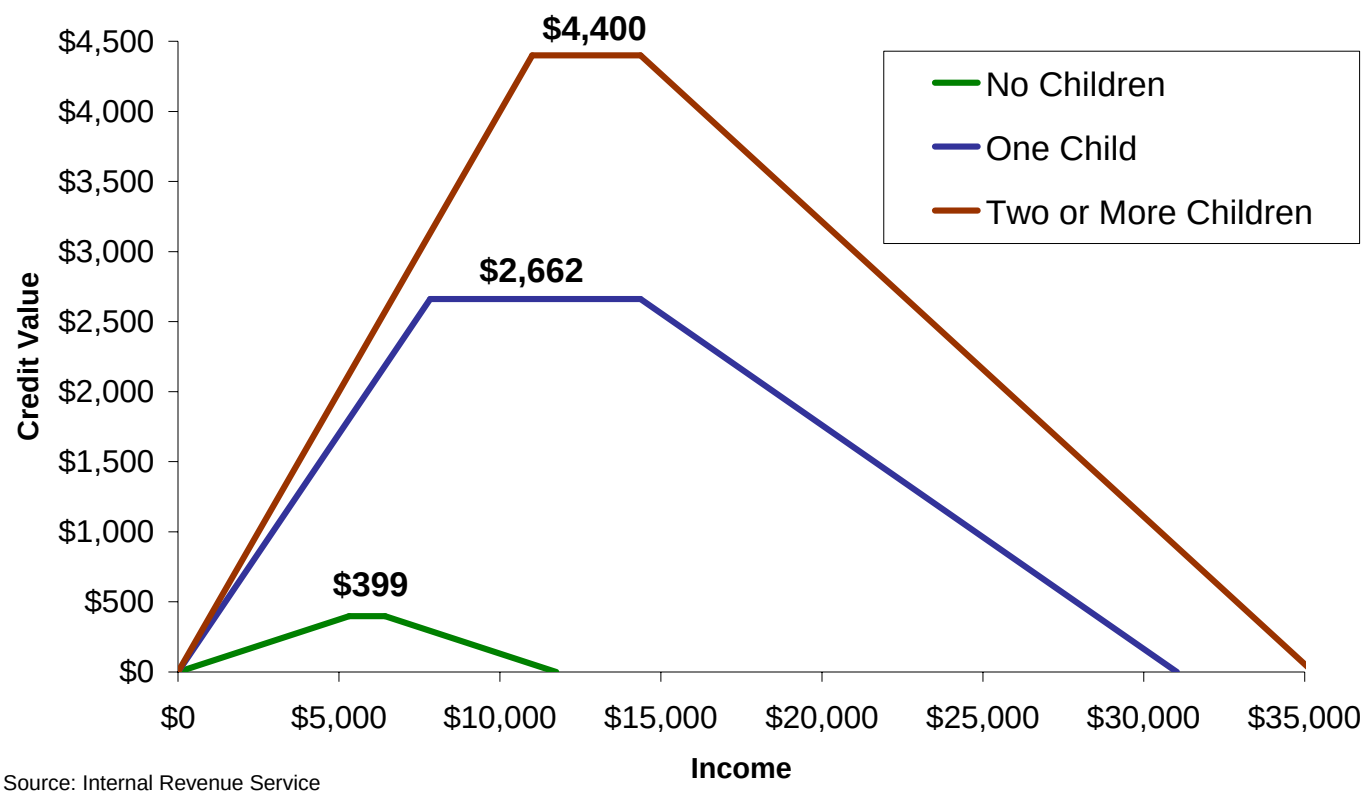
EARNED INCOME TAX CREDIT

- Can be claimed by families with children with incomes up to roughly 200% poverty; childless workers with incomes up to roughly 100% poverty (see chart)
- 80% of claimants have children in their home (receive 97% of the benefits)
- Dynamics of receipt reflect economic conditions primarily, participation secondarily

About IRS data

The EITC actually embodies three separate tax credits based on income and family type

Value of EITC by income and number of children, 2005



Source: Internal Revenue Service

* Married couples filing jointly are eligible for slightly higher credit amounts in the "phase-out" range of the EITC.

The WHY: Important factors distinguish IRS income data from ACS/CPS/SAIPE income data (4)

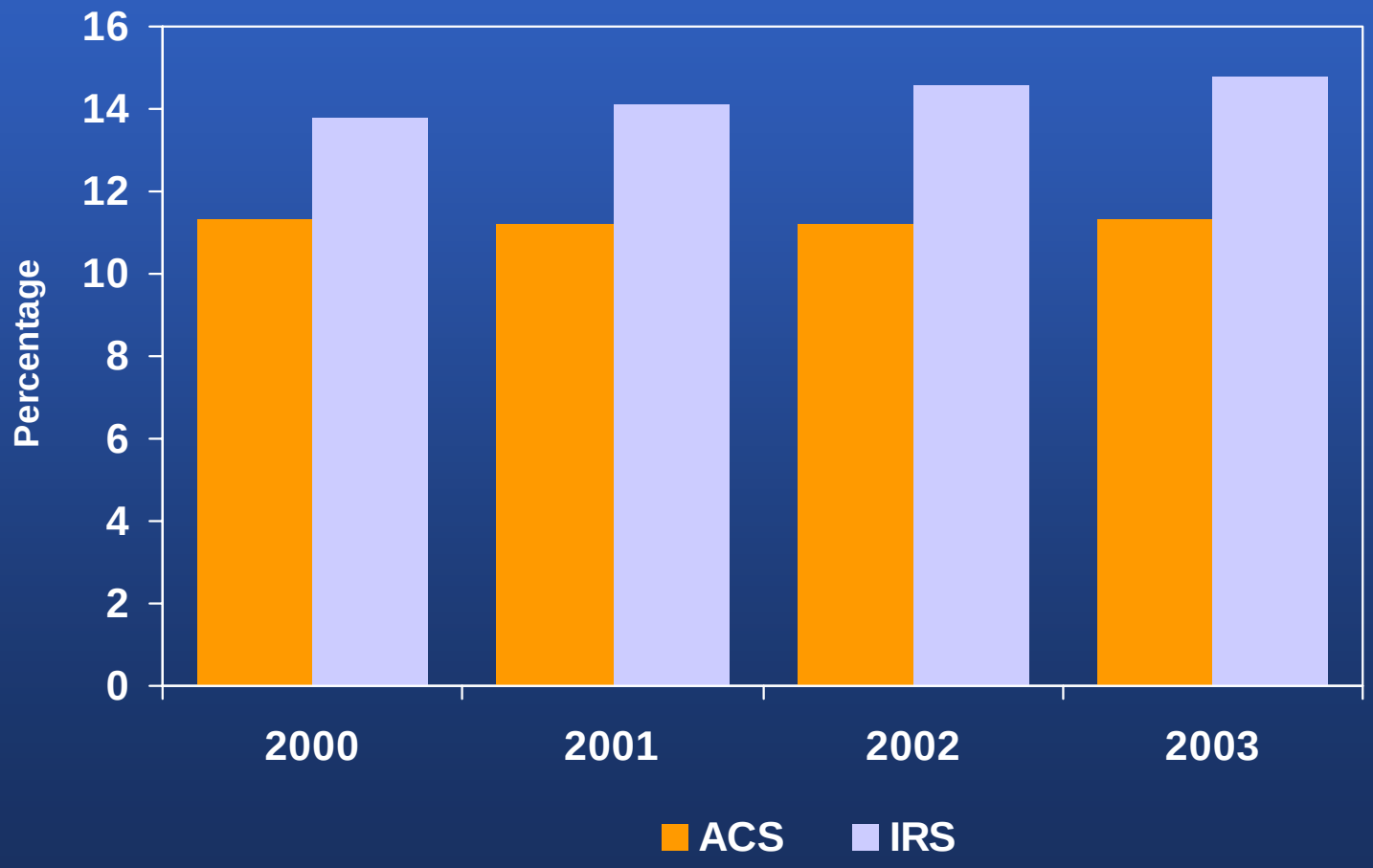
FILING STATUS

- Four possible statuses: single, married filing jointly, married filing separately, head of household
- Not all heads of household are single parents
  - May be other relative (aunt/uncle, grandparent)
  - May be caring for other dependent (parent, sibling)
- Not all single parents are heads of household
  - Some file as single (don't pass "support test")
  - Some are married, filing separately (didn't live apart from spouse for second half of year)

About IRS data

IRS shows slow growth in heads of household; ACS shows little change in families with related children

Percentage of taxpayers that are heads of HH vs. percentage of HH that are families with related kids, 2000-2003



Source: Brookings analysis of Internal Revenue Service data

Assuming these don't totally discourage you…

Four examples of how you might use IRS data in your work

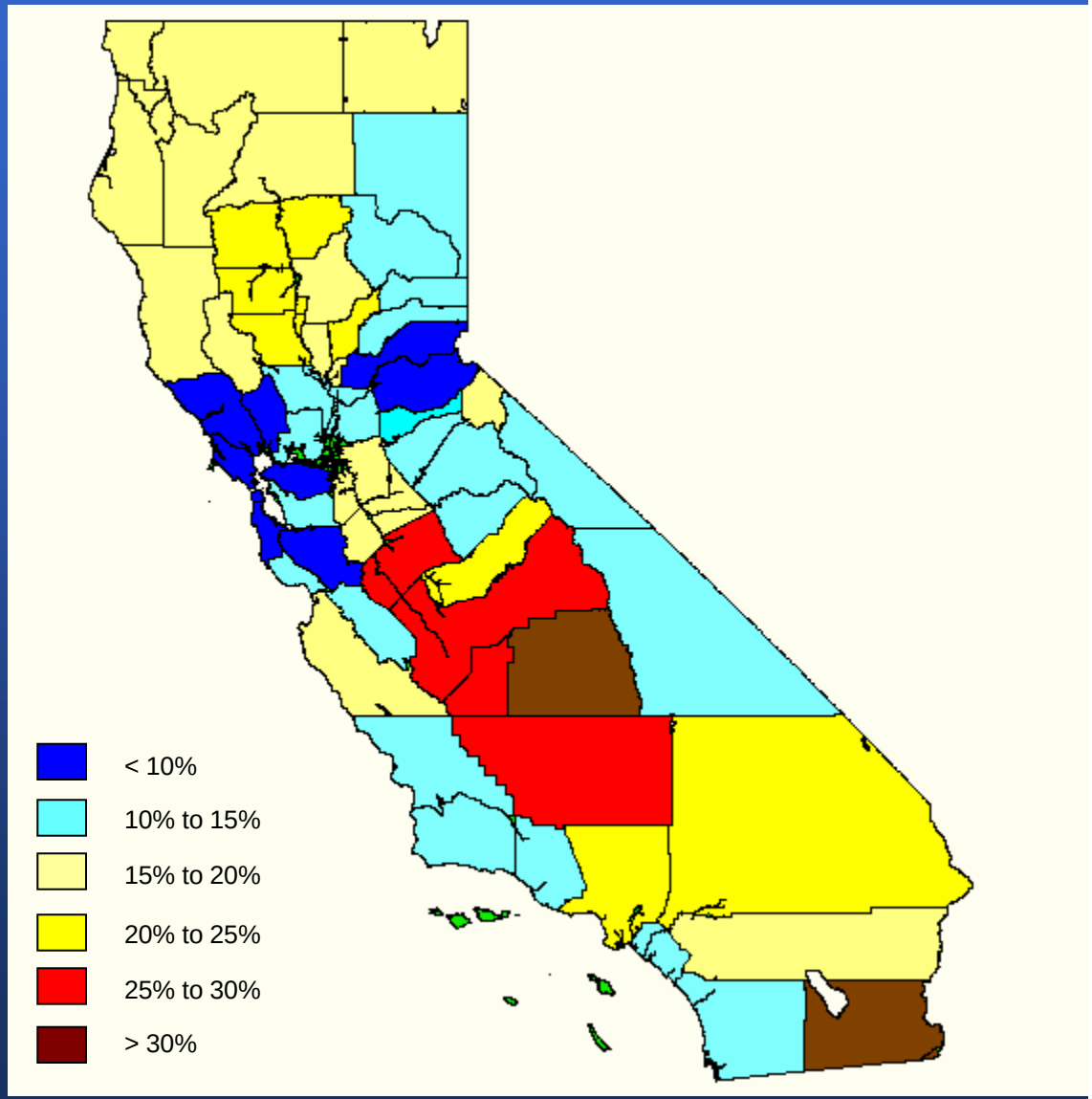
1. To describe levels and changes in “working poverty” (EITC)
2. To describe the characteristics of low-income working families
3. To show income changes over time—esp. for places
4. To demonstrate the potential value of similar state investments

Using IRS data

#1—Show regional variation in the incidence of "working poverty"

% of taxpayers receiving EITC by county, CA, 2003

Source: Brookings analysis of Internal Revenue Service data



Using IRS data

#1—Track local change in working poverty

Top ten cities by change in % of taxpayers receiving EITC, CA, 2003

Source: Brookings analysis of Internal Revenue Service data

		% Receiving EITC		
City		TY 2000	TY 2003	Change
1	Antioch	9.1%	13.3%	4.2%
2	Milpitas	6.1%	9.9%	3.8%
3	San Gabriel	20.4%	24.2%	3.8%
4	Monterey Park	17.9%	21.7%	3.8%
5	San Leandro	8.4%	11.7%	3.3%
6	Fairfield	12.6%	15.8%	3.2%
7	Union City	8.0%	11.0%	3.0%
8	Rosemead	26.6%	29.6%	3.0%
9	Santa Clara	4.5%	7.4%	2.9%
10	Vista	15.7%	18.6%	2.9%

Using IRS data

#2—Examine the characteristics of low-income working families

Filing status of EITC recipients by county, CA, 2003

Source: Brookings analysis of Internal Revenue Service data

Highest in Married-Couple Families

County	% MFJ
Sutter	40.9%
Modoc	40.8%
Glenn	40.5%
Colusia	40.5%
Imperial	40.3%
Yuba	38.0%
Sierra	37.4%
Mariposa	37.2%
Tehama	36.2%
Siskiyou	36.0%
CA Total	**28.5%**

Highest in Single-Parent Families

County	%HH
Tulare	53.7%
Inyo	53.1%
Fresno	52.2%
Solano	51.8%
Kings	50.2%
Alpine	49.4%
San Benito	48.4%
San Bernardino	48.1%
San Joaquin	48.1%
Sacramento	47.8%
CA Total	**45.6%**

Using IRS data

#3—Show income changes across time

Change in median AGI, CA cities, 2000-2003

Source: Brookings analysis of Internal Revenue Service data

*in 2003 dollars

	City	2000*	2003	Change
1	Indio	21,672	24,167	11.5%
2	Fairfield	34,102	37,540	10.1%
3	Madera	22,140	23,788	7.4%
4	Clovis	33,084	35,386	7.0%
5	Perris	22,856	24,378	6.7%
204	Mountain View	57,986	49,597	-14.5%
205	Laguna Hills	37,711	32,249	-14.5%
206	Los Gatos	74,003	62,505	-15.5%
207	Los Altos	99,707	79,459	-20.3%
208	San Rafael	44,585	35,025	-21.4%
	CA Total	31,748	30,981	-2.4%

Using IRS data

#4—Demonstrate the impact of similar state investments

Value of a 20% refundable state EITC, CA counties, 2006*

Source: Brookings analysis of Internal Revenue Service data

*based on 2003 receipt, adjusted for projected value of EITC nationwide in 2006

	County	Recip.	Value ($000s)	CA Estimate ($000s)
1	Los Angeles	815,020	1,483,859	296,772
2	San Diego	180,790	303,838	60,768
3	San Bernardino	152,363	291,973	58,395
4	Orange	153,946	255,893	51,179
5	Riverside	132,678	250,971	50,194
6	Fresno	84,030	166,430	33,286
7	Sacramento	83,953	145,708	29,142
8	Kern	67,752	136,790	27,358
9	Alameda	68,216	103,971	20,794
10	Santa Clara	62,216	91,350	18,270
	CA Total	2,384,888	$4,206,081	$841,216

Next Steps

- Obtain data from IRS SPEC Territory Manager: http://www.cbpp.org/eitc-partnership/territory-manager.htm
- Read guidance, ask questions: aberube@brookings.edu
- Consider what topics might make sense for your organization to study further
- Provide feedback to IRS, me about what additional information might be useful
- File your 1040 for 2005!

www.brookings.edu/metro/eitc