Improving the Accuracy and Reliability of ACS Estimates for Non-Standard Geographies Used in Local Decision Making

Warren Brown, Joe Francis, Xiaoling Li,
and Jonnell Robinson

Cornell University
Program on Applied
Demographics

Outline

- Goal – Accurate and Reliable Estimates
- Urban Neighborhoods and Rural Areas
- Problem 1: Standard Errors Too Large
- Problem 2: Spatial Mismatch
- No Perfect Solutions – Best Approximation and Proceed With Caution

Quality Data for Local Decisions

Greater Syracuse Tenants Network

TENANT POWER

Forever Affordable

ADIRONDACK COMMUNITY HOUSING TRUST

Urban

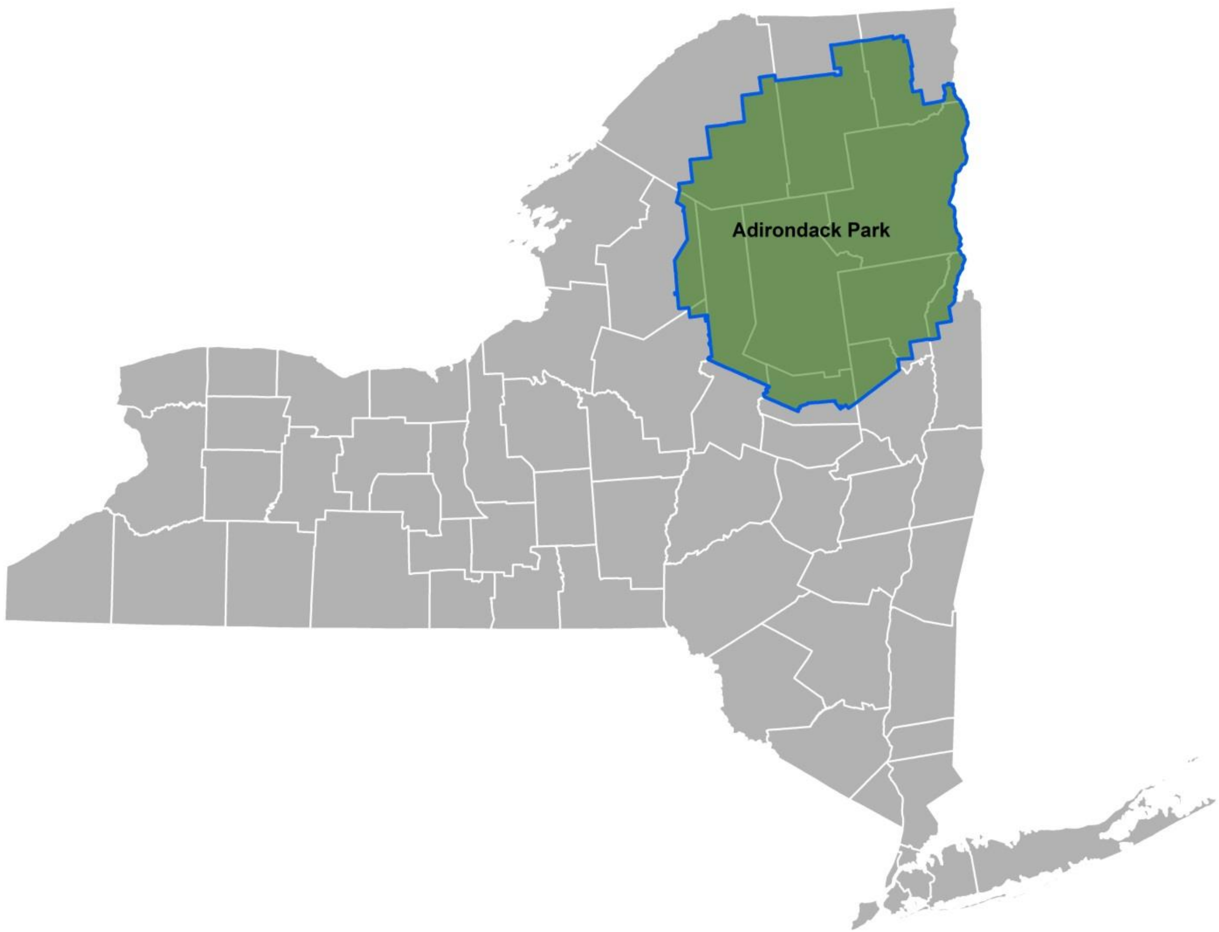
Rural

Urban Neighborhoods

City of Syracuse and "Tomorrow's Neighborhoods Today" (TNT)

Rural Areas

Adirondack Park in New York State



Problem #1 Unreliable Estimates

Small Samples

+

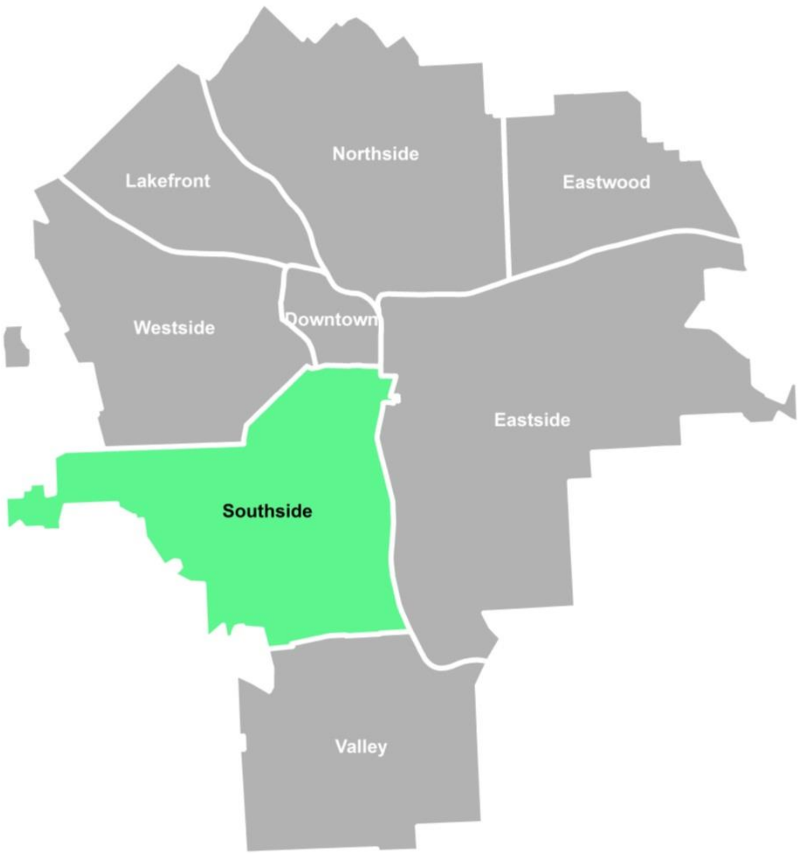
Small Areas

=

Large Standard error

Example Areas Illustrating Problems

Syracuse: Southside TNT



Adirondack Park: Essex County

Measures of Reliability

- Standard Error (SE) = Std Dev / $\sqrt{n}$

- Margin of Error (90% CI) = 1.645 x SE

- Coefficient of Variation (%) = 100 x (SE/Estimate)

Coefficient of Variation

Expresses Standard Error as a Percentage of the Estimate

No hard and fast rules, but the lower the better

- CV < 15% **Good**
- CV 15% - 29% **Fair**
- CV > 30% **Poor**

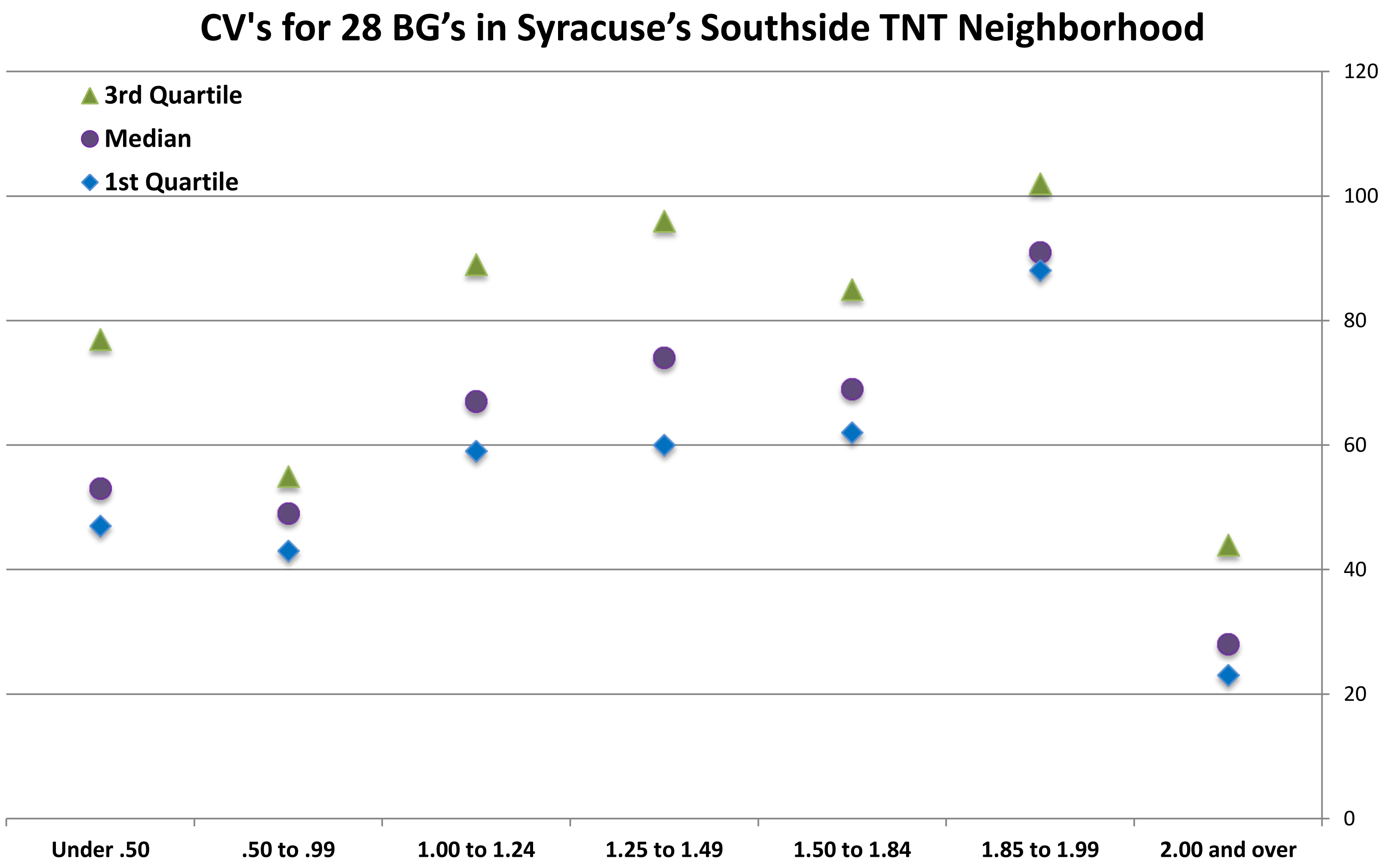
This is the measure we are using to assess reliability of the ACS estimates.

ACS 2008-2012 Estimates: Ratio of Income to Poverty Level (Table C17002)

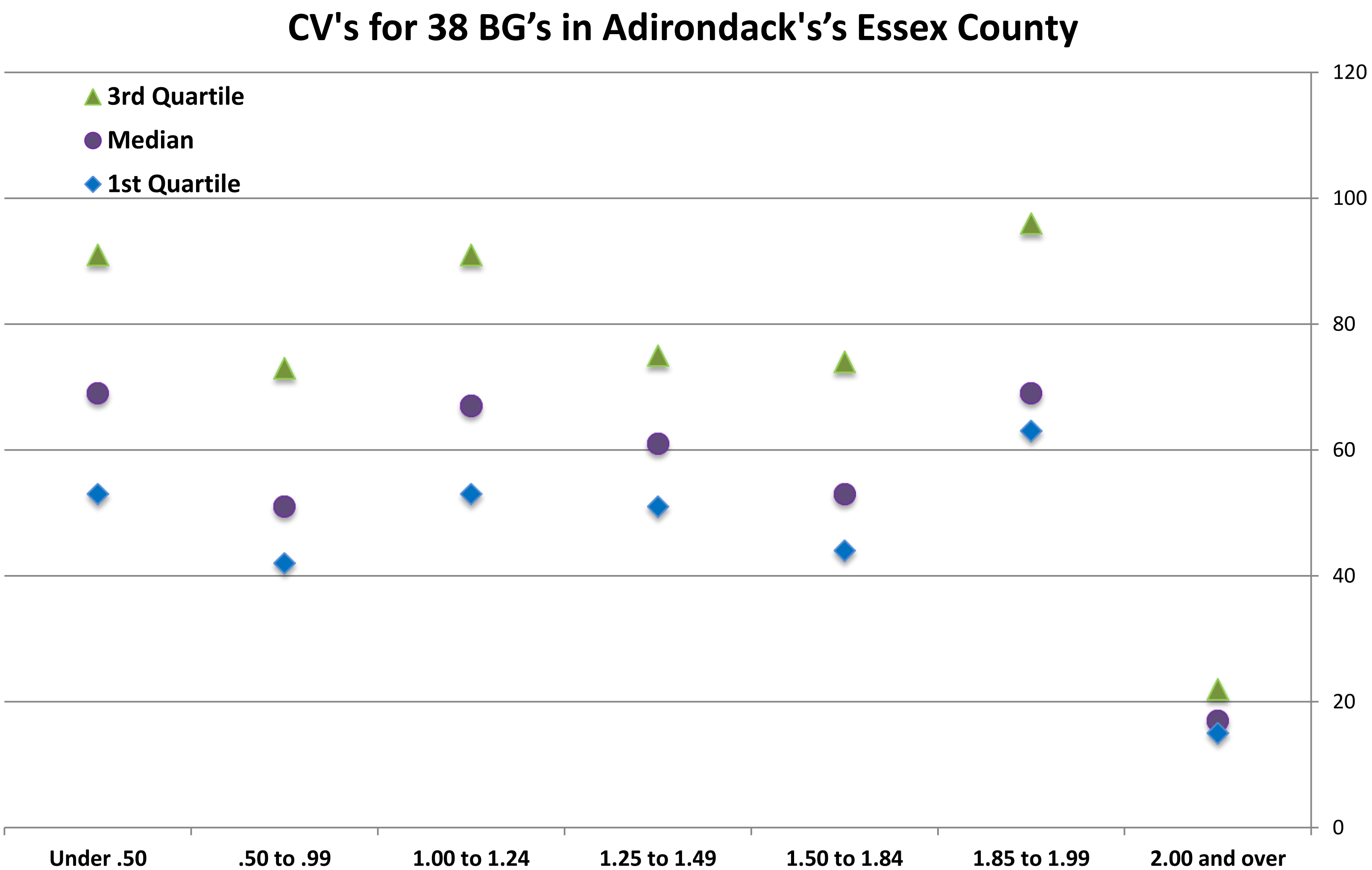
CV's for 28 BG's in Syracuse's Southside TNT Neighborhood

Block Group	Under .50	.50 to .99	1.00 to 1.24	1.25 to 1.49	1.50 to 1.84	1.85 to 1.99	2.00 and over
50002	n/a	62	n/a	96	94	n/a	13
53001	n/a	53	61	52	54	106	31
48002	365	47	n/a	67	69	n/a	14
48001	152	91	94	n/a	86	n/a	15
59002	74	51	87	65	62	91	28
54003	57	56	182	87	92	n/a	52
49002	73	95	122	96	69	77	28
51003	54	52	60	59	74	n/a	37
52001	49	41	80	61	94	n/a	47
52003	81	38	72	83	65	81	48
57001	78	53	102	95	66	111	20
57002	42	51	87	99	60	81	22
61011	53	46	53	45	85	n/a	26
52002	83	48	59	79	52	n/a	26
50001	52	69	n/a	64	62	102	19
51002	47	46	56	n/a	54	203	26
59001	40	111	91	n/a	72	90	41
54002	48	42	32	98	69	n/a	33
51001	83	48	67	n/a	n/a	91	28
58002	41	45	63	102	88	n/a	42
58003	49	55	92	58	69	n/a	28
54001	82	51	74	98	90	93	54
42002	45	22	67	48	77	n/a	37
49001	51	60	43	66	117	n/a	24
54004	59	42	50	59	54	91	66
42001	34	41	66	85	73	101	59
58001	47	42	89	97	75	n/a	18
53002	32	43	59	69	58	88	67

ACS 2008-2012 Estimates: Ratio of Income to Poverty Level (Table C17002)



ACS 2008-2012 Estimates: Ratio of Income to Poverty Level (Table C17002)



Simple Solution: Combine and Collapse

Increase the effective sample size by:

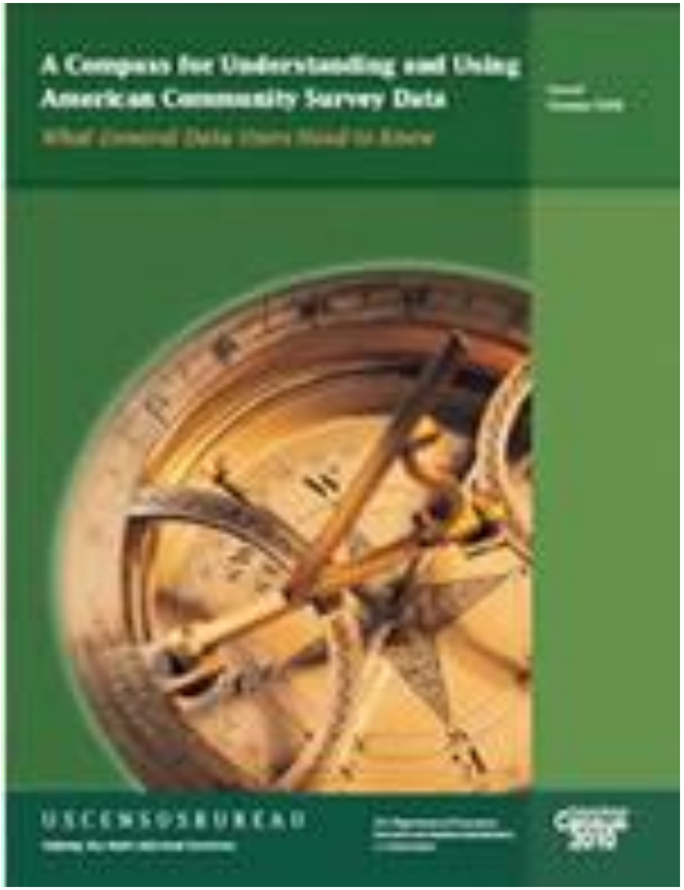
- Combining geographic areas
- Collapsing detailed categories

Formula to approximate combined/collapsed standard error:

$$SE(\hat{X}_1 \pm \hat{X}_2) \approx \sqrt{\left[SE\left(\hat{X}_1\right)\right]^2 + \left[SE\left(\hat{X}_2\right)\right]^2}$$

Census Bureau References

Compass Series



ACS Methods Page

Accuracy of the Data

A basic explanation of the sample design, estimation methodology, and accuracy of the data

- 2010-2012 & 2008-2012 Multiyear Accuracy (US) [PDF 319KB]
- 2010-2012 & 2008-2012 Multiyear Accuracy (Puerto Rico) [PDF 361KB]
- 2012 ACS 1-year Accuracy of the Data (US) [PDF 625KB]
- 2012 PRCS 1-year Accuracy of the Data (Puerto Rico) [PDF 604KB]

ACS Estimates Aggregator

http://www.psc.isr.umich.edu/dis/acs/estimates_aggregator/

University of Michigan

PSC

Population Studies Center

INSTITUTE FOR SOCIAL RESEARCH

RESEARCH | PEOPLE | PUBS | NEWS & EVENTS | TRAINING | DATA | INFO SERVICES | CENTER SERVICES | ABOUT

Home > Data Servic

ACS Estimates Aggregator

Instructions

This tool allows one to either (a)combine counts across several units of geography or to (b)collapse cells in a table to generate new estimates and margins of error. This is useful for creating robust statistics for health areas, neighborhoods, state economic areas, etc. Follow link for further discussion of **ACS Data Quality issues**.

Copy and paste the information from output from American FactFinder: Name, Estimates & Margin of Error pairs. Replace the sample tab-delimited data below.

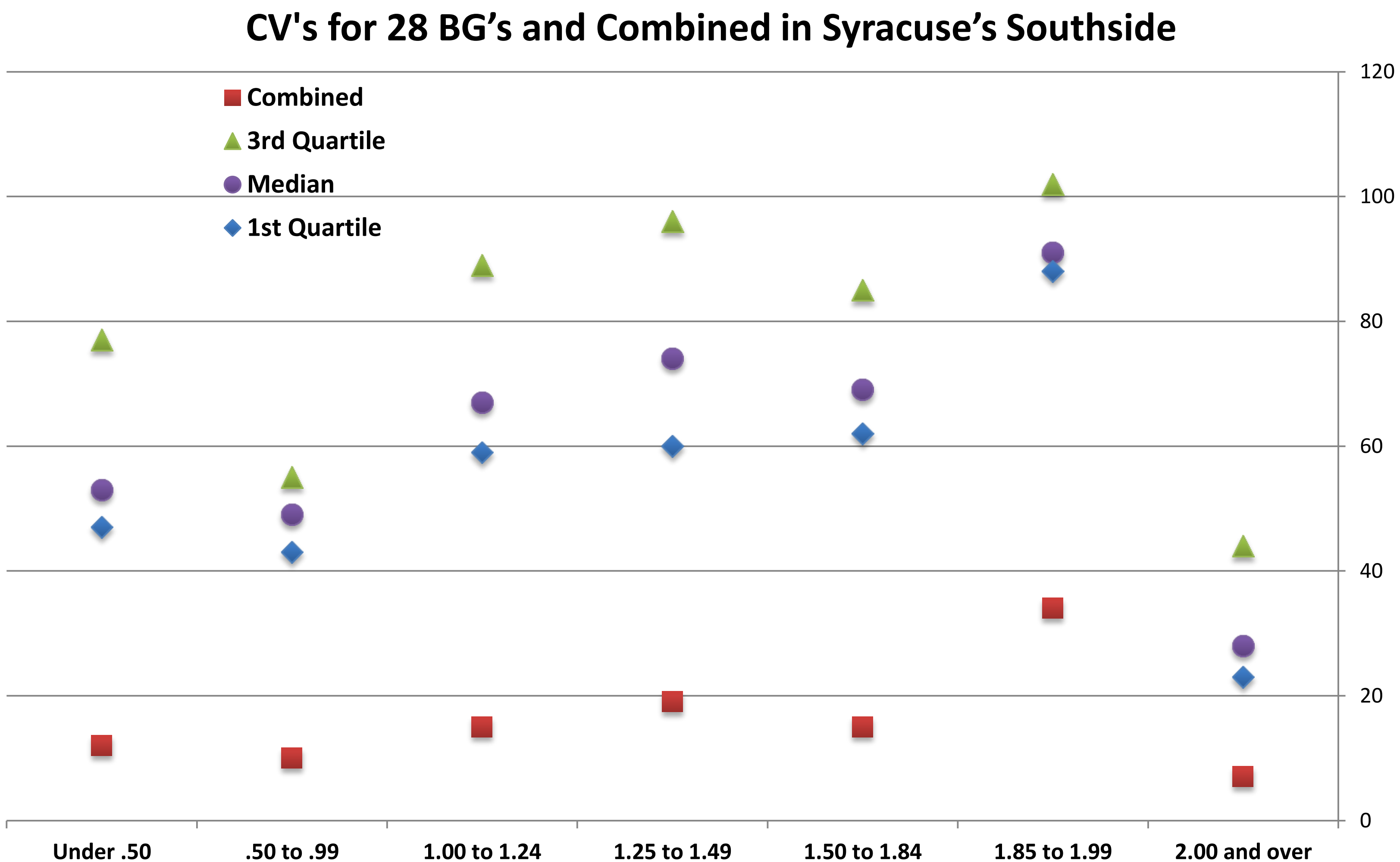
```
Name (tab) Estimate1 (tab) MoE1 (tab) Estimate2 (tab) MoE2 (tab) Estimate3 (tab) MoE3...
```

```
a    286  70   876  56   286  70
b    102  40   456798    102  40
c    399  60   456  39   399  60
d    77   29   864  34   77   29
```

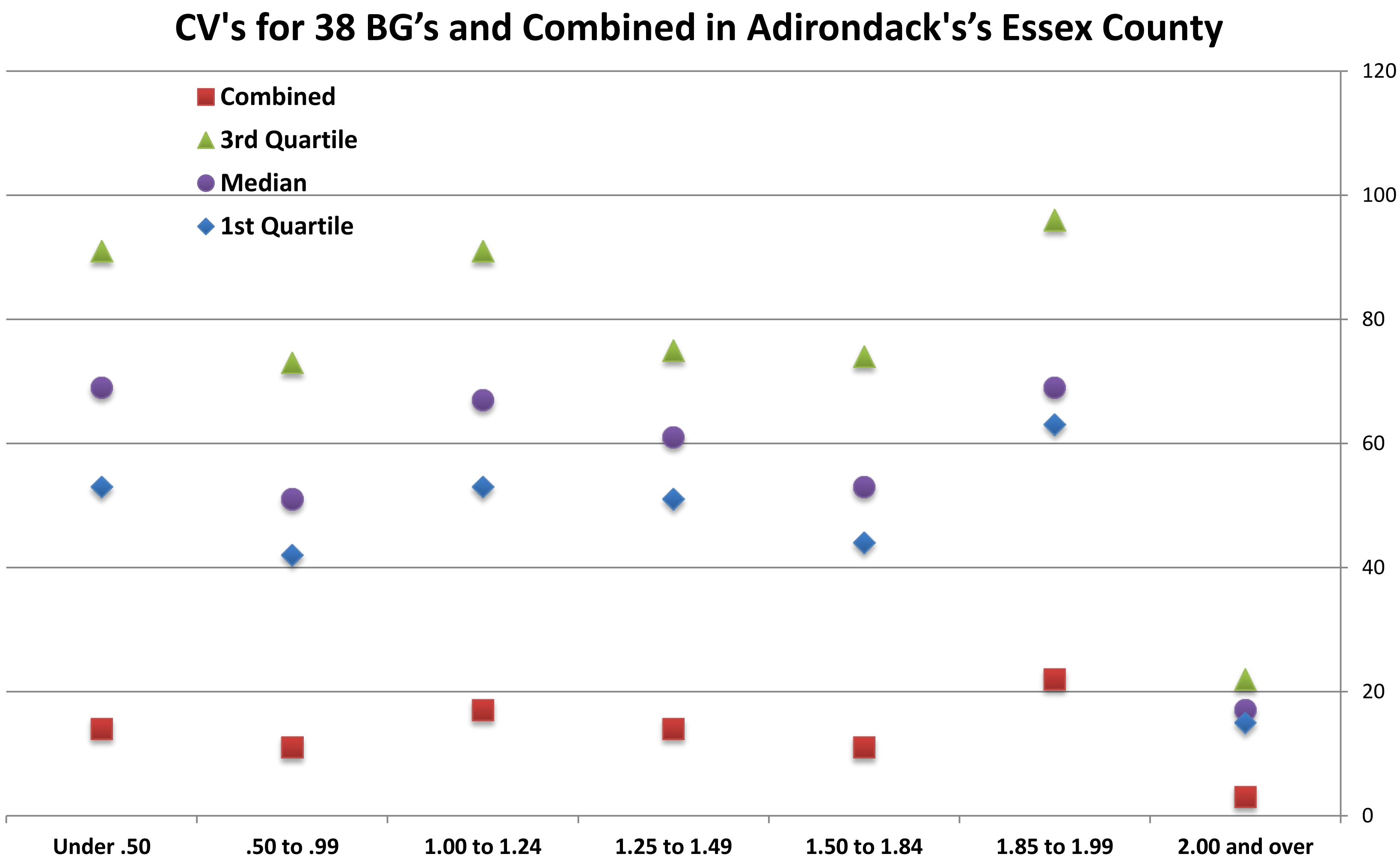
Counts

Formulas are drawn from Appendix 11 from the U.S. Census Bureau, A Compass for Understanding and Using American Community Survey Data.

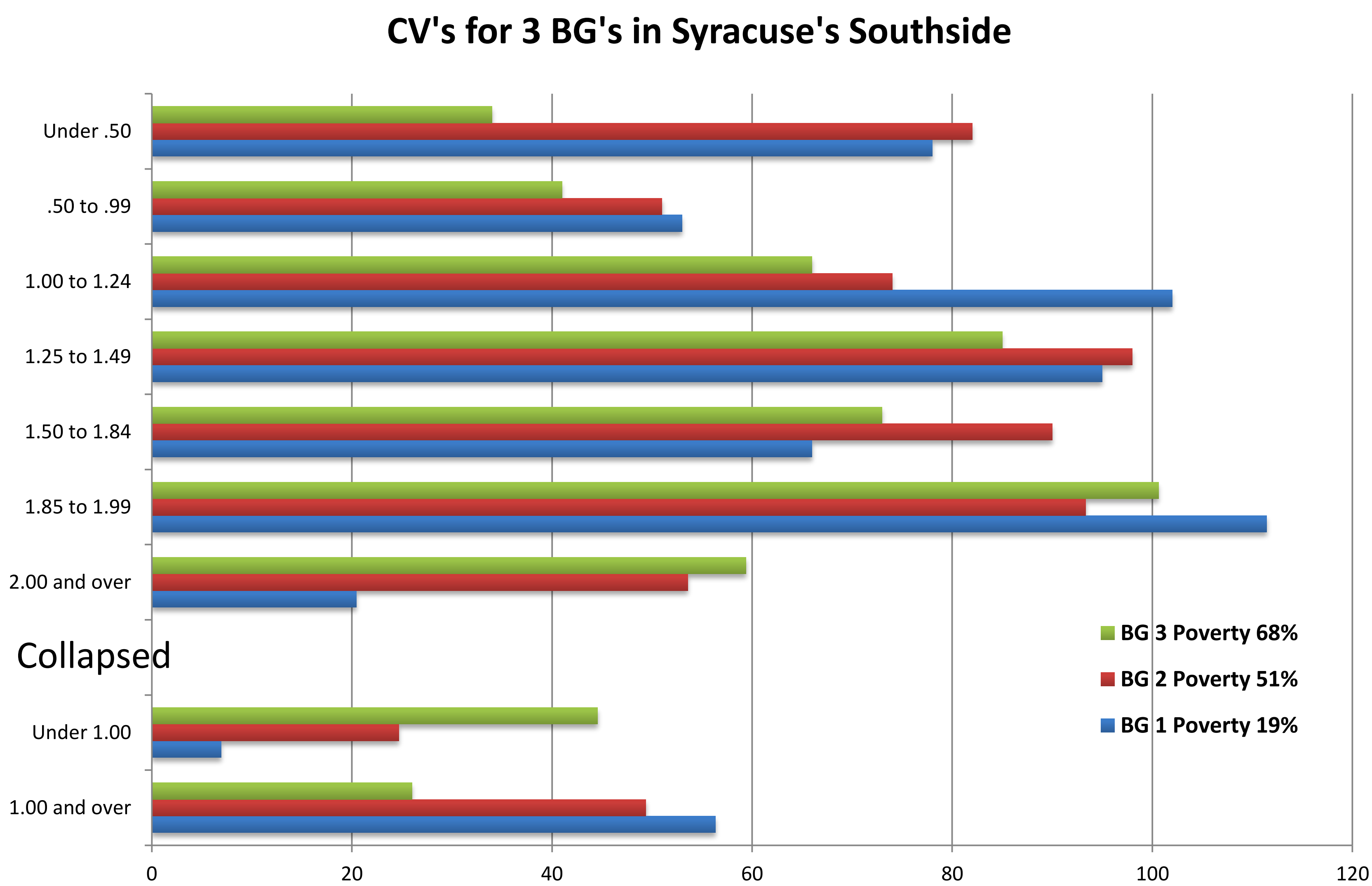
Combine Block Groups



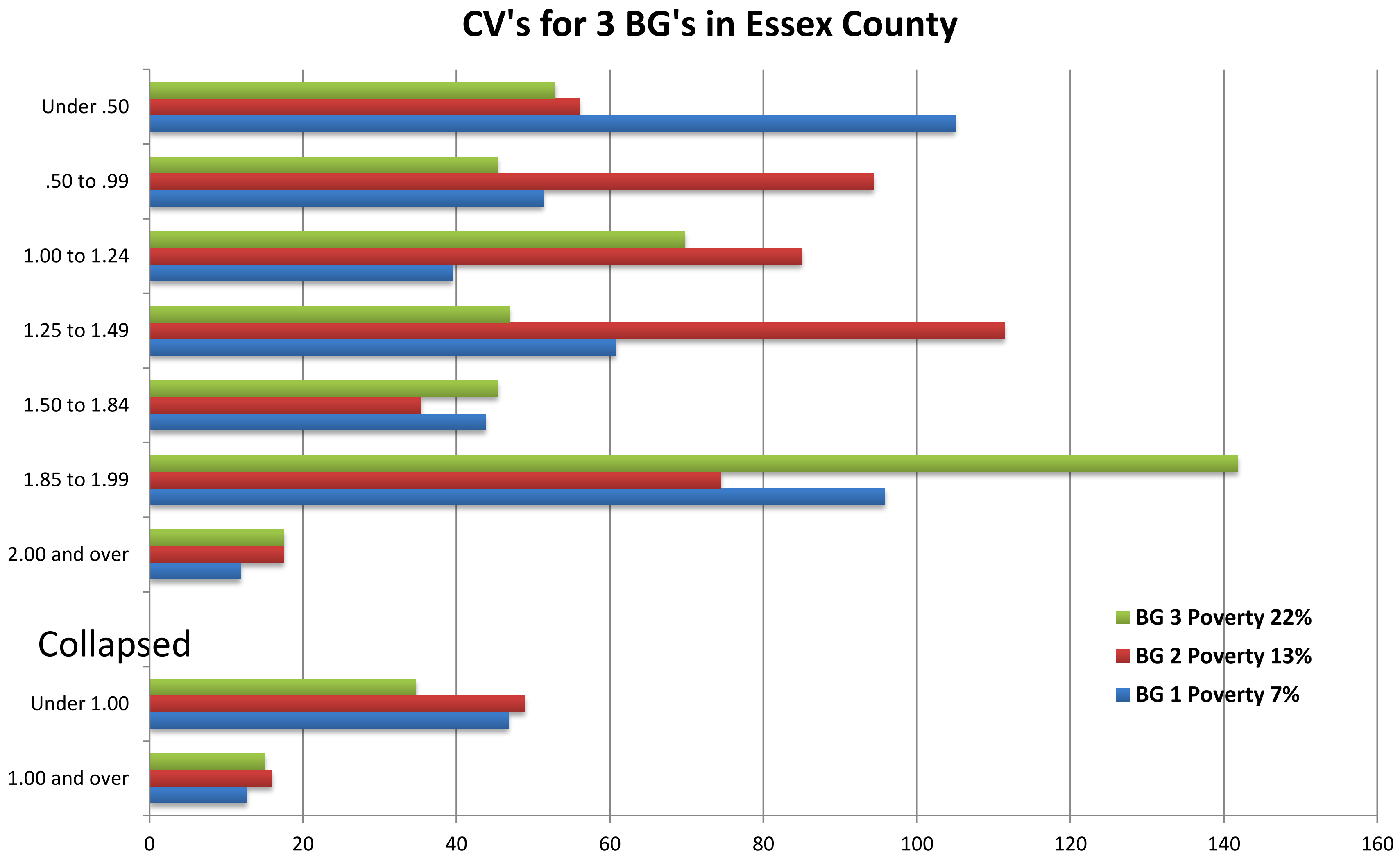
Combine Block Groups



Collapse Categories



Collapse Categories



Problem Solved? – Not Really

- Simple solutions to sampling error render "approximate" solutions with no accurate means to assess quality of the new estimates.
- Not able to determine statistically significant differences between:
  - Two or more areas
  - Change over time for one area

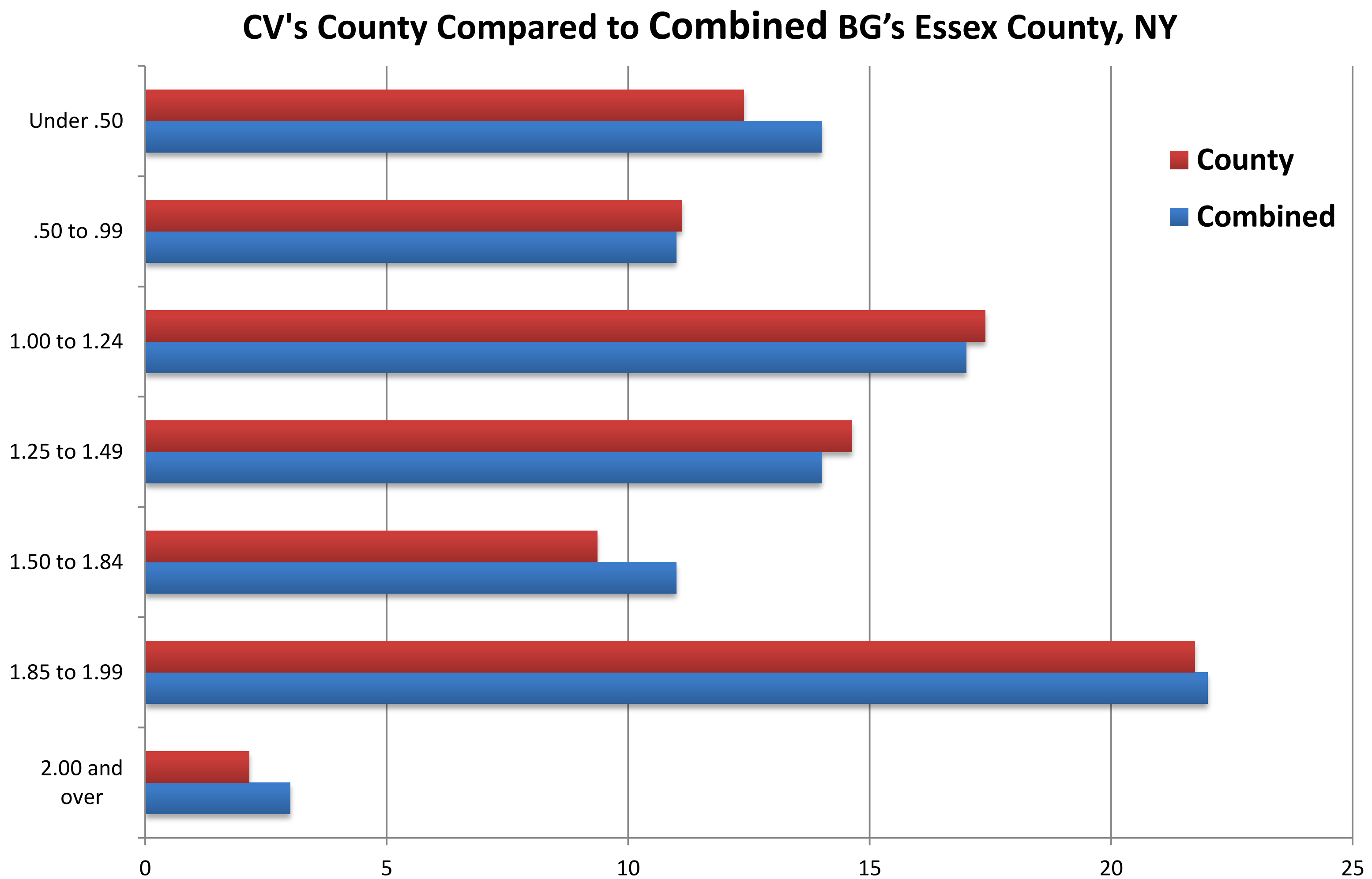
Bias Due to Missing Term

Bias in calculation of Standard Error due to the absence of a covariance term.

$$SE(\hat{X}_1 \pm \hat{X}_2) = \sqrt{\left[SE(\hat{X}_1)\right]^2 + \left[SE(\hat{X}_2)\right]^2 \pm 2cov(\hat{X}_1, \hat{X}_2)}$$

Direction of bias may be positive or negative depending on the sign of the covariance.

Assess how much error



Proceed with Caution

- Use the largest type of census geography possible
- Use a collapsed version of a detailed table
- Create estimates and SEs using the Public Use Microdata Sample (PUMS)
- Request a custom tabulation, a fee-based service offered under certain conditions by the Census Bureau.

Problem #2 Square Peg in a Round Hole

Boundaries of planning areas don’t match standard census geography

Spatial Mismatch

A common problem faced by demographers dealing with local areas is that:

1. Geographies of interest (e.g. neighborhoods, watershed boundaries, protected land preserves, local labor markets) don't conform to Census Geographies like tracts or block groups.
2. Hence published tract or block group summary statistics for those geographies of interest aren't accurate.
3. This problem is present whether dealing with decennial census, ACS or annual estimates data.

Here we will be dealing with 2008-12 ACS data.

Spatial Mismatch

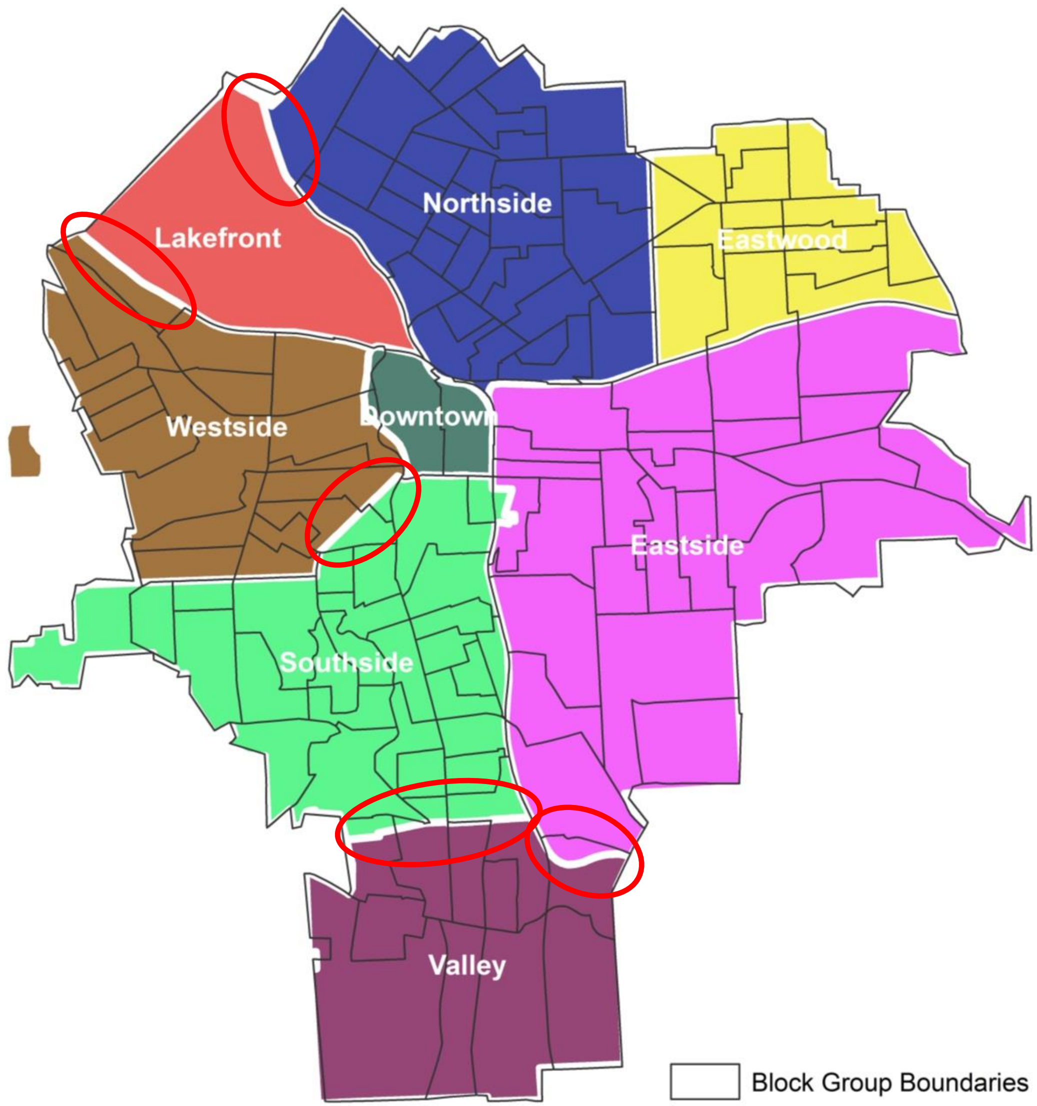
If block group or tract ACS information, like housing units or population characteristics, are not allocated when the Block Group or tract is intersected by a boundary of interest then some proportion of those block group/tract data are assigned incorrectly to the wrong geography.

Four possible approaches that have been taken:

- Completely Ignore the mismatch; hope for best
- Pick some Block Groups to include
- Systematic Area proportional weighting
- Dasymetric mapping

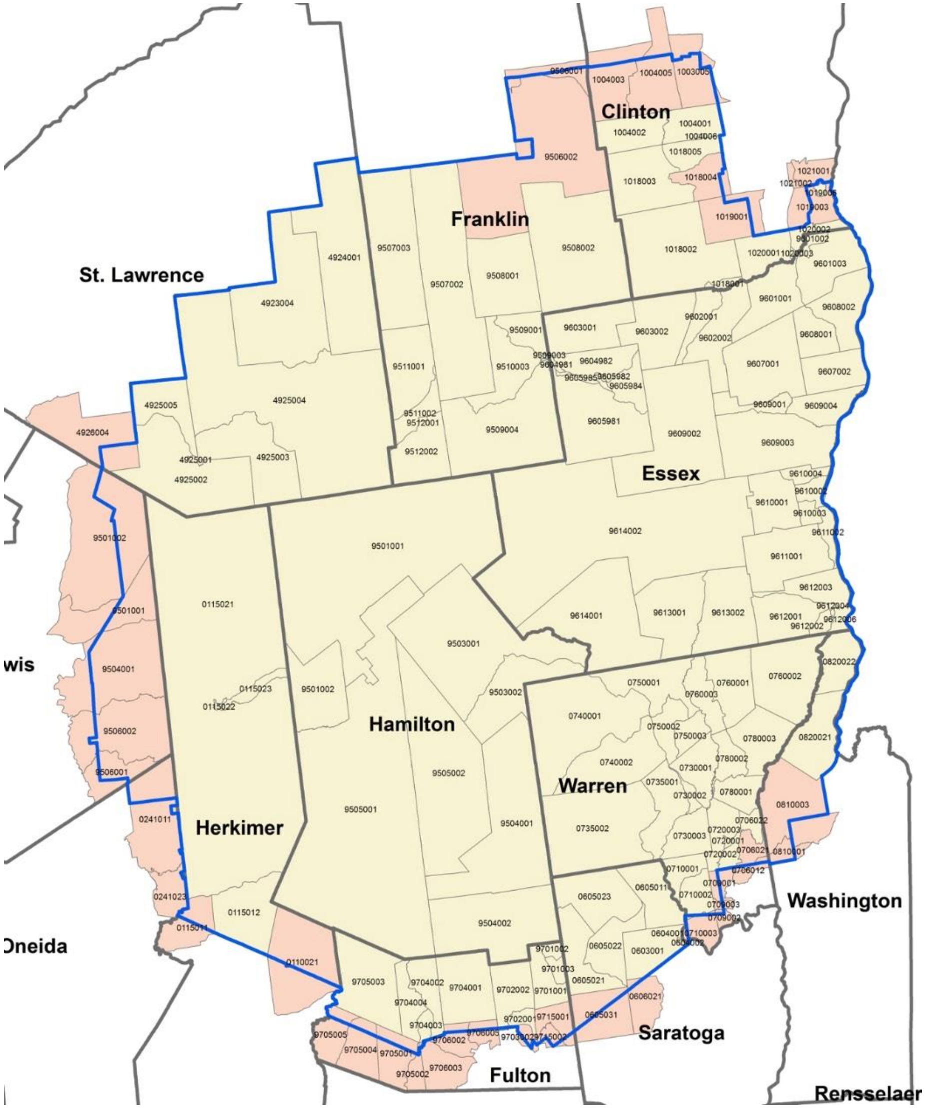
Case 1: Syracuse TNT Zones

Miss-Match of TNT Zones and Block Groups



Adirondack Park Boundary

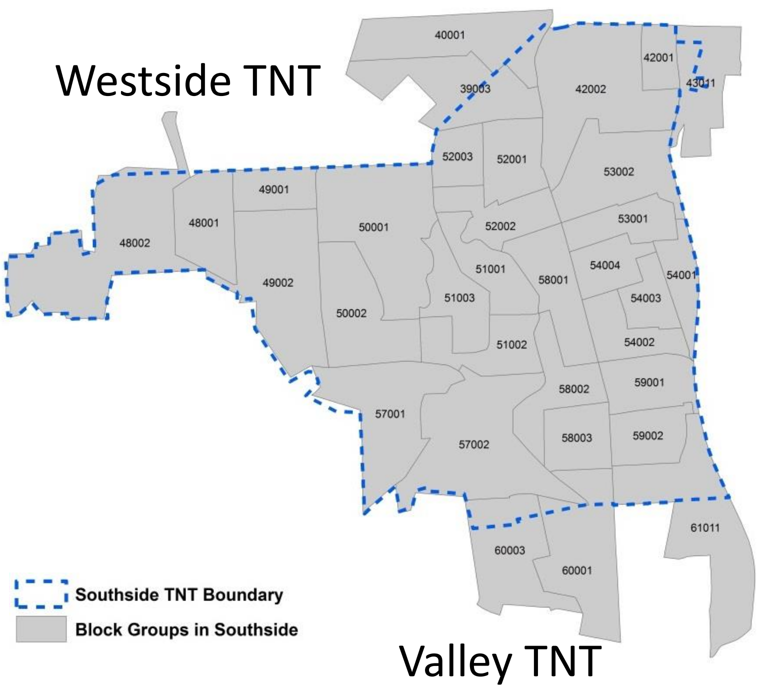
Park Boundary, the Blue Line, intersects Block Groups



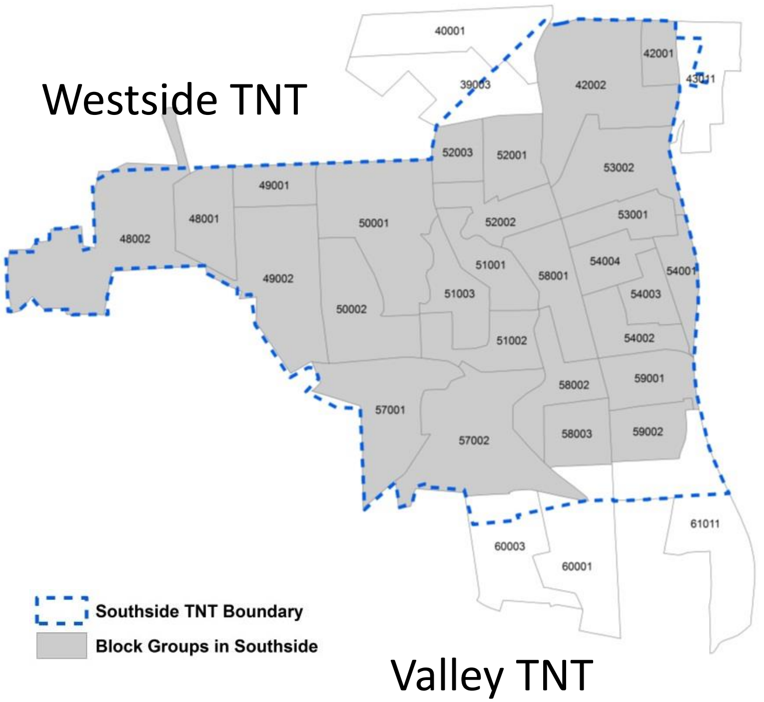
Ignore the Mismatch

May work if small amount of boundary mismatch but causes increasing amount of error in direct relationship to amount of mismatch.

Option A: Include if Crossed



Option B: Exclude if not Totally Inside



Ignore the Mismatch

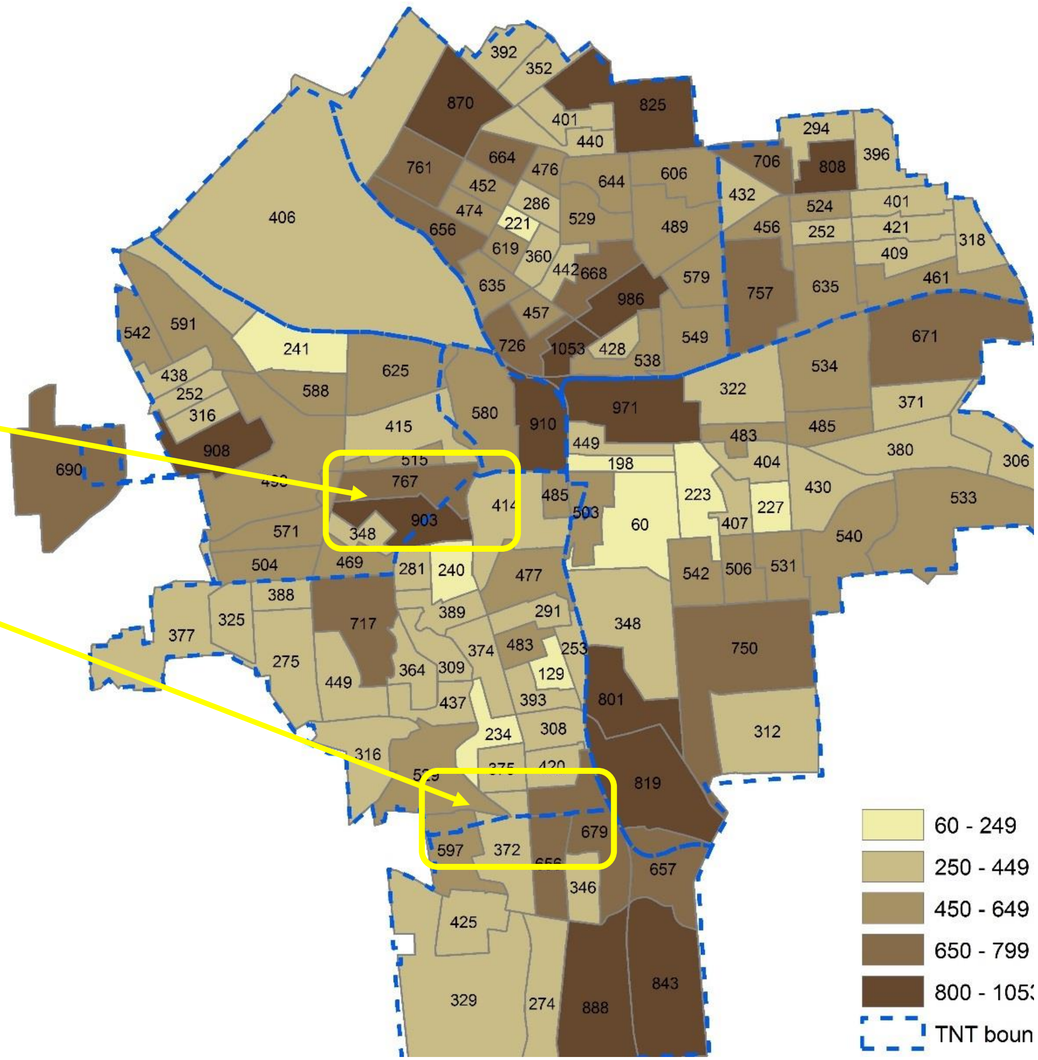
Southside TNT HUs for BG
Totally within: 10032

Option A:

Include crossed BGs—3318

40001→767 HUs
39003→903 HUs
60003→597 HUs
60001→372 HUs
61011→679 HUs

Southside TNT HUs: 13350
for 33.1% increase



Option B:

Exclude BGs—3318

Southside TNT HUs: 10032

Pick Some BGs to Include

Researcher may select some but not all BGs to include.

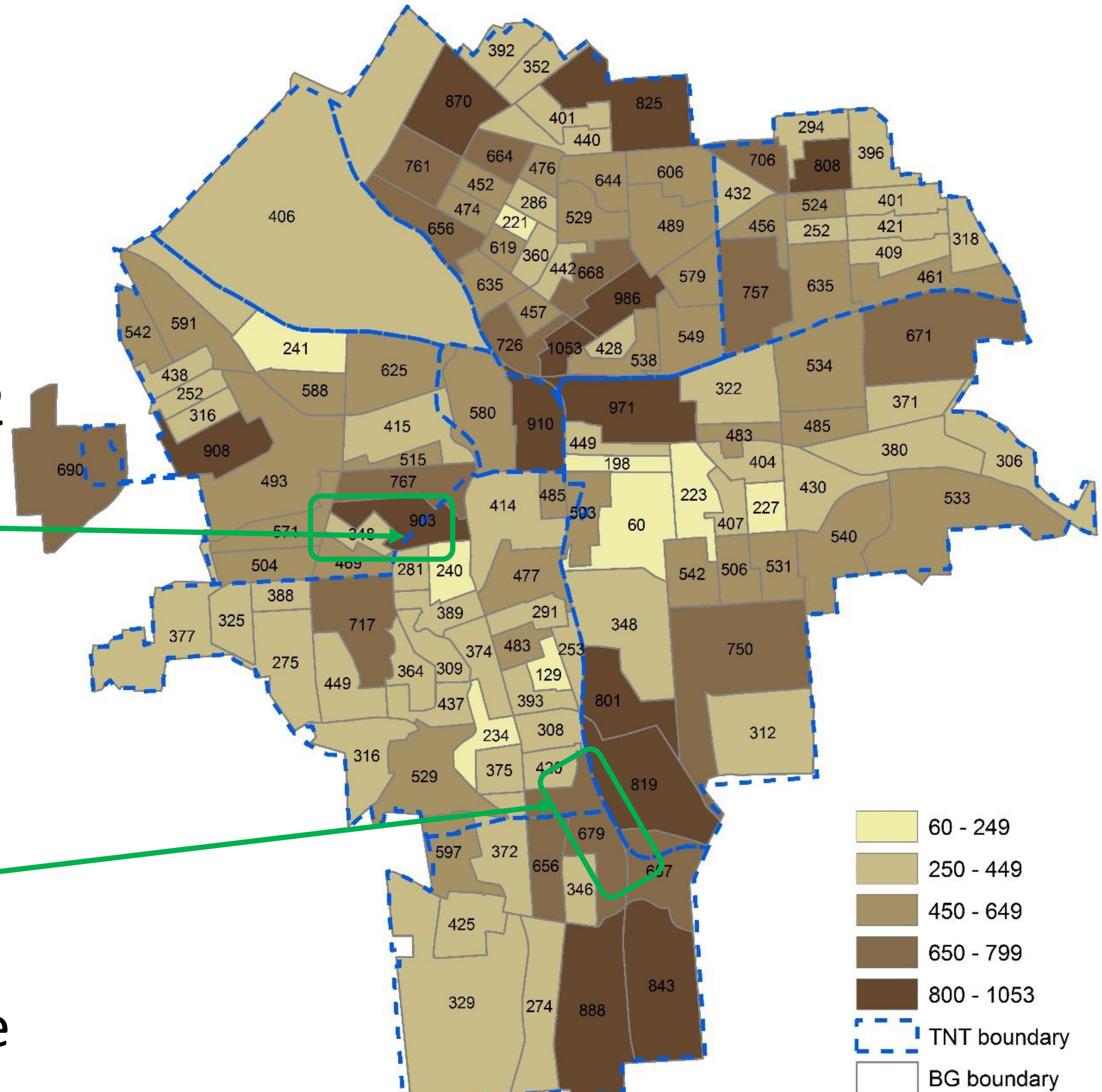
Southside TNT HUs for BG Totally within: 10032

Include BG 39003: 903
10032 + 903= 10935
for 9% increase

Or

Include BG 61011: 679
10032 + 679= 10711

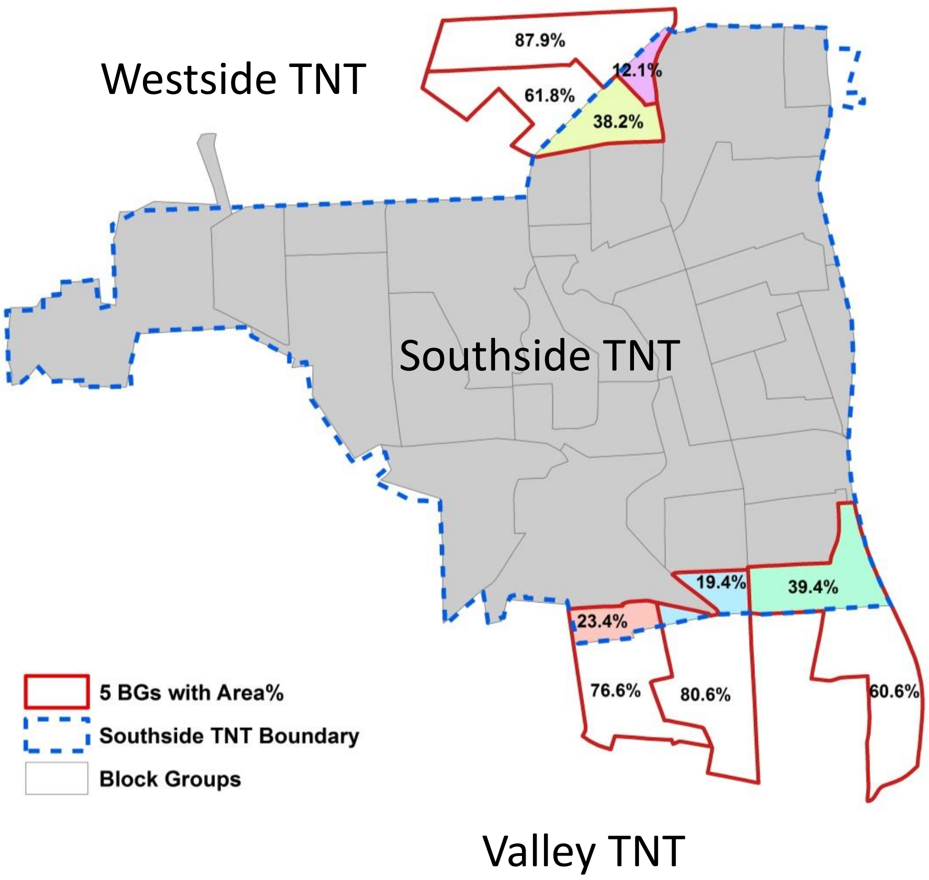
for 6.8% increase



Area Proportional Allocation

Area Proportional Weighted allocation" where the proportion of a block group's land area falling inside the boundary of the area of interest (e.g. TNT) is used to proportionally allocate the population.

However this procedure assumes that the land area in the block group is equally usable and used. Yet we know this not always the most accurate reflection of actual land usage in lots of block groups and tracts.



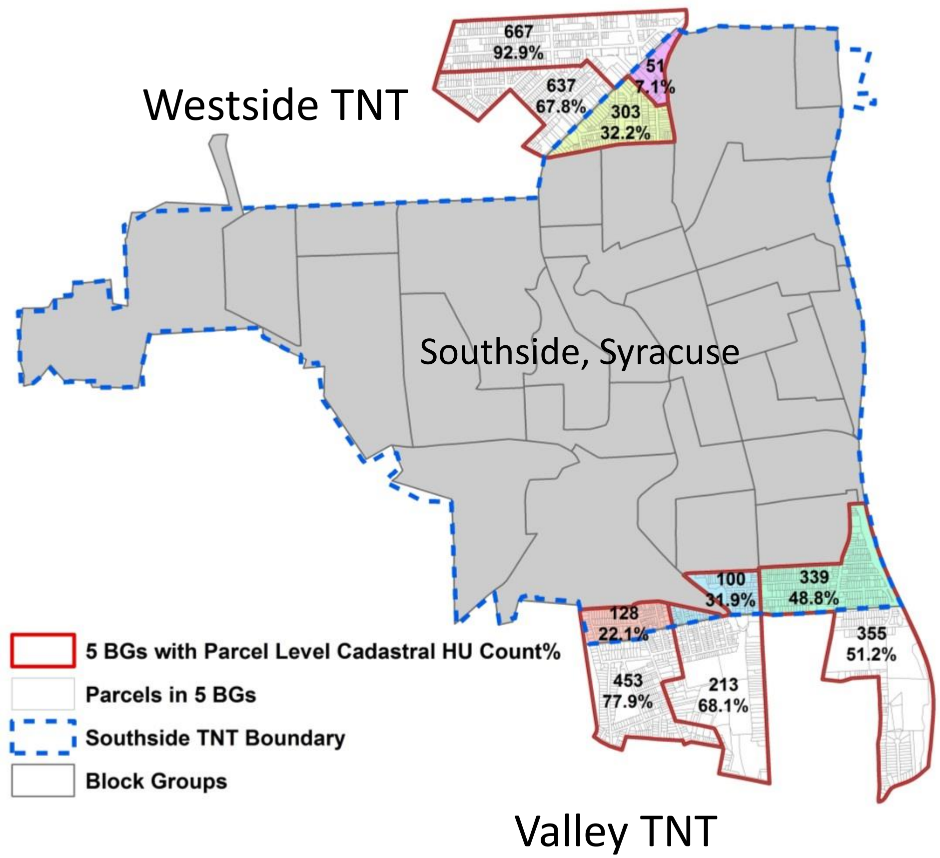
Area Proportional Allocation

To evaluate performance of area proportional allocation, compare the percentages of Census HUs in split block group with the percentage from ACS allocated via area proportional weighting.

Block Group	2010 Census HUs	ACS HU	Ground	Neighbor-hood	2010 Census HU%	2010 HUs	Area Weight %	Allocated ACS HUs Using Area%	Ground Verification	Ground %
39003	843	903	757	Southside	31%	260	38%	345	222	29%
				Westside	69%	583	62%	558	535	71%
40001	729	767	619	Southside	7%	48	12%	93	43	7%
				Westside	93%	681	88%	674	576	93%
60001	311	372	317	Southside	32%	99	19%	72	109	34%
				Valley	68%	212	81%	300	208	66%
60003	592	597		Southside	20%	119	23%	140	127	
				Valley	80%	473	77%	457	?	
61011	677	679	572	Southside	51%	346	39%	268	310	54%
				Valley	49%	331	61%	411	262	46%

Dasymetric Mapping

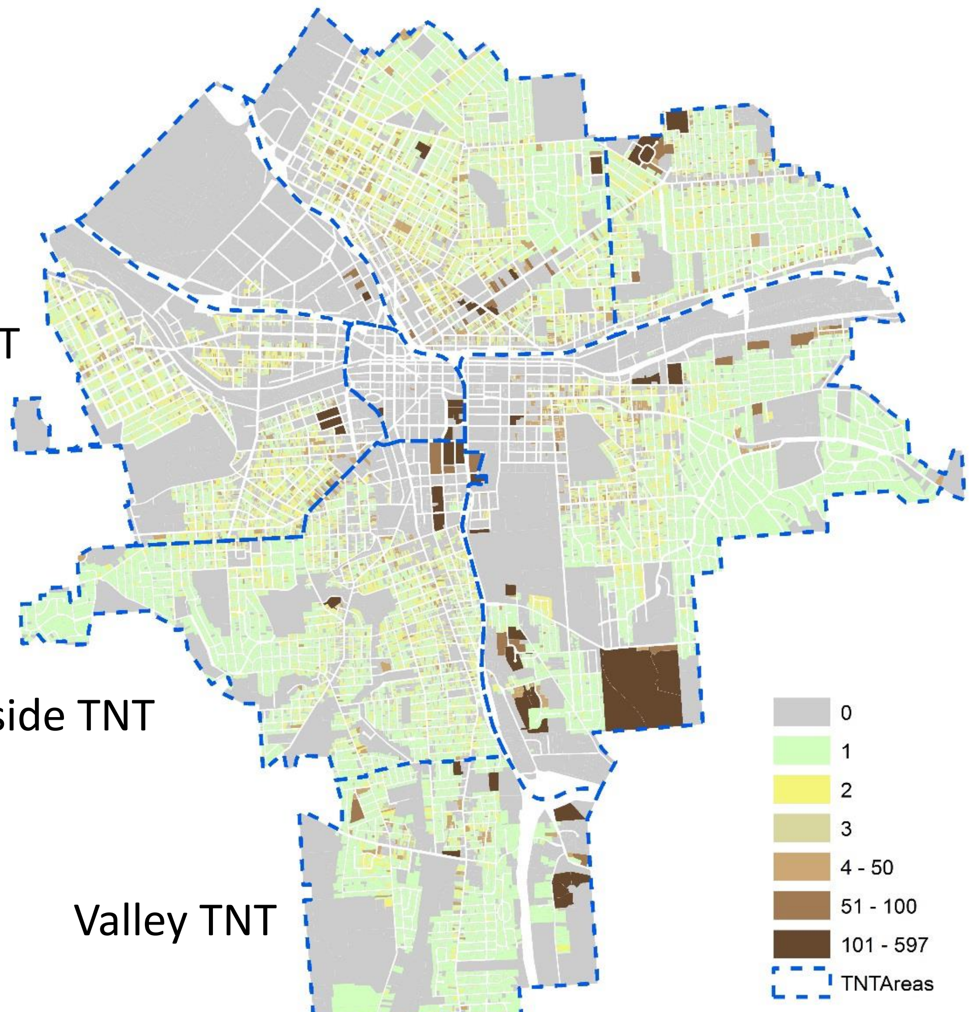
Dasymetric mapping is generally a better solution. It uses administrative records like data on land use of property tax records in an urban setting. Knowing where in a block group residences are and are not allows dasymetric mapping to improve the decisions about inclusions /exclusions of HUs, and error of those decisions.



Dasymetric Mapping

As this tax parcel map shows, sometimes one can determine for each tax parcel not only whether it is residential (not gray) but type of residential unit.

Housing Units Map -- Parcel Level



Dasymetric Mapping Allocation

To evaluate performance of the dasymetric mapping allocation, compare the percentages of Census HUs in split block group with the percentage from ACS allocated via dasymetric mapping procedures.

Block Group	2010 Census HUs	ACS HU	Ground	Neighbor -hood	2010 Census HU%	2010 HUs	Dasymetric %	Allocated ACS HUs Using Dasymetric %	Ground Verification	Ground %
39003	843	903	757	Southside	31%	260	32%	291	222	29%
				Westside	69%	583	68%	612	535	71%
40001	729	767	619	Southside	7%	48	7%	54	43	7%
				Westside	93%	681	93%	713	576	93%
60001	311	372	317	Southside	32%	99	32%	119	109	34%
				Valley	68%	212	68%	253	208	66%
60003	592	597		Southside	20%	119	22%	132	127	
				Valley	80%	473	78%	465	?	
61011	677	679	572	Southside	51%	346	49%	332	310	54%
				Valley	49%	331	51%	347	262	46%

Which Procedure is Better?

No perfect solution. However, several findings of note:

1. In every instance percentages from Dasymetric allocation are closer to percentage of 2010 Census HUs in each split BG.

Block Group	2010 Census HUs	ACS HU	Ground	Neighbor-hood	2010 Census HU%	2010 HUs	Dasymetric %	Allocated ACS HUs Using Dasymetric%	Area %	Allocated ACS HUs Using Area%	Ground Verification	Ground %
39003	843	903	757	Southside	31%	260	32%	291	38%	345	222	29%
				Westside	69%	583	68%	612	62%	558	535	71%
40001	729	767	619	Southside	7%	48	7%	54	12%	93	43	7%
				Westside	93%	681	93%	713	88%	674	576	93%
60001	311	372	317	Southside	32%	99	32%	119	19%	72	109	34%
				Valley	68%	212	68%	253	81%	300	208	66%
60003	592	597		Southside	20%	119	22%	132	23%	140	127	
				Valley	80%	473	78%	465	77%	457	?	
61011	677	679	572	Southside	51%	346	49%	332	39%	268	310	54%
				Valley	49%	331	51%	347	61%	411	262	46%

Which Procedure is Better?

No perfect solution. However, several findings of note:

2. In all BGs, the percentages of HUs assigned to each split BG via Dasymetric allocation is closer to % via ground verification.

Block Group	2010 Census HUs	ACS HU	Ground	Neighbor-hood	2010 Census HU%	2010 HUs	Dasymetric %	Allocated ACS HUs Using Dasymetric %	Area %	Allocated ACS HUs Using Area%	Ground Verification	Ground %
39003	843	903	757	Southside	31%	260	32%	291	38%	345	222	29%
				Westside	69%	583	68%	612	62%	558	535	71%
40001	729	767	619	Southside	7%	48	7%	54	12%	93	43	7%
				Westside	93%	681	93%	713	88%	674	576	93%
60001	311	372	317	Southside	32%	99	32%	119	19%	72	109	34%
				Valley	68%	212	68%	253	81%	300	208	66%
60003	592	597		Southside	20%	119	22%	132	23%	140	127	
				Valley	80%	473	78%	465	77%	457	?	
61011	677	679	572	Southside	51%	346	49%	332	39%	268	310	54%
				Valley	49%	331	51%	347	61%	411	262	46%

Which Procedure is Better?

No perfect solution. However, several findings of note:

3. In all but one BG, the number of HUs allocated to each split BG via Dasymetric allocation is closer to 2010 Census HUs.

Block Group	2010 Census HUs	ACS HU	Ground	Neighbor-hood	2010 Census HU%	2010 HUs	Dasymetric %	Allocated ACS HUs Using Dasymetric%	Area %	Allocated ACS HUs Using Area%	Ground Verification	Ground %
39003	843	903	757	Southside	31%	260	32%	291	38%	345	222	29%
				Westside	69%	583	68%	612	62%	558	535	71%
40001	729	767	619	Southside	7%	48	7%	54	12%	93	43	7%
				Westside	93%	681	93%	713	88%	674	576	93%
60001	311	372	317	Southside	32%	99	32%	119	19%	72	109	34%
				Valley	68%	212	68%	253	81%	300	208	66%
60003	592	597		Southside	20%	119	22%	132	23%	140	127	
				Valley	80%	473	78%	465	77%	457	?	
61011	677	679	572	Southside	51%	346	49%	332	39%	268	310	54%
				Valley	49%	331	51%	347	61%	411	262	46%

Which Procedure is Better?

No perfect solution. However, several findings of note:

4. In all but two BG, the number of HUs assigned to each split BG via Dasymetric allocation is closer to ground verification.

Block Group	2010 Census HUs	ACS HU	Ground	Neighbor-hood	2010 Census HU%	2010 HUs	Dasymetric %	Allocated ACS HUs Using Dasymetric%	Area %	Allocated ACS HUs Using Area%	Ground Verification	Ground %
39003	843	903	757	Southside	31%	260	32%	291	38%	345	222	29%
				Westside	69%	583	68%	612	62%	558	535	71%
40001	729	767	619	Southside	7%	48	7%	54	12%	93	43	7%
				Westside	93%	681	93%	713	88%	674	576	93%
60001	311	372	317	Southside	32%	99	32%	119	19%	72	109	34%
				Valley	68%	212	68%	253	81%	300	208	66%
60003	592	597		Southside	20%	119	22%	132	23%	140	127	
				Valley	80%	473	78%	465	77%	457	?	
61011	677	679	572	Southside	51%	346	49%	332	39%	268	310	54%
				Valley	49%	331	51%	347	61%	411	262	46%

Westside TNT Neighborhood:
Where Dasymetric Mapping Didn't Work

Neighborhood
in Transition

Vacant
Housing

Public Housing

Vacant Lot

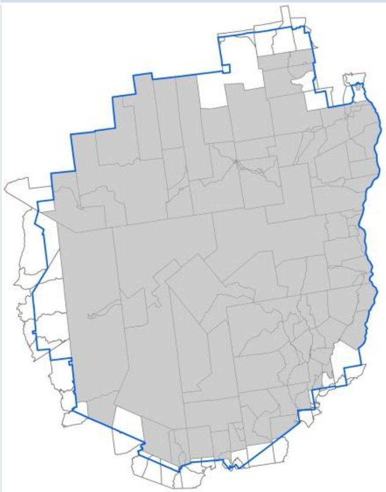
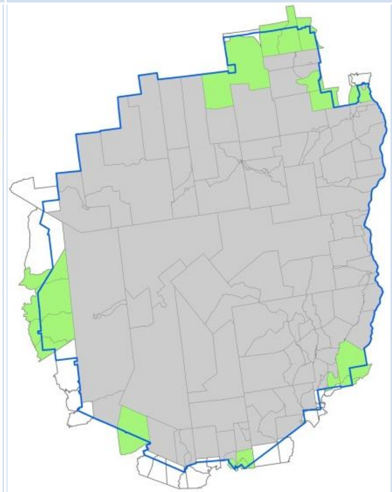
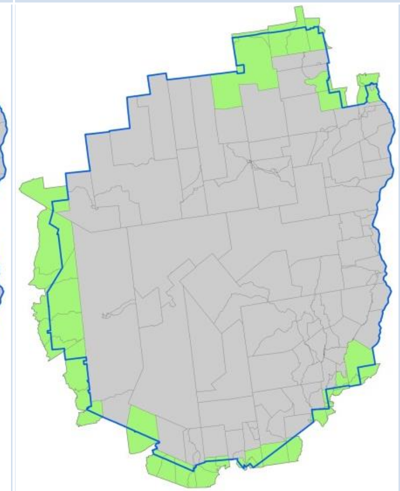
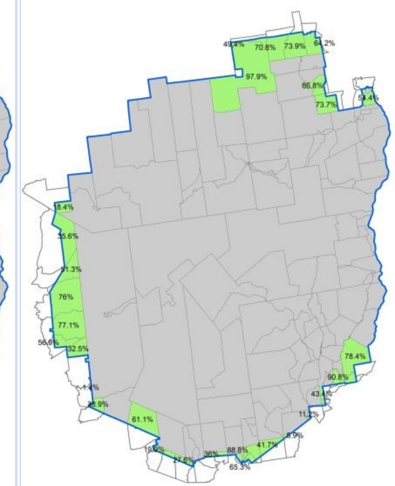
Valley TNT Neighborhood:
Where Dasymetric Mapping Worked Well

Stable, Semi-Suburban Neighborhood

Typical Streets

Newer Construction

Spatial Mismatch in Adirondack Park

	Ignore the Mismatch Approach			Area% Approach
	BGs in the park only	BGs of > 50% area in the park	BGs in and cross the park	BGs in and area% Allocation for crossing
				
Pop Estimate	116771	137569	167344	138663
% diff from 2010Census	-10.5%	5.4%	28.2%	6.3%
CV	1.22%	1.13%	1.00%	1.20%

Future Work

1. Conduct dasymetric mapping analysis for Adirondack Park
2. Compare allocation methods results
3. Compare cadastral dasymetric mapping with environmental constraint dasymetric mapping.
4. Explore use of these techniques for more complex task of allocating population by characteristics such as income and poverty.