Impacts of Sample Sizes in the American Community Survey

Northwestern University Transportation Center

City Drivers Stuck in Slow Lane (Chicago Tribune March 31, 2005)

"The ACS helps ease some of the data withdrawal experienced by hard-core census geeks, but not entirely."

Public Misunderstanding About ACS

☐ "The 10 year number is going to continue to be the gold standard."

☐ "We will have a lot of data (from ACS), but there is still nothing like the census itself."

Presentation Topics

- ☐ Review sampling in ACS and relative standard errors for estimates
- ☐ Evaluate several possible alternative ACS sampling scenarios
- ☐ Several exhibits from the 1999-2001 ACS-Census 2000 Comparison Study
- ☐ Three MPO case studies to measure impact of ACS on CTPP Part 3

Errors in Sample Estimates

Cities with Worst Commute	
New York City, NY	38.3
Chicago, IL	33.2
Newark, NJ	31.5
Riverside, CA	31.2
Philadelphia, PA	29.4
Baltimore, MD	29.0
Los Angeles, CA	29.0
Miami, FL	29.0
San Francisco, CA	28.5
Washington, DC	28.4

Travel Time Standard Errors
(ACS Data Profiles 2003)

Cities with Worst Commute	
New York City, NY	38.3 **±0.30**
Chicago, IL	33.2 **±0.45**
Newark, NJ	31.5 **±2.85**
Riverside, CA	31.2 **±2.09**
Philadelphia, PA	29.4 **±0.55**
Baltimore, MD	29.0 **±1.33**
Los Angeles, CA	29.0 **±0.45**
Miami, FL	29.0 **±1.12**
San Francisco, CA	28.5 **±0.30**
Washington, DC	28.4 **±0.42**

How Different are Commute Times?

- ☐ 90 city pairings of average commute times
- ☐ Calculate standard errors for differences
- ☐ Calculate 90% confidence interval for rejecting null hypothesis that times are the same
- ☐ Compare differences in commute times against 90% confidence interval
- ☐ Only 15 of 90 pairings are significantly different
  - ■ New York vs. all other cities
  - ■ Chicago vs. Philadelphia-Washington, DC

ACS Data Collection
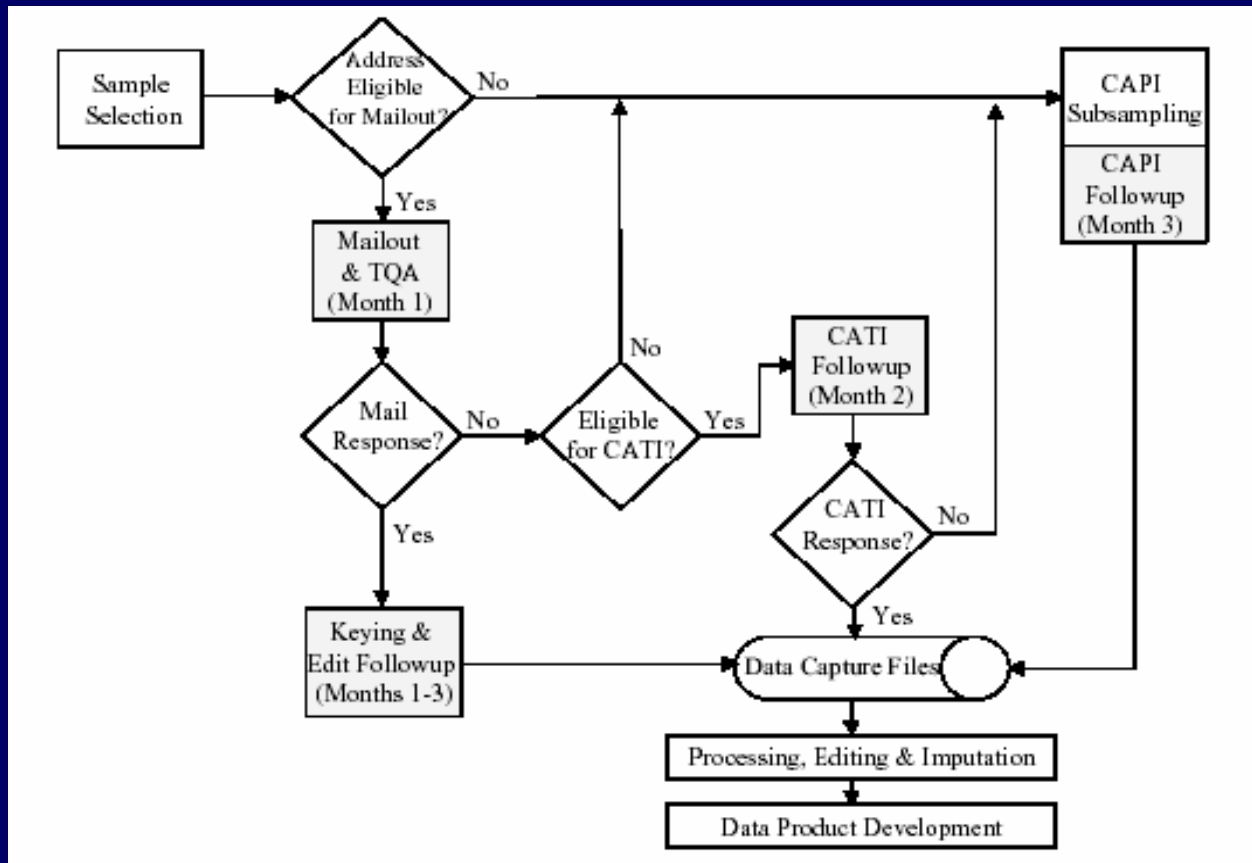
Housing Unit Samples

Area Type	Census 2000	ACS
Blocks in Smallest Gov. Units <200 Occupied Housing Units 200-800 Occupied Housing Units	1 in 2	1 in 10 1 in 13.3
Blocks in Small Gov. Units 800-1200 Occupied Housing Units	1 in 4	1 in 26.7
Blocks in Large Tracts >2000 Occupied Housing Units	1 in 8	1 in 53.3
All Other Blocks	1 in 6	1 in 40

Computer Assisted Personal Interviewing Sample

| CAPI Eligible H.U.	Rate
Without Mailing Address	2 in 3
In Tracts with Response Rate	
Less Than 35 Percent	1 in 2
Between 35 and 50 Percent	2 in 5
Rate Between 50 and 60 Percent	1 in 3
Greater Than 60 Percent (Initial Mail-Out Reduced by 8%)	1 in 3

ACS Survey Responses

Type	Percent
Mail-back Questionnaires	65%-70%
Computer Assisted Telephone Interview	10%-15%
Computer Assisted Personal Interview	20%-25%

ACS Data Availability

Summary Level Population	Estimates	Available
>65,000	Single Year	2006 on
20,000-65,000	3-Year Average	2008 on
Tracts, Block Groups	5-Year Average	2010 on
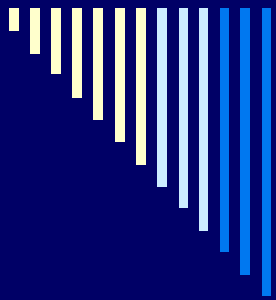
ACS and Census 2000 Estimates

Estimate Source	Approximate Sample Size	Relative Standard Error
Census 2000 Long-Form	1 in 6	1.00
ACS 1-Year	1 in 40	2.79
ACS 3-Year Average	1 in 13.3	1.61
ACS 5-Year Average	1 in 8	1.25

Paper's Evaluation of ACS Standard Errors

- ☐ Based on distributed questionnaires not completed interviews - increase by 10%-15%
- ☐ Ignores weighting of estimates to equal control totals
  - ■ Population and housing unit estimates in Census 2000 areas with 200 or more completed questionnaires have no error
  - ■ Fewer areas in ACS would be similarly weighted
- ☐ Adjustments to estimates to reconcile large and small area estimates

Alternative ACS Sampling Scenarios

- ☐ Restricted funding
  - ■ 50% reduction
  - ■ 25% reduction
  - ■ No Computer Assisted Personal Interviewing
  - ■ No Computer Assisted Telephone Interviewing or Computer Assisted Person Interviewing
- ☐ Missing year of data collection
- ☐ Voluntary participation
- ☐ Other
  - ■ GAO proposal
  - ■ 7-year averaging

GAO Proposal

- Increase sampling rate to 1 in 25 in year before, during, and after decennial census

- Small area 3-year average estimate nearly equal to 5-year average estimate within one year of decennial census

- Possibly lower sample rate during seven off years

Alternative Scenarios: Standard Error Evaluation

| Scenario	Relative Standard Errors
Restricted Funding	12%-42% Worse
Missing Year	23% Worse (3-Year Average) 12% Worse (5-Year Average)
Voluntary	9%-12% Worse
GAO Proposal	1% Worse (3-Year vs. 5-Year Average)
7-Year Average	15% Better (7-Year vs. 5-Year Ave.) 6% Worse (7-Year vs. Census 2000)

ACS Test Site and Census 2000 Comparisons

☐ Thirty-six ACS test sites in 31 counties during 1999-2001

☐ Extensive comparisons between test site results and Census 2000 completed by Census Bureau

☐ Different sample rates from fully implemented ACS, also some variation by test site

☐ Following comparisons reflect adjustment of 1999-2001 ACS sample sizes to roughly equal full ACS

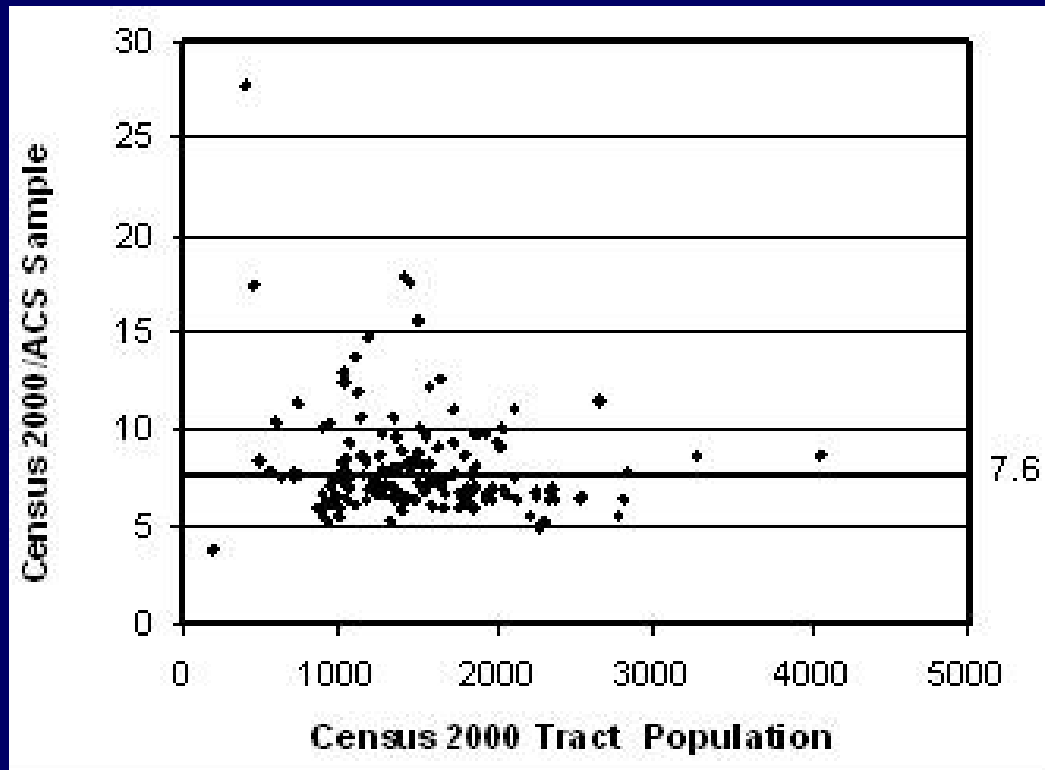# 90% C.I. for Tract Estimates of Public Transportation Commuters
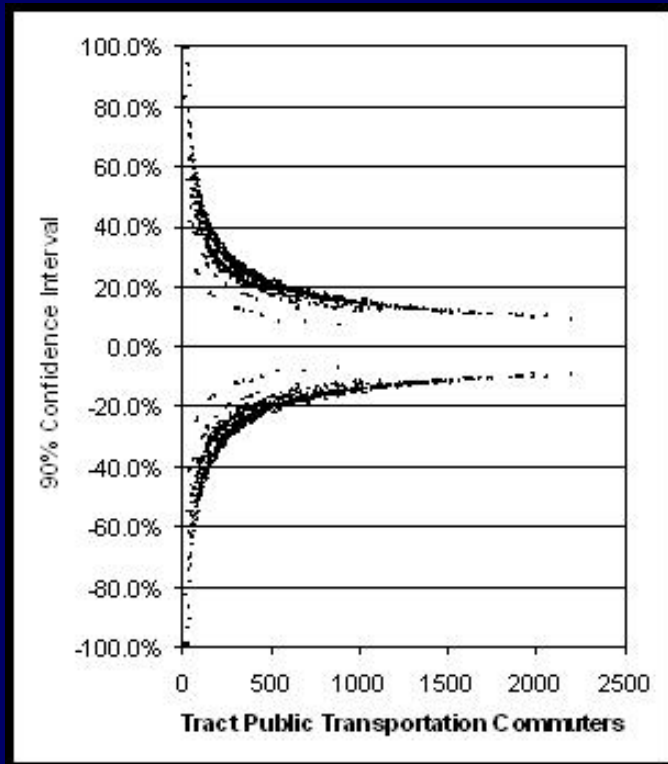
Census 2000 Long-Form   1999-2001 ACS Estimates
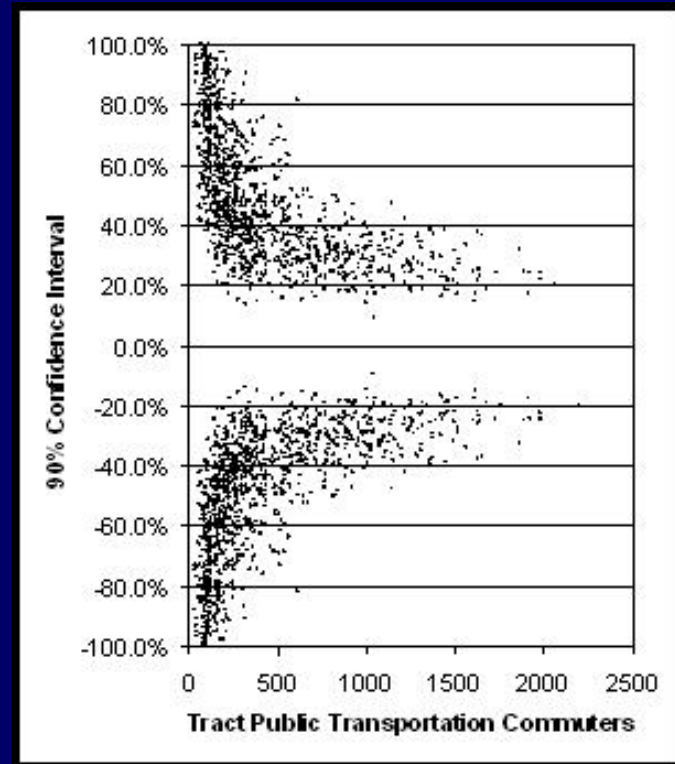
Years of ACS Data Needed to Match Census 2000 Sample

Lake County Illinois Test Site

Effect of ACS on CTPP Part 3

☐ Three Illinois MPO case studies

- Chicago Area Transportation Study

- TriCounty Regional Planning Commission

- Kankakee Area Transportation Study

☐ Suppression of data

- Five tables (Tables 3-03 through 3-07) suppressed if workers $\leq$ 3

- Zeroed values with suppression flag

☐ Simulate effect of ACS by sampling Census 2000 CTPP

Simulation Approach

1. Read 2000 CTPP interchange
2. Rounding: determine high and low values
3. With HU sample rate, estimate upper and lower bounds on sampled workers
4. Randomly determine workers in interchange
5. For each worker in interchange, randomly determine if in sub-sample of 2000 CTPP matching ACS sample (0.75 probability)
6. Determine if reduced sample changes suppression (new workers ≤ 3)
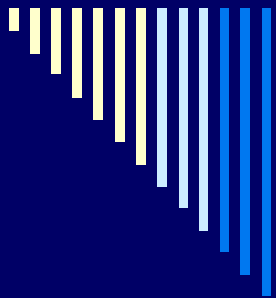
CATS

1843 Tracts

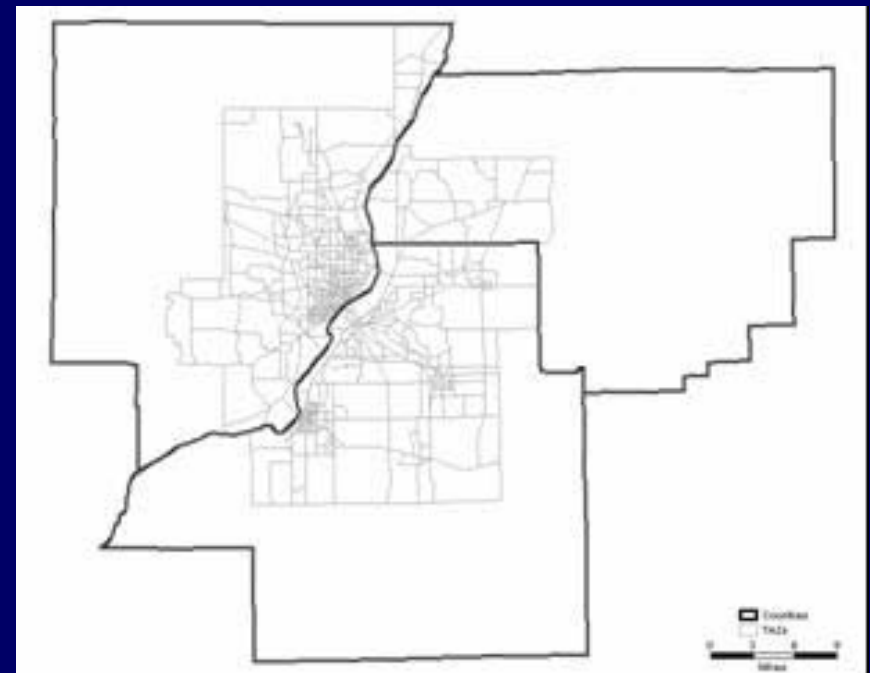### 6167 TAZs

Census Data for Transportation Planning
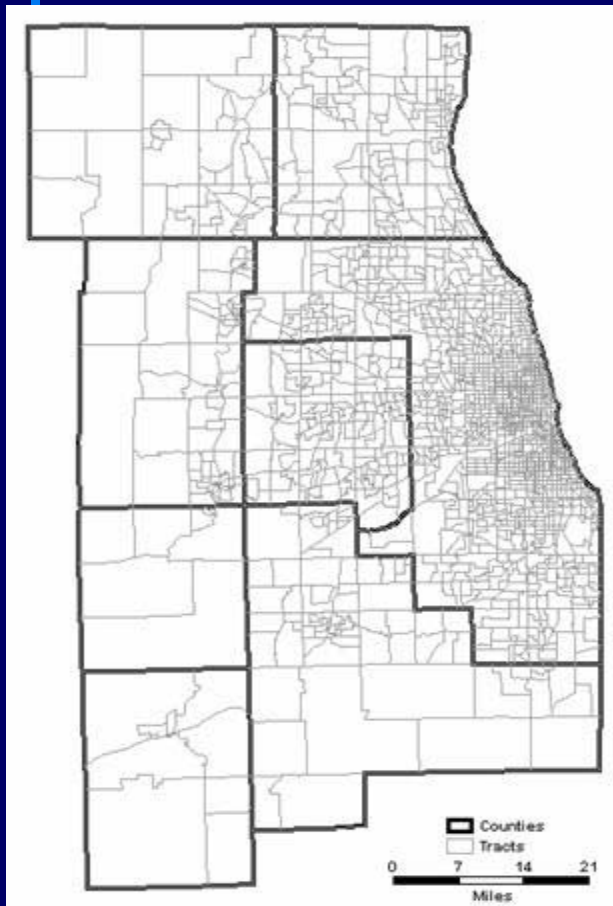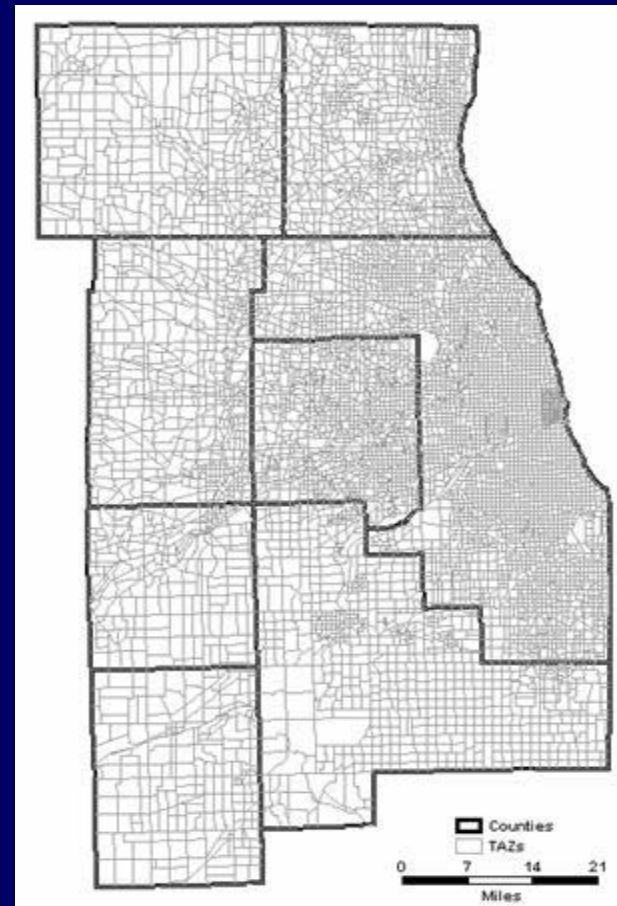
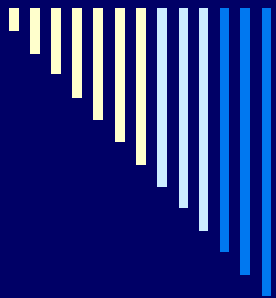# TriCounty RPC

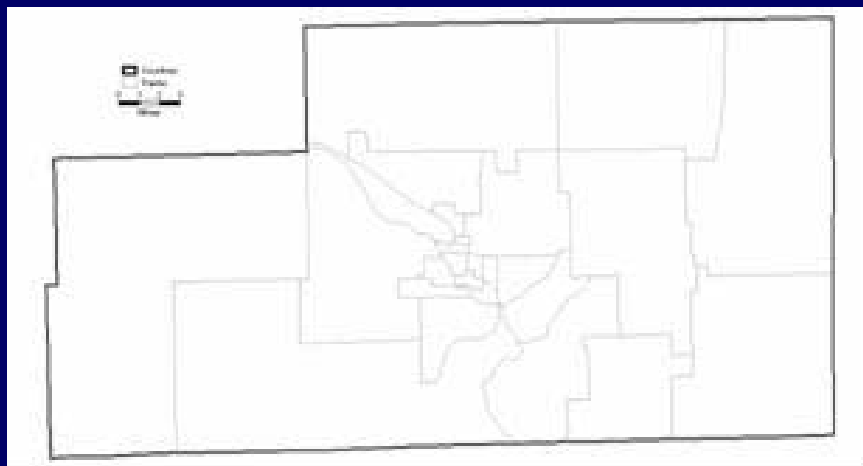87 Tracts   526 TAZs
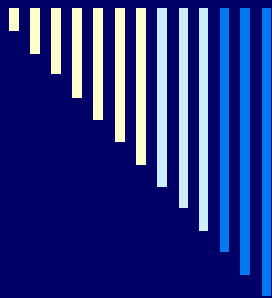
KATS

26 Tracts   195 TAZs
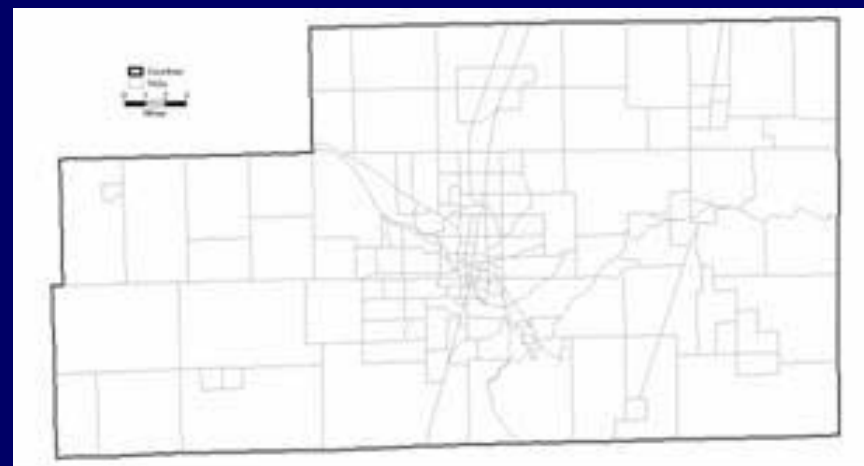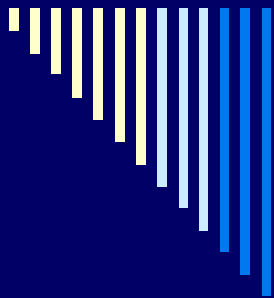
Journey-to-Work Interchanges

Internal Interchanges	-----2000 CTPP-----			--Simulated ACS--		
	CATS	Tri-County	KATS	CATS	Tri-County	KATS
Tract to Tract						
CTPP/Max. Records	6.5%	56.7%	85.2%	5.4%	51.3%	80.9%
Unsuppressed/ CTPP Records	18.3%	50.3%	72.0%	16.4%	46.1%	67.6%
TAZ to TAZ						
CTPP/Max. Records	0.9%	5.3%	9.7%	0.7%	4.3%	8.2%
Unsuppressed/ CTPP Records	6.7%	11.9%	16.4%	5.7%	10.1%	13.4%

Journey-to-Work Interchanges Weighted by 2000 CTPP Workers

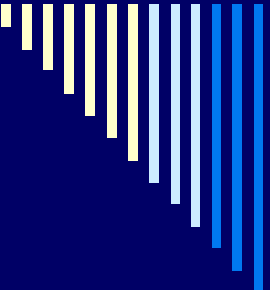| Internal Interchanges	-----2000 CTPP-----			--Simulated ACS--		
	CATS	Tri-County	KATS	CATS	Tri-County	KATS
Tract to Tract						
CTPP/Max. Workers	98.1%	99.5%	100.0%	89.9%	97.6%	99.3%
Unsuppressed/ CTPP Workers	55.1%	88.3%	96.3%	52.0%	84.9%	94.4%
TAZ to TAZ						
CTPP/Max. Workers	96.6%	97.1%	99.9%	82.6%	86.1%	89.2%
Unsuppressed/ CTPP Workers	24.7%	37.5%	41.3%	22.8%	34.6%	35.7%

Conclusions: Standard Errors in ACS Small Area Estimates

☐ Estimates of increased standard errors due to sample size alone are conservative and may not be most important contributor

☐ Most important impacts:

- Small proportions of larger populations (non-motorized, transit, work at home)

- Tails of distributions (vehicle ownership, workers in households)

- Transportation studies involving subpopulations (environmental justice, specialized transit)

Conclusions: Tracking Regional Socioeconomic Changes

- ☐ Few differences between annual estimates for small areas will be statistically significant

- ☐ Difference between two estimates has larger standard error than single year estimate

- ☐ Generally can only track changes for some large area estimates

Conclusions: ACS Methodology and Sampling

☐ Greatly depends on mail-back of questionnaires

☐ CAPI is sample of sample (more housing units eligible for CAPI reduces overall sample)

☐ Mail-back participation may vary

- ■ Between decennial census
- ■ Over time

Conclusions: Alternative Sampling Procedures

- ☐ Major impacts from reduced samples due to possible interruptions and cost-cutting
- ☐ GAO proposal:
  - ■ ACS would benefit from publicity surrounding decennial census
  - ■ ACS estimates close to decennial census and can use for ACS control totals
  - ■ Three-year vs. five-year average
  - ■ Variable sample rate and staffing requirements during 10 year cycle

Conclusions: CTPP Part 3

- Important issue is suppression, not ACS sample
    - TAZ level tables with suppression are of little value for most MPOs
    - For larger MPOs, tract level tables with suppression appear to be of limited use
- Unsuppressed Part 3 tables are modestly affected by ACS sample, but still should be useful

Implications for MPO and State DOT Planners

- Little past awareness of errors in census estimates
- Research on how errors are transmitted through model calibration and validation
- Discontinuities in estimates
  - Current vs. usual residence
  - Procedures for surveying large households
  - Other methodological differences?
- Agency staffing

Final Questions?

- ☐ Do MPO and state DOT planners actually want annual small area long-form estimates?
  - ■ Plan updates and model calibration/validation driven by multiyear planning cycles
  - ■ Same base year often used for several planning cycles
  - ■ Many annual releases will go unused
- ☐ Larger area estimates more useful for tracking changes and work program planning