Outline

• Motivation
• Testing Equivalence in General
• Application to Certifying APC Data for NTD Reporting
• Comments on Federal Register Notice
• Contact Info
National Transit Database (NTD)

- U.S. public transit agencies that benefit from Federal transit grant programs report their data to the NTD on:
  - Resource used
  - Service provided
  - Service consumed

- Service consumed includes:
  - Unlinked passenger trips (UPT), i.e., boardings
  - Passenger miles traveled (PMT)

- Service consumed: 100% count or estimated
  - Annual UPT: increasingly a 100% count
  - Annual PMT typically estimated

- Minimum 10% precision & 95% confidence

Load-Based PMT for Each Trip

1. Collect detailed field data:
   - Boardings at each stop
   - Alightings at each stop

2. Balance trip-total boardings and alightings

3. Calculate loads between stops

4. Calculate PMT between stops

5. Sum PMT between stops across all stops
Manual Method for Field Data

• Human ride checks or in-vehicle video
• Advantages
  – Can ALSO observe the following
    • Passengers from the previous trip at 1st stop
    • Passenger loading between each pair of stops
    • Passengers continuing to the next trip at last stop
  – Minimum measurement error when done properly
• Disadvantages
  – High labor cost
  – Limited coverage (i.e., a small sample of trips each year)
• Estimate annual PMT through random sampling

Automated Method

• Automatic passenger counter (APC)
• Advantages
  – Low labor cost
  – Wide coverage (i.e., all trips with APC)
• Disadvantages
  – Cannot also observe the following
    • Passengers from the previous trip at 1st stop
    • Passenger loading between each pair of stops
    • Passengers continuing to the next trip at last stop
  – Potentially high measurement error in PMT
  – Worst with interlined and loop services
  – Wide variation in accuracy across applications
• Estimate annual PMT through
  – Using all valid APC data, or
  – Using random sampling
Components of Errors in Annual PMT

- Manual Data
  - Zero measurement error (at least assumed so)
  - Pure random error from sampling
  - Sampling plans are based on 10% precision
- APC Data
  - Using all valid APC data
    - Minor random error due to large sample size
    - Mainly measurement error
  - Using random sampling
    - Large random error due to small sample size
    - Measurement error can be significant
    - Sampling plans must account for measurement error: sampling precision = 10% - measurement error

Motivation

Certification of APC Data

- NTD requires APC data be certified for when used for reporting
- NTD requires proof of equivalence, i.e., APC and manual data differ within an acceptable difference
- NTD requires using paired observations, i.e., collecting both APC and manual data from the same sample of vehicle trips
Potential Methods for Certification

• Non-Statistical Testing
  – Test if APC and manual data differ within an acceptable difference **numerically**
  – Referred to as *Numerical Testing* here

• Statistical Testing
  – Traditional t-test:
    • referred to as *Difference Testing*
  – A well established test in medicine
    • test if APC and manual data differ within an acceptable difference **statistically**
    • referred to as *Equivalence Testing*

**Basics of Statistical Testing**

• A pair of hypotheses:
  Null \( (H_0) \) vs Alternative \( (H_1) \)

• Objective is to prove \( H_1 \) through rejecting \( H_0 \)

• Not rejecting \( H_0 \) means not enough evidence to reject it

• Not rejecting \( H_0 \) does not prove it statistically

• \( H_1 \) must be based on what one wants to prove:
  – To detect difference requires equality in \( H_0 \) and inequality in \( H_1 \) (e.g., traditional t-test)
  – To prove “equivalence” requires “non equivalence” in \( H_0 \) and “equivalence” in \( H_1 \)
Basic Terms and Notations

- \( n \) paired observations: \( y_{ai} \) (automated) and \( y_{mi} \) (manual)
- Mean for manual data: \( \bar{y}_a = \frac{\sum y_{ai}}{n} \)
- Difference for each pair: \( d_i = y_{ai} - y_{mi} \)
- Mean difference: \( \bar{d} = \frac{\sum d_i}{n} \) (or \( \bar{d} \) in percent terms)
- Standard deviation of difference: \( s_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1} \)
- Standard error of mean difference: \( s_{\bar{d}} = \frac{s_d}{\sqrt{n}} \)
- Significance level for statistical testing: \( \alpha \) (e.g., 5%)
- Population mean difference: \( \mu_d \) (the true value)
- Acceptance criterion for Numerical Testing: \( \delta \)
- Acceptance criterion for Equivalence Testing: \( \theta \)

Testing Equivalence

Numerical Testing

- Pre-select a specific value for \( \delta \) (i.e., acceptance criterion)
- Compare \( \bar{d} \) to \( \delta \) numerically:
  - Accept equivalence of APC data if \( |\bar{d}| \leq \delta \)
- Simple
- Ignore variability in \( \bar{d} \) across samples
- Consider precision but not significance
- High probability of accepting equivalence by chance
- Do not ensure 10% precision at 95% confidence
Difference Testing

- Hypotheses: H₀: μ₀ = 0; H₁: μ₀ ≠ 0
  
- Designed to detect difference through rejecting equality

- Failing to reject H₀ if
  - \( t = \frac{|d|}{s_d} \leq z_{1-\alpha/2} \)
  - Or \((1-\alpha)\%\) confidence interval includes 0

- Failing to reject equality is mistaken with proving equality (equivalence)

- Poor precision or small sample favors this mistaken equivalence

Equivalence Testing

- Pre-select acceptance criterion θ: mean difference not exceeding \( \theta \) is acceptable
  - Based on prior knowledge and application at hand

- Hypotheses: H₀: |μ₀| < -θ or μ₀ > θ; H₁: |μ₀| < θ

- Designed to prove equivalence through rejecting not being equivalent

- Use well established Two One-Sided Test (TOST)

- Reject H₀ (i.e., accept equivalence at \( \alpha \)) if
  - \( t_1 = \frac{d-(-\theta)}{s_d} > z_{1-\alpha} \) and \( t_2 = \frac{d-\theta}{s_d} < -z_{1-\alpha} \)
  - Or \((1-2\alpha)\%\) confidence interval falls within \((-\theta, \theta)\)

- Avoid issues with Numerical or Difference Testing
Comparison via Confidence Intervals

Equivalent?

<table>
<thead>
<tr>
<th>Difference Testing (includes 0)</th>
<th>Equivalence Testing ($\pm \theta$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Testing Equivalence

Comparison via Acceptance Zones

Testing Equivalence
Pre-Selection of Acceptance Criterion for Equivalence Testing

- Highly encourage using all valid APC data:
  - A reasonable value for $\theta = 9\%$, allowing a small amount of up to 1% random error due to <100% APC coverage and invalid APC data.

- If using random sampling for estimation:
  - Pre-select $\theta = 6\%$, leaving adequate room for minimum sampling precision
  - More difficult for APC data to be certified
  - Minimum sampling precision = 10% - actual $|\bar{d}|$
    - If actual $|\bar{d}| = 5\%$, sampling precision must be increased to 5%.
      As a result, annual sample size would quadruple over what 10% sampling precision would require.

Application to Certifying APC Data
Pre-Selection of Acceptance Criterion for Numerical Testing

- NTD has proposed acceptance criterion $\delta = 5\%$
- The 10% precision and 95% confidence requirements are statistical in nature but this test is non-statistical
- Passing test does not ensure meeting the requirements
- Highly encourage using all valid APC data
  - Consider cap standard error of mean difference at 4.5%
- If using random sampling for estimation:
  - Minimum sampling precision for sampling plans MUST be increased to 10% minus actual $|\bar{d}|$ in percent terms
    - If actual $|\bar{d}| = 3\%$, sampling precision must be increased to 7%. As a result, annual sample size would double over what 10% sampling precision would require.
  - Consider cap standard error of mean difference at 3%

Application to Certifying APC Data

Using All Valid APC Data: Zone based

Application to Certifying APC Data
# Using All Valid APC Data: Confidence Interval Based

### Difference Testing

<table>
<thead>
<tr>
<th>Agency</th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>H₀ Not Rejected?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.23%</td>
<td>1.05%</td>
<td>1.1%</td>
<td>5.3%</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>5.16%</td>
<td>1.52%</td>
<td>2.1%</td>
<td>8.2%</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>-0.12%</td>
<td>0.95%</td>
<td>-2.0%</td>
<td>1.8%</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>1.06%</td>
<td>2.06%</td>
<td>-3.0%</td>
<td>5.1%</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>4.74%</td>
<td>1.13%</td>
<td>2.5%</td>
<td>7.0%</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>12.42%</td>
<td>1.83%</td>
<td>8.8%</td>
<td>16.1%</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>1.10%</td>
<td>2.14%</td>
<td>-3.1%</td>
<td>5.3%</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>-0.36%</td>
<td>1.35%</td>
<td>-3.0%</td>
<td>2.3%</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>-0.15%</td>
<td>1.08%</td>
<td>-2.3%</td>
<td>2.0%</td>
<td>Yes</td>
</tr>
</tbody>
</table>

# Using All Valid APC Data: Confidence Interval Based (θ = 9%)  

### Equivalence Testing

<table>
<thead>
<tr>
<th>Agency</th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Equivalent?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.23%</td>
<td>1.05%</td>
<td>1.5%</td>
<td>5.0%</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>5.16%</td>
<td>1.52%</td>
<td>2.6%</td>
<td>7.7%</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>-0.12%</td>
<td>0.95%</td>
<td>-1.7%</td>
<td>1.5%</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>1.06%</td>
<td>2.06%</td>
<td>-2.4%</td>
<td>4.5%</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>4.74%</td>
<td>1.13%</td>
<td>2.9%</td>
<td>6.6%</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>12.42%</td>
<td>1.83%</td>
<td>9.4%</td>
<td>15.5%</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>1.10%</td>
<td>2.14%</td>
<td>-2.5%</td>
<td>4.7%</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>-0.36%</td>
<td>1.35%</td>
<td>-2.6%</td>
<td>1.9%</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>-0.15%</td>
<td>1.08%</td>
<td>-1.9%</td>
<td>1.6%</td>
<td>Yes</td>
</tr>
</tbody>
</table>
### Using Random Sampling: Acceptance criterion $\theta = 6\%$

<table>
<thead>
<tr>
<th>Agency</th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Equivalent?</th>
<th>Sampling Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.23%</td>
<td>1.05%</td>
<td>1.5%</td>
<td>5.0%</td>
<td>Yes</td>
<td>6.77%</td>
</tr>
<tr>
<td>2</td>
<td>5.16%</td>
<td>1.52%</td>
<td>2.6%</td>
<td>7.7%</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>3</td>
<td>-0.12%</td>
<td>0.95%</td>
<td>-1.7%</td>
<td>1.5%</td>
<td>Yes</td>
<td>9.88%</td>
</tr>
<tr>
<td>4</td>
<td>1.06%</td>
<td>2.06%</td>
<td>-2.4%</td>
<td>4.5%</td>
<td>Yes</td>
<td>8.94%</td>
</tr>
<tr>
<td>5</td>
<td>4.74%</td>
<td>1.13%</td>
<td>2.9%</td>
<td>6.6%</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>6</td>
<td>12.42%</td>
<td>1.83%</td>
<td>9.4%</td>
<td>15.5%</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>7</td>
<td>1.10%</td>
<td>2.14%</td>
<td>-2.5%</td>
<td>4.7%</td>
<td>Yes</td>
<td>8.90%</td>
</tr>
<tr>
<td>8</td>
<td>-0.36%</td>
<td>1.35%</td>
<td>-2.6%</td>
<td>1.9%</td>
<td>Yes</td>
<td>9.64%</td>
</tr>
<tr>
<td>9</td>
<td>-0.15%</td>
<td>1.08%</td>
<td>-1.9%</td>
<td>1.6%</td>
<td>Yes</td>
<td>9.85%</td>
</tr>
</tbody>
</table>

### Application to Certifying APC Data

### Using Random Sampling: Acceptance criterion $\delta = 5\%$

<table>
<thead>
<tr>
<th>Agency</th>
<th>Mean Difference</th>
<th>Equivalent?</th>
<th>Sampling Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.23%</td>
<td>Yes</td>
<td>6.77%</td>
</tr>
<tr>
<td>2</td>
<td>5.16%</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>3</td>
<td>-0.12%</td>
<td>Yes</td>
<td>9.88%</td>
</tr>
<tr>
<td>4</td>
<td>1.06%</td>
<td>Yes</td>
<td>8.94%</td>
</tr>
<tr>
<td>5</td>
<td>4.74%</td>
<td>Yes</td>
<td>5.26%</td>
</tr>
<tr>
<td>6</td>
<td>12.42%</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>7</td>
<td>1.10%</td>
<td>Yes</td>
<td>8.90%</td>
</tr>
<tr>
<td>8</td>
<td>-0.36%</td>
<td>Yes</td>
<td>9.64%</td>
</tr>
<tr>
<td>9</td>
<td>-0.15%</td>
<td>Yes</td>
<td>9.85%</td>
</tr>
</tbody>
</table>
• Newly proposed APC rules do not ensure 10% precision & 95% confidence levels
• To protect the integrity of these statistical requirements:
  – Require using *Equivalence Testing*
  – Encourage using all valid APC data over random sampling for estimation
  – When using all valid APC data, acceptance criterion $\theta = 9\%$
  – When using random sampling:
    ▪ Use a stricter acceptance criterion $\theta = 6\%$
    ▪ Require agencies to account for measurement error of APC data in sampling plans
    ▪ sampling precision = 10% - absolute value of actual mean difference

Comments on FR Notice

• Newly proposed *Numerical Testing* does not ensure 10% precision & 95% confidence levels
• If stay with *Numerical Testing*, improve it with the following additional requirements:
  – Encourage using all valid APC data over random sampling for estimation
  – When using all valid APC data, cap mean difference at $= 4.5\%$
  – When using random sampling:
    ▪ Cap mean difference at $= 3\%$
    ▪ Require agencies to account for measurement error of APC data in sampling plans
    ▪ sampling precision = 10% - absolute value of actual mean difference

Comments on FR Notice
Having an Excel template for agency to use

- Agencies enter paired data and method for estimation
- Results from the template:
  - Certification outcome
  - If failing, indicate reasons for the failure:
    - Mean difference too big
    - Differences vary too much (i.e., standard error of mean difference too big)
    - Both are too big
  - Produce minimum sampling precision if using random sampling

For additional questions or comments, contact me:

- E-mail: xchu@cutr.usf.edu
- Phone: 813-974-9831