

Discussion-piece #17  
Methods for Collection of Data on Transit Riders  
Federal Transit Administration  
June 6, 2006

1. Motivations. Decisionmaking on major transit projects – that have price tags in the hundreds of millions, often billions, of dollars – deserves reliable information on the costs, benefits, and other impacts of those projects. Prominent among the benefits are the service improvements and resulting mobility increases for existing transit riders and for travelers attracted to transit by those improvements. Informed decisionmaking requires an understanding of the current role of transit in a metropolitan area and a reliable forecast of the proposed project’s contribution to mobility as part of transit’s future role in the area. Both of these requirements usually mean that useful data on the characteristics and travel patterns of current transit riders is essential. Consequently, FTA has proposed a requirement in the upcoming New Starts regulation implementing SAFETEA-LU that the predicted transportation benefits for projects proposed for entry into preliminary engineering must be grounded in a recent survey of riders of the existing transit system.
2. Purpose of supporting guidance. To fulfill the purposes of the proposed regulatory requirement, the data on current riders must be sufficient to the tasks of (1) illuminating the various roles that transit plays in the area and (2) ensuring that the travel forecasting models grasp those roles. Consequently, if the eventual regulation includes the data requirement, the supporting guidance will define the conditions that establish sufficiency. The guidance would likely cover four dimensions to ensure a sufficient data collection effort:
  - Relevant markets and transit services;
  - Sample size and distribution;
  - Sample control and expansion; and
  - Necessary data items.

The guidance would also offer some general technical observations on the careful design and efficient implementation of rider surveys. As with other FTA guidance, the discussions are intended to represent good practice as it applies generally. Because no guidance can anticipate the specific conditions in all situations, FTA would make the usual recommendation that potential project sponsors discuss the implications of the guidance in their context and reach agreement on any appropriate departures from the general principles it endorses.

3. Relevant markets and transit services. The dual needs to understand the major roles transit plays in a particular area and to ensure that the local travel models understand those roles typically require the collection of data from all existing transit travel markets and transit services. In unusual circumstances, a more limited scope might be sufficient. Forecasts for alternative long-distance services and facilities for outlying commuter markets, for example, might be sufficiently supported by data on current ridership patterns in those markets, particularly if the local bus system serves primarily to provide basic mobility to transit dependent travelers in the regional core. Similarly, forecasts for downtown circulation

facilities might be best informed by data on current transit riders making lunchtime and other circulation trips within downtown rather than a regional survey. With either example, however, a generally serviceable regional model set is likely to be important because no transit project exists in perfect isolation. Consequently, a systemwide survey may still be desirable unless good evidence already exists on the usefulness of the model set in making transit forecasts.

4. Sample size and distribution. The need to understand travel patterns may lead to sample designs that are quite different from the traditional design oriented towards descriptive statistics on the riders of individual routes. The traditional focus on individual routes means that higher sampling rates are needed on lower volumes routes – leading to an allocation of samples that favors marginal markets. Travel patterns are characterized by important origin-destination (really production-attraction) flows, stratified at least by trip purpose (home-based-work, home-based-non-work, and non-home-based, perhaps) and the rider's income or auto-ownership. Where the existing transit system includes guideway facilities, separate tabulations of the travel patterns using those facilities are extremely desirable. An efficient sample for New Starts purposes may therefore be one allocated in direct proportion to the size of travel markets – and in direct proportion to volumes on individual transit routes since the travel markets themselves are usually not observable prior to survey data collection. The goal is to have a sufficient number of observations to support the number reported in any important cell in the expanded tabulations. An estimate of 1,000 transit trips in a cell is probably not helpful if it is derived from only two underlying observations with sample-expansion weights that average 500!
5. Sample control and expansion. Rider surveys have traditionally depended heavily on counts of survey forms handed out (on specific bus trips or train trips, or at specific stations, or at other selected locations). Often, these counts are the primary (and perhaps the only) control used in expansion of the sample. The advantage of this strategy is that it requires little additional effort (and therefore cost) beyond the effort needed anyway to distribute the survey forms. The principal difficulties with this strategy are that it ignores the effects of sampling error and implicitly assumes a uniform response rate among the various travel markets using the transit system. For example, survey forms distributed on a sampled train trip might be controlled only against the total number of forms distributed on that trip. The resulting expansion factor (forms-distributed divided by useable forms returned) would then apply to all useable forms returned from that trip. But, if the access-mode distribution at each station is of interest (and it is), then the survey will be able to produce that estimate based only on the rail-line-wide expansion factors. Sampling error alone is likely to produce estimates well off the mark. Even more problematic are the effects of non-response bias. The likelihood that a sampled rider will return a useful form depends on the rider's literacy, education level, success in finding a seat on the train, duration of their time on the train, and possibly other influences. Consequently, upper-middle-class riders boarding at outlying stations making long seated trips are likely to respond at higher rates than lower-income riders boarding at close-in stations and standing on the train for a few stops before their destination station. Expansion factors derived from and applied to all riders boarding an individual train will inevitably over-estimate: the number of boardings at outlying stations, the number of cars parked in park-ride lots, the average trip-time and trip-distance on the rail line, and the average income of riders – and underestimate other ridership characteristics.

The remedy for these problems is the collection of a rich set of count information that can be used to control sample expansion. Counts of distributed surveys should be as un-aggregated as possible; counts of forms distributed by station, for example, would help to address both sampling error and non-response bias. Supplemental counts are crucial: the number of cars parked in each park-ride lot, for example, would substantially address the impact of non-response biases on access-mode distributions.

6. Necessary data items. Table 1 suggests a starting-point set of necessary and optional data items for surveys. These items are draft and subject to revision.

Table 1. Data Items from Rider Surveys	
Necessary data items	Optional data items
- Trip origin and destination (O&D)	- First boarding (on) location
- Activity purposes at the origin & destination	- Last alighting (off) location
- Trip access and egress modes (O&D)	- Fare payment method
- Park/ride location	- Frequency of transit use
- All transit lines used for the trip	- Other household characteristics
- Driver's license (or ability to drive)	- Other personal characteristics
- Household vehicles	- Satisfaction with service
- Household workers	
- Household income	
<b>----- subject to revision -----</b>	

It is important to note the item on all transit lines used for the trip, plus the access and egress modes from and to the actual origin and destination (rather than the access mode to the boarding point for the current line and the egress mode from the alighting point for this line).

7. Careful design and implementation. General observations on useful data collection are likely to deal with (1) pilot testing of the survey form, (2) avoidance of round-trip reporting by respondents, (3) the importance of a well developed sample-expansion plan, (4) the crucial role of a count program in conjunction with the survey, (5) special strategies for special markets, (6) recruitment, training, and retention of competent survey crews, and probably other topics.
8. FTA technical support. FTA is making arrangements to provide contractor support to project sponsors and other agencies who will be conducting rider surveys in preparation for New Starts forecasting. This support can address the development of survey forms, the preparation of sampling plans, calculation of survey-expansion weights, quality control, and other aspects of the design of the survey.