

MIT Open Access Articles

Using Smart Card Fare Payment Data To Analyze Multi-Modal Public Transport Journeys in London

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Seaborn, Catherine, John Attanucci, and Nigel H. M. Wilson. "Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data." *Transportation Research Record: Journal of the Transportation Research Board* 2121.-1 (2009): 55-62.

As Published: <http://dx.doi.org/10.3141/2121-06>

Publisher: Transportation Research Board of the National Academies

Persistent URL: <http://hdl.handle.net/1721.1/69577>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



**Using Smart Card Fare Payment Data
To Analyze Multi-Modal Public Transport Journeys in London**

By

Catherine Seaborn
Halcrow Group
Vineyard House, 44 Brook Green
Hammersmith, London W6 7BY
United Kingdom
E-mail: cseaborn@alum.mit.edu

John Attanucci (corresponding author)
Research Associate
Center for Transportation and Logistics
Massachusetts Institute of Technology
77 Massachusetts Avenue, Room 1-274
Cambridge, MA 02139
Phone: (617) 253-7022
E-mail: jattan@mit.edu

Nigel H. M. Wilson
Professor
Department of Civil and Environmental Engineering
Massachusetts Institute of Technology
77 Massachusetts Avenue, Room 1-238
Cambridge, MA 02139
Phone: (617) 253-5046
E-mail: nhmw@mit.edu

Word count: 7,400 words (including abstract, tables and figures, acknowledgements and references)

Date submitted: August 1, 2008

Revised: November 15, 2008

Revised for publication: March 15, 2009

ABSTRACT

This paper contributes to an emerging literature on the application of smart card fare payment data to public transportation planning. The research objective is to identify and assess complete, multi-modal journeys using “Oyster” smart card fare payment data in London. Three transfer combinations: bus-to-Underground, Underground-to-bus, and bus-to-bus are considered in order to formulate recommendations for maximum elapsed time thresholds to identify transfers between journey stages for each passenger on the London network. Recommended elapsed time thresholds for identifying transfers are: 20 minutes for Underground-to-bus, 35 minutes for bus-to-Underground, and 45 minutes for bus-to-bus, but a range of values that account for variability across the network are assessed. Key findings about bus and Underground travel in London include: an average of 2.3 daily public transportation journeys per passenger, 1.3 journey stages per public transportation journey, and 23 percent of Underground journeys include a transfer to or from a bus. The application of complete journey data to bus network planning is used to illustrate the value of new information that would be available to network planners by using smart card fare payment data.

INTRODUCTION

Transport for London's (TfL) planning guidelines emphasize better system integration across modes (1). In particular, TfL's *Interchange Plan* (2) categorizes 614 interchange facilities into five groups ranging from major central London termini to interchanges of local importance, and then prioritizes them for infrastructure improvement. This prioritization is based primarily on qualitative analysis and the plan states that reliable data on such basic metrics as the number of people transferring at each facility was not available. In addition to key interchange facilities, over 700 intersecting bus routes provide opportunities for on-street transfers in London.

TfL currently has well-developed systems for assessing passenger demand between stops on a bus route and on *parallel* routes, but information on passenger demand between stops on a route and *intersecting* routes or Underground stations, and between ultimate origins and destinations is less well provided for. The current sources of stop-level passenger demand data are a rolling (every six years) origin-destination survey as well as a survey that records boardings, alightings and loads at 400 key bus stops on a two-year cycle. Thus, bus network planners at TfL utilize intermittent surveys and route-level Electronic Ticketing Machine data (a farebox record at point of entry on the bus), as well as experiential knowledge, to evaluate bus routing and service changes.

With the ongoing challenge of developing the bus network to meet the needs of Londoners, smart card data can be used to expand information on passenger demand which might include the following:

- Passenger flows between intersecting routes to provide support for direct links that reduce the need for transfers;
- Transfer volumes from bus routes to an Underground station to show which routes are the most important means of accessing the station and adjust station design and/or bus routing accordingly;
- Comparison of Underground-to-bus transfer times, controlling for scheduled bus frequency at an Underground station, to highlight reliability or crowding problems;
- Evidence of multi-modal journeys (e.g., bus-Underground-bus) to support route redesign such as creating direct bus links that reduce the need to transfer and relieve congestion on the Underground; and
- Identification of repeated daily individual passenger travel on a route to indicate strong reliance on that service.

The goal of this paper is to demonstrate that smart card fare payment data can be of value in improving bus network planning by being representative of passenger demand across the TfL network. Extensive service improvements and an associated 52 percent growth in ridership between 2001 and 2007 are testimony to the high quality of bus services in London; however, this growth has been achieved with only modest advances in the methods and data systems used for network planning (3).

Smart cards, such as the "Oyster" card in London, are owned by individuals and generally record the time and place of every transaction the cardholder makes on the public transportation system, for example a bus boarding or Underground station exit. Several types of analyses can be done with smart card data, including estimating origin-destination matrices, measuring passengers' behavioural reactions to service changes, and evaluating service quality. The key contribution of this paper, however, is to develop a methodology for describing passenger transfer behaviour to, from and within the bus network in London using smart card data to identify travel patterns. The results are compared with survey data on aggregate travel patterns.

The widespread adoption of the smart card fare payment system in London and the potential application of resultant data to bus network planning should be of interest to other public transportation agencies implementing smart card fare payment systems. Smart cards have been adopted by approximately 22 public transportation agencies in Europe and more than 30 cities in Asia but studies of the application of the resultant data to bus network planning are only beginning to emerge (4). A few public transportation agencies are integrating smart card data analysis into their daily operations and planning, for example the Seoul Metro Company (4) and the London Underground at TfL (5). However, smart card data has not been used to study how passengers travel across multiple modes in London,

although some similar analysis has been done in Chicago (6). To the best of our knowledge, this research is the first comprehensive attempt to combine bus and Underground journey stage data derived from smart card fare payment transactions into complete journeys using informed maximum elapsed time assumptions to identify transfers. The availability of complete journey information, albeit approximate, would be an advance in knowledge for network planners in evaluating the costs and benefits of changes to the bus network. Moreover, data derived from Oyster smart cards may be less expensive (data collected for fare payment purposes), more timely (available almost immediately) and more accurate (not subject to survey errors) than data from conventional sources.

PRIOR RESEARCH

Analysis of automated fare collection data is an emerging theme in public transportation literature. Using the Chicago Transit Authority as an example, Utsunomiya, Attanucci and Wilson (7) discuss the potential usage of, and barriers to, increased data availability after smart card implementation in public transportation agencies, concluding that agencies need to tailor their smart card implementation plan to make the most of the increased data availability it offers and that smart card penetration as a fare payment method is the key to its effective use for the analysis of passenger behavior.

Bagchi and White (8) examine three cases of smart card implementation in small bus networks in the United Kingdom. They find that the advantages of smart card data include larger samples than existing data sources and the ability to analyze travel behavior over longer periods, but there are also limitations, including in the case of bus travel in which cards are only validated upon entry to the system (i.e., bus boarding) and the absence of certain types of information such as journey purpose. They conclude that smart card data cannot replace existing survey methods for data collection but may complement them. Additionally, the authors estimate smart card turnover rates and trip rates per card, and infer the proportion of all bus boardings to linked trips (i.e., with bus-bus transfers based on a 30-minute threshold) in each network. In a similar study for a larger city, Hoffman and O'Mahony (9) use a 90-minute threshold to link bus journey stages as recorded by magnetic stripe electronic ticketing technology. The highest rate of transfers in this study occurred between 18 and 28 minutes after boarding the first bus. Okamura, Zhang and Akimasa (10) define a transfer as two journey stages that are provided by different operators and occur within 60 minutes at the same location and go on to analyze transfer wait time at major transit hubs. In sum, a range of transfer time assumptions between 30-90 minutes has been used in previous studies for linking bus journey stages to form complete journeys.

Trépanier et al. (11), Cui (6), Zhao et al. (12) and Chan (13) all demonstrate how to estimate the destination of individual bus or rail passengers and thus develop full origin-destination matrices using smart cards based on two assumptions: (1) a passenger's journey stage destination is the first stop of their following journey stage, and (2) at the end of the day, passengers return to the stop where they first boarded. Further to this work, Chu and Chapleau (14) develop methods for enriching smart card data for transit demand modeling including inferring the arrival time of bus runs at the stop level using schedule constraints and linking journey stages based on both location and time constraints. Thus, they avoid the need to make arbitrary transfer time assumptions, but the methodology is complex and computationally intensive.

LONDON APPLICATION

Smart card penetration is crucial to its effective use for the analysis of travel behavior. In London, 73 percent of all journeys on the TfL network are made using the "Oyster" smart card system (2007) (15). Raw data collected from the card readers on each bus (entry only) and at each Underground gate (both entry and exit) are compiled into a sequenced journeys table by TfL. In this format, each record represents one journey stage by bus, Underground, or another mode accepting Oyster.

A journey stage refers to a component of a complete trip and is bounded by the start or end of a journey or a transfer. "Thus, a journey stage is made by a single mode of transport [and vehicle] within a trip that may comprise several journey stages by different modes [and vehicles]" (4, p. i). For example, a separate journey stage occurs each time a passenger boards a different bus. In 2006, 28 million journey

stages occurred daily in Greater London, which has an area of 1,584 square kilometers and a population of about 7.5 million residents (4, 16). Approximately 37 percent of these journey stages were made using public transportation, including bus or tram (19%), Underground (10%), and rail (8%). Public transportation is used for two-thirds of all weekday, work-related journey stages, demonstrating its critical role in the economy of the city. Journey stages by public transportation grew 18 percent between 2000 and 2005 while travel by private vehicle has declined since 2002 (17). The London public transportation network includes: 275 Underground stations on 12 lines, 8,200 buses serving over 700 routes, light rail, tram and ferry services, and 8 major rail termini (15).

Data

A sequenced journeys table for a typical weekday (Wednesday, November 14, 2007) was supplied by TfL for this research. The sequenced journeys table includes over 8 million records, each representing one journey stage on the TfL network, although Underground journey stage records may actually comprise two or more stages as Underground transfers occur “behind the barrier” and are therefore not recorded by Oyster. The vast majority of journey stages recorded by Oyster are by bus or Underground, but 4% are on other modes including intercity rail, tram and light rail. These records were excluded because of data scarcity, leaving almost 8 million records for bus and Underground, including about 5.1 million bus journey stages.

The attributes of each journey stage that were used for this research are: Day, Card ID (encrypted), Journey Stage Sequence Number, Mode, Start Location, End Location, Start Time, and End Time. A further 2% of records were classified as errors or outliers and removed (18).

This research compares travel patterns derived from smart card data to the London Travel Demand Survey (LTDS), which is used to estimate network-wide origin-destination information for all travel in Greater London. This annual survey of London residents reaches approximately 8,000 households (12,000 individuals) and was last completed in April 2006. LTDS does not provide sufficient detail on passenger behaviour to be useful for route-level network planning; rather it is employed primarily to calibrate strategic transportation models. Nonetheless, aggregate travel patterns derived from Oyster data can be compared with metrics from LTDS.

METHOD FOR IDENTIFYING COMPLETE JOURNEYS

To create a network-wide dataset of complete, multi-stage journeys using smart card data, it is necessary to make assumptions about acceptable elapsed time thresholds between sequential journey stages for each passenger. This is difficult because the spatial accuracy of bus boarding locations is limited to the route level and there is no information on bus alighting time or location. The lack of a link between a vehicle and its location at the time of a fare payment transaction makes it difficult to link journey stages. Nonetheless, there is potential value in forming complete journeys based on elapsed time thresholds for the purposes of identifying and better understanding passenger behavior with respect to bus transfers.

Definition of a Transfer

A transfer is defined as the act of changing between modes or between services (i.e., vehicles) of the same mode. But, it is not quite so simple. Passengers may engage in a spectrum of time-consuming activities between journey stages, ranging from an activity that is the purpose of the journey to a non-travel activity incidental to the transfer, for example buying a newspaper. In the case of incidental activities, the passenger would consider the two journey stages to be part of the same complete journey so they should be linked together; however, in the case of an activity that was the main purpose of travel, the journey stages should not be linked into a complete journey even if the activity duration is very short.

One limitation of smart card data is that there is no way of determining what activities people are engaged in between journey stages. So, it is necessary to estimate time thresholds between journey stages within which most people are likely to be transferring, allowing time for incidental activities only. The elapsed time threshold would be the maximum allowable transfer time (or bus in-vehicle plus wait time) for two sequential stages to be considered part of the same journey.

This concept is illustrated in Figure 1, in which the horizontal axis represents the elapsed time threshold, and the vertical axis represents the share of transfers that are pure transfers as the time threshold increases. In other words, if the time threshold is set at x (or less) then all potential transfers are pure transfers, whereas if the time threshold is set between x and y then an increasing share of transfers will include time-consuming incidental activities. Elapsed time thresholds above y would misclassify a large share of non-transfers as transfers.

The formation of complete journeys based only on time thresholds between journey stages may result in some sequential stages being linked when in reality no transfer took place. Transfers that include incidental activities between journey stages should be classified as transfers because this reflects passengers' perceptions of their travel experience rather than simply a minimum time threshold needed to transfer. The difference between incidental and purposeful activities between journey stages cannot be identified using smart card data, so it is necessary to rely on a network-wide analysis of actual transfer times to determine which journey stages to link together. The assumption is that as long as the elapsed time thresholds are fairly representative for the system as a whole, then when a particular route is reviewed, the actual journey patterns for passengers on that route relative to network norms, as well as the physical and scheduling context, can be considered. In sum, the general goal in determining appropriate elapsed time thresholds is to include all pure transfers and incidental activity transfers whilst minimizing the number of non-transfers.

Identification of Transfers

To determine which journey stages to link into complete journeys, three transfer scenarios were examined: Underground-to-bus, bus-to-Underground, and bus-to-bus. Potential Underground-to-bus transfers are characterized by the time in minutes between Underground station exit and bus boarding. Both the station exit and bus boarding must be recorded by a transaction, or "tap", with a unique Oyster card. For a potential bus-to-Underground transfer, the "transfer" time threshold includes bus travel time in addition to the time taken to walk to the Underground station ticket gates after alighting from the bus. Finally, for potential bus-to-bus transfers, the time threshold includes not only the in-vehicle travel time on the first bus stage, but also the wait time for the second bus.

The following ranges of recommended elapsed time thresholds for identifying transfers in the London network were determined from an analysis of typical distributions of potential transfer times across the network (18):

- Underground-to-bus transfers from 15 to 25 minutes between Underground station exit and bus boarding;
- bus-to-Underground transfers, including bus in-vehicle time, from 30 to 50 minutes between bus boarding and Underground station entry; and
- bus-to-bus transfers, including bus in-vehicle time and waiting time for the second bus stage, from 40 to 60 minutes from first bus boarding to second bus boarding.

Elapsed time distributions of potential transfers were compared by time period, across fare zones, and by station type for each modal combination. Slight differences across these categories are accounted for in the range of recommended elapsed time thresholds.

Method for Linking Journey Stages Based on Maximum Elapsed Time Thresholds

An iterative method was used to link bus and Underground journey stages for each of 2.4 million passengers. A query first runs through each smart card ID and flags the journey stages, ordered chronologically, that should be linked to their successor based on the specified modal combination and elapsed time threshold. Sequential bus journey stages that satisfy the elapsed time threshold are not linked if they are on the same route because this is likely to be a return journey. The query then iterates a second time to number each complete journey for each passenger. Note that a complete journey may be comprised of a single journey stage.

It takes several hours to generate a network-wide dataset of complete journeys from a sequenced journeys file for a single day using the Structured Query Language (SQL) code developed for this

research. Once the dataset is generated, any subset of routes or stations of interest can be extracted in a few minutes without having to re-identify complete journeys. The research process to generate the results presented in this paper spanned several months but after the methodology was developed it was adapted for a different version of SQL (used by TfL) in a few weeks. Appropriate sampling techniques for smart card data in London should be explored in order to reduce the computational resource demands of using the entire dataset.

As presented in Table 1, three complete journey datasets were created for the entire network based on the mid-point and extremes of the range of elapsed time thresholds for each interchange type. Assuming the ranges of elapsed time thresholds are a reasonable reflection of actual transfer behavior, the results from the three datasets are expected to be similar. The three complete journey datasets are compared with LTDS to verify whether they are similar to survey-based results and could therefore be used for network planning.

RESULTS OF LINKING JOURNEY STAGES TO FORM COMPLETE JOURNEYS

Aggregate results of low, middle, and high elapsed time threshold approaches are now compared with results derived from TfL publications and the 2005/06 LTDS. The metrics used for comparison are: total journeys, journeys per passenger, stages per journey, and modal patterns. Other metrics, or the same metrics disaggregated by time of day or journey duration, could also be used to compare the two datasets however there are sample size issues in further disaggregating LTDS. The comparison presented in this paper is intended to be a broad-brush analysis in order to determine whether it is worthwhile to develop network planning applications as a further means of assessing the quality of the Oyster card data.

Total Daily Journeys

There are approximately 4 million average weekday journeys on the Underground network (19). Comparable statistics for average *weekday* bus journeys were not available, but there were 3.2 million average daily bus journeys in 2006/07 (4). (Note that average weekday bus journey stages are 38 percent higher than average weekend day bus journey stages.) Therefore, total daily weekday journeys for both modes is approximately 7.6 million (November 2007). However, travel by smart card only represents 73 percent of daily journeys, so the complete journeys datasets created using the elapsed time thresholds should include approximately 5.55 million complete journeys.

As shown in Table 2, the three sets of elapsed time thresholds yield a total of 5.3 to 5.5 million daily complete journeys (i.e., could include Underground-to-bus, bus-to-Underground or bus-to-bus transfers). These total complete journeys figures are all within 5 percent of the expected value of 5.55 million complete journeys. The high set of elapsed time thresholds may underestimate complete journeys slightly by linking too many journey stages together. The results suggest that the proposed ranges of elapsed time thresholds are indeed appropriate in London, but three other metrics are examined to determine which thresholds are preferred.

Public Transportation Journeys per Passenger

The authors' calculations using LTDS data show that those Greater London residents who travel by bus or Underground make an average of 2.05 complete journeys on these modes per weekday. As summarized in Table 2, the sets of elapsed time thresholds yield 2.23 to 2.33 daily journeys per passenger. These results are slightly higher than the expected value based on LTDS. This difference could be explained by London residents (LTDS data) making slightly fewer daily journeys by public transportation per day than the average TfL passenger (Oyster data), or by the difference in methodologies used to identify complete journeys.

To better understand the differences in estimates of daily public transportation journeys per passenger, the shares of passengers who make one to six journeys per day are examined. Figure 2 shows that using elapsed time thresholds at the mid-point of the suggested ranges, 51 percent of passengers make two journeys per day, and a further 37 percent make one or three journeys per day whereas, based on LTDS, 71 percent of passengers make two public transportation journeys per day and only 26 percent

make one or three journeys per day. Results for the high and low ends of the elapsed time threshold ranges (not shown) are similar. The discrepancy between the smart card and LTDS results is not negligible, and is difficult to explain. If the smart card methodology resulted only in too many people making one journey per day, then the elapsed time thresholds could be decreased to address the discrepancy. Conversely, if the smart card approach resulted only in too many people making three or more journeys per day then the elapsed time thresholds could be increased to form fewer journeys. However, since both the number of passengers making one journey and the number of passengers making three (or more) journeys per day are higher than expected, there is no obvious direction to adjust the elapsed time thresholds to make the results more consistent with LTDS.

The linking of journey stages into complete journeys using smart card data is based on simple elapsed time assumptions and does not explicitly take into account passengers' perceptions of what constitutes a complete journey as would be the case in LTDS. Future versions of the methodology could use different elapsed time thresholds per station or apply more spatial restrictions on which journey stages become linked. Time thresholds could also vary by time of day, for example peak versus off-peak, because people making trips with different purposes might tolerate different transfer times or longer trip durations. These adjusted methods would help to more accurately identify true complete journeys but some discrepancy with LTDS is still expected due to the differences in methodology.

Overall, there is some cause for concern that the complete journeys methodology does not accurately combine journey stages into two journeys per day for a sufficient number of passengers. On the other hand, smart card data constitutes a very large sample of public transportation users and this may simply reflect a greater variety of travel behavior than is captured by the small sample from LTDS. In addition, the passenger survey "recall" method used in LTDS might also be biased to produced more "typical" day results of two journeys (to and from work) per day. Moreover, this metric should be considered in combination with the other metrics presented in this section in evaluating the elapsed time threshold approach.

Stages per Public Transportation Journey

LTDS data indicates that in London the average number of stages per public transportation journey is 1.25. This value is similar to empirical results from previous analyses of smart card data that find a typical implied ratio of bus boardings to complete journeys on bus-only networks of 1.21 to 1.25 (8, 9).

Table 2 shows that the average number of stages per journey for complete journeys formed using the sets of elapsed time thresholds is 1.30 to 1.36. These values are slightly higher than the expected value from LTDS, especially given that LTDS identifies Underground-to-Underground transfers. Previous smart card analyses also suggest that the proposed elapsed time thresholds may be too high and therefore the best option may be the low set of thresholds.

Additionally, Figure 3 shows that the distribution of one-, two-, and three-stage journeys across passengers is almost identical between LTDS and the complete journeys formed using the low end of the elapsed time threshold ranges. This result further supports adoption of the low end of the elapsed time thresholds for identifying complete journeys.

Modal Patterns

Finally, the share of Underground journeys that include at least one bus journey stage is used as a metric to compare modal patterns. LTDS data indicates that 23 percent of Underground journeys include at least one bus journey stage. As shown in Table 2, at the low end of the elapsed time thresholds 23 percent of Underground journeys include bus stages. Moreover, in the case of complete journeys formed using the middle and high sets of elapsed time thresholds, 25 percent of Underground journeys include bus stages. Based on this metric, it is preferable to use the low end of the proposed ranges but results from the middle or high end of the ranges are not drastically different from expected values.

Summary

There are two possible reasons for some discrepancy between LTDS and smart card results. First, LTDS is a survey of London residents only whereas the smart card data captures anyone who travels on the TfL network using Oyster, whether or not they live in London. Second, the LTDS travel diary methodology for a small, statistically-constructed sample reveals how London residents perceive and recall their travel by public transportation as opposed to the network-wide representation of actual travel by bus and Underground obtained from smart card data.

The stages per journey and number of Underground journeys with bus stages metrics indicate that elapsed time thresholds at the low end of the ranges yield results closest to the expected values, but the daily public transportation journeys per passenger metric indicates that the middle or high end of the ranges yield results closer to the expected value. Nonetheless, in all cases, the results within the ranges are similar. Since there is no clear winner and the middle of range is very close to the expected values (which are approximate themselves), a complete journeys dataset generated from the mid-range values was used for the bus network planning application described below.

APPLICATION TO BUS NETWORK PLANNING

Smart card data can provide new information to network planners at TfL in a timely manner. New information derived from forming complete journeys using smart card data can be divided into two categories: (1) contextual knowledge about a route or station that may be quantitative or qualitative, and (2) quantitative inputs to the cost-benefit models currently employed to evaluate frequency, capacity, or restructuring changes to bus routes.

Contextual knowledge about a route or group of routes in the vicinity of an Underground station that may be gained from examining complete journeys data includes:

- connectivity of the route or station with the public transportation network,
- intermodalism of journeys on the route or through the station,
- bus access journey duration, and
- Underground-to-bus transfer time.

Quantitative inputs to the cost-benefit models used by bus network planners at TfL to evaluate proposed changes in the capacity or service frequency of a route may also be augmented by information from complete journeys data. With this data, it is possible to estimate the number of people transferring from one route to another over the course of a day or during any time period of interest. If service is increased on a low-frequency route, there may be greater benefit for transferring passengers than those who access the route by walking; they have less control over their arrival time because it is constrained by the schedule of the first service. The wait time savings of transferring passengers may be weighted more heavily in the cost-benefit model than those of non-transferring passengers. Conversely, a decrease in service on the second route could have greater disbenefit for transferring passengers and should be assessed accordingly.

Moreover, for two routes with a unique intersection, the complete journeys data can be used to determine how many passengers would no longer need to transfer if a new direct link were added. Additional passenger benefits in terms of wait time and transfer penalty savings could then be estimated.

Preliminary application of complete journeys data to selected bus routes (not shown) demonstrated that expected variation in connectivity and intermodalism across routes is borne out in the complete journeys for these routes, and that the expected intersecting routes and stations also emerge from the data. Preliminary application to Underground stations provides examples of the value of comparing bus access times and transfer times across routes, as well as identifying unexpected journey patterns such as a high volume of two-stage bus journeys made to access an Underground station. *In all cases, the simple metric of volume of interchanges between intersecting routes is a fundamental result.* Volume of interchanges is one of several important factors in assessing bus route restructuring, among others including revenue generation, crew scheduling, and service reliability on a particular corridor.

CONCLUSIONS

This paper contributes to an emerging body of literature on the application of smart card data to public transportation planning and addresses the issue of planning for transfers.

A simple methodology for obtaining new, valuable information from smart card data for application to bus network planning in London is developed. This methodology demonstrates that passenger flows between intersecting routes can be quantified with smart card data. This information might support rerouting decisions or be an input for cost-benefit. The quantification of passenger flows between a bus route and an Underground station could inform bus station design and other aspects of planning services to and from a station, and further quantitative information such as the percentage of inter-modal journeys on a route, the heaviest transfer locations on a route, and the repetition of individual travel patterns can help inform bus route planning.

Early work established ranges of elapsed time thresholds for linking journey stages to form complete journeys for the entire London network based on an analysis of elapsed times between journey stages for individual passengers across locations and time periods (18). Complete journeys formed by linking smart card journeys stages according to different time thresholds were compared to metrics from a travel diary survey. The results suggest that times at the low end or mid-point of each range are generally appropriate for creating complete journeys in London. Further informed by selected network planning applications, the recommended elapsed time thresholds for the TfL network are 20 minutes for Underground-to-bus transfers, 35 minutes for bus-to-Underground transfers, and 45 minutes for bus-to-bus transfers.

Key results about travel patterns from the smart card data are that approximately 23 percent of Underground journeys involve bus segments and 75 percent of bus or Underground journeys are single-stage journeys (although the Underground journeys may include “behind the barrier” transfers). Additionally, 20 percent of complete journeys are two-stage journeys and about 5 percent include three (or more) stages (see Figure 3). These types of metrics could be monitored to assess the impacts of measures intended to influence travel behavior, for example new fare policy. Smart card data is readily available and the added cost of querying the data for this type of information is likely to be smaller, as well as have greater coverage, than data from bespoke surveys. It would complement the depth of information about passenger characteristics and journey purpose gained from the more limited travel diary surveys currently used by TfL.

Ideally, different elapsed time thresholds would be applied to identify transfers at each Underground station or during different time periods. This would reduce the error rate of linking journey stages that actually form separate journeys or, conversely, of not linking journey stages that are actually part of complete journeys. Station- or time-specific elapsed time thresholds do not solve the limitation of route-level spatial accuracy for bus journeys, but this could be addressed through TfL’s new Automated Vehicle Locator (AVL) system (the “iBus” system). At its most basic level, the combination of AVL and smart card data would enable the location of the bus to be identified at the time a passenger boards the bus and taps their smart card on the reader. This type of stop-specific spatial information would add incredible richness to the smart card data to a level of detail that would not only make information on transfers, especially bus-to-bus, more precise by allowing for spatial thresholds as well as time thresholds, but also might eventually modify the sampling methodology for survey-based data collection that is currently used to develop passenger origin-destination matrices.

ACKNOWLEDGMENT

This research was conducted as part of a research agreement between TfL and the Massachusetts Institute of Technology. The financial support of TfL is gratefully acknowledged. The authors would also like to express their thanks to Rosa McShane, John Barry, and H el ene Bataille, all of TfL, for their significant contribution to the research and applications described in this paper. The views expressed in this paper are those of the authors.

REFERENCES

1. TfL. *Intermodal transport interchange for London: Best practice guidelines*. Transport for London, 2001.
2. TfL. *Interchange Plan – improving interchange in London*. Transport for London, 2002.
3. Transport for London, *London Travel Report 2007*. Transport for London, 2008.
4. Park, J. Y. and D.J. Kim. The Potential Use of Smart Card Data to Define the Use of Public Transit in Seoul. In *87th Annual Meeting of the Transportation Research Board*. CD-ROM. Transportation Research Board of the National Academies, Washington, DC, 2008.
5. TfL. *Traffic Report – Period 9 (11 November 2007 – 8 December 2007), Fares & Ticketing*. Transport for London, 2007.
6. Cui, A., *Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems*. Master of Science in Transportation Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2006.
7. Utsunomiya, M., J. Attanucci, and N. Wilson. Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning. In *Transportation Research Board: Journal of the Transportation Research Board, No. 1971*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 119-126.
8. Bagchi, M. and P.R. White. The Potential of Public Transport Smart Card Data. In *Transport Policy*, Vol. 12, 2004, pp. 464-474.
9. Hofmann, M. and M. O'Mahony. Transfer Journey Identification and Analyses from Electronic Fare Collection Data. In *Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems Meeting*, Vienna, 2005.
10. Okamura, T., J. Zhang, and F. Akimasa. Finding Behavioral Rules of Urban Public Transport Passengers by Using Boarding Records of Integrated Stored Fare Card System. In *Proceedings of the 10th World Conference on Transportation Research*, Istanbul, 2004.
11. Trépanier, M., N. Tranchant, and R. Champleau. Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System. In *Journal of Intelligent Transportation Systems: Technology, Planning and Operations*, Vol. 11, no. 1, 2007, pp. 1-14.
12. Zhao, J., A. Rahbee, and N.H.M. Wilson. Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems. In *Computer-Aided Civil and Infrastructure Engineering*, Vol. 22, no. 5, 2007, pp. 376-387.
13. Chan, J. *Rail transit OD matrix estimation and journey time reliability metrics using automated fare data*. Master of Science in Transportation Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2007.
14. Chu, K. K. A. and R. Champleau. Enriching Archived Smart Card Transaction Data for Transit Demand Modeling. In *87th Annual Meeting of the Transportation Research Board CD-ROM*. Washington, DC, 2008.
15. Mayor of London. *TfL Fact Sheet*. Transport for London, 2007.
16. Greater London Authority. *London's changing population: Diversity of a world city in the 21st century*. Greater London Authority, 2005.
17. TfL. *London Travel Report 2006*. Transport for London, 2007.
18. Seaborn, C.W. *Application of Smart Card Fare Payment Data to Bus Network Planning in London, UK*. Master of Science in Transportation Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2008.
19. Mayor of London. *Transport – facts and figures*. Greater London Authority, 2008.

FIGURE 1 Conceptual diagram of elapsed time threshold versus transfer type.

TABLE 1 Ranges of Elapsed Time Thresholds for Identifying Transfers

TABLE 2 Travel Metrics for Smart Card Complete Journey Data Versus Survey Data

FIGURE 2 Daily public transportation journeys per passenger.

FIGURE 3 Stages per complete journey by time threshold range versus survey data

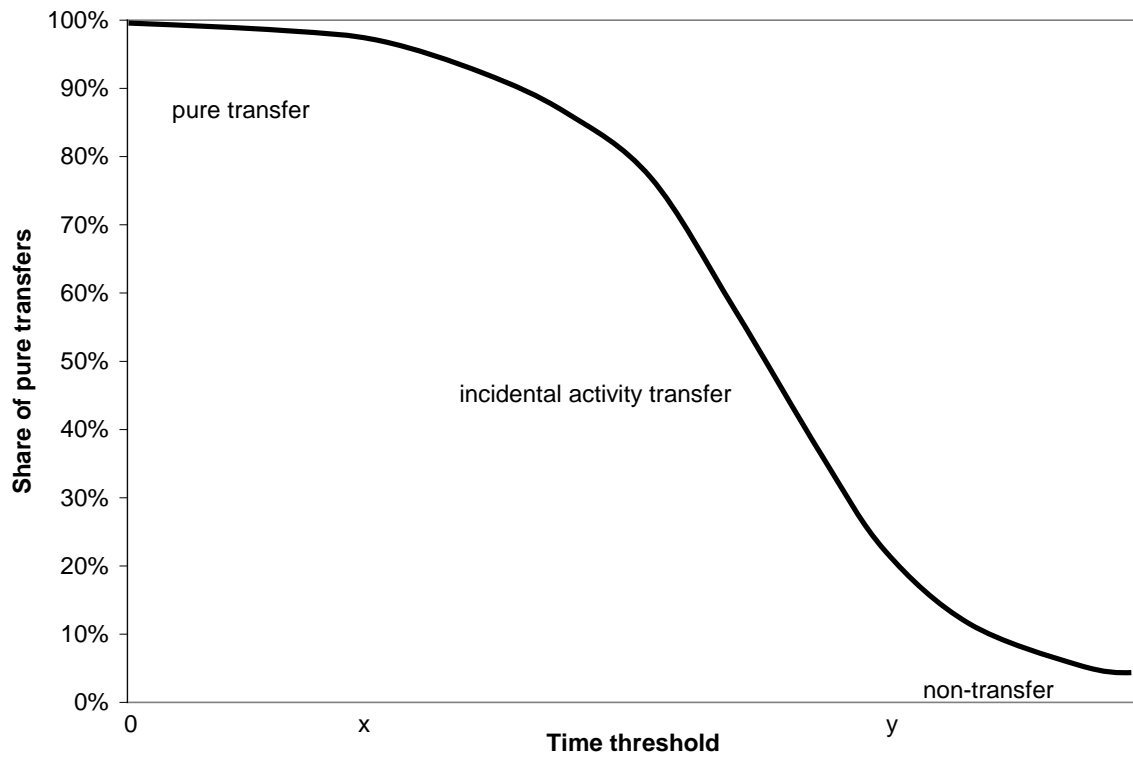


FIGURE 1 Conceptual diagram of elapsed time threshold versus transfer type.

TABLE 1 Ranges of Elapsed Time Thresholds for Identifying Transfers

Transfer Type	Elapsed Time Threshold		
	<i>Low range</i>	<i>Middle range</i>	<i>High range</i>
Underground-to-bus	15 minutes	20 minutes	25 minutes
Bus-to-Underground	30 minutes	40 minutes	50 minutes
Bus-to-bus*	40 minutes	50 minutes	60 minutes

*Limited to journey stages on different routes.

TABLE 2 Travel Metrics for Smart Card Complete Journey Data Versus Survey Data

Elapsed Time Thresholds	Travel Metric			
	Total daily journeys (million)	Daily PT* journeys per passenger	Average stages per PT* journey	Underground journeys with bus stage (%)
Low range	5.5 (1%)**	2.33 (14%)	1.30 (4%)	23 (0%)
Mid range	5.4 (3%)	2.27 (11%)	1.34 (7%)	25 (9%)
High range	5.3 (5%)	2.23 (9%)	1.36 (9%)	25 (9%)
<i>Expected value from survey</i>	<i>5.55</i>	<i>2.05</i>	<i>1.25</i>	<i>23</i>

* PT = public transportation

** difference from expected value (based on travel survey) shown in brackets

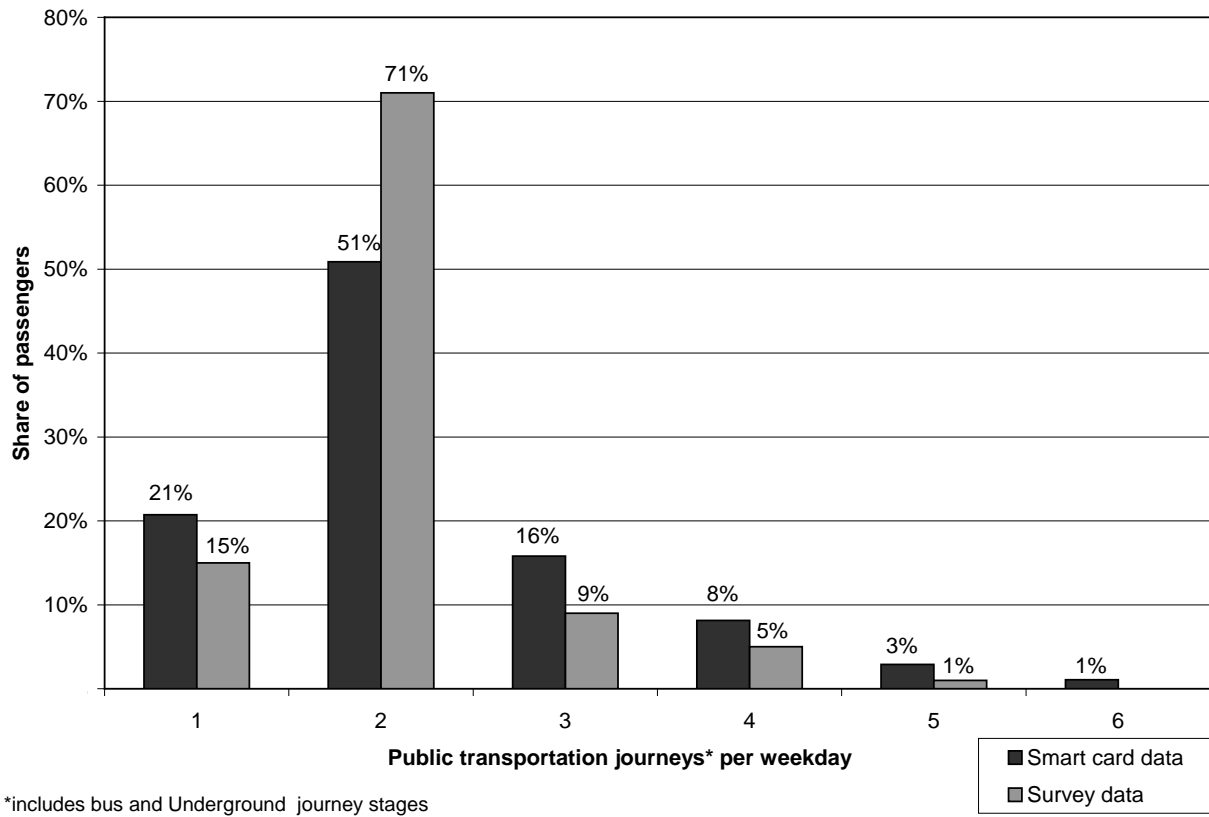


FIGURE 2 Daily public transportation journeys per passenger.

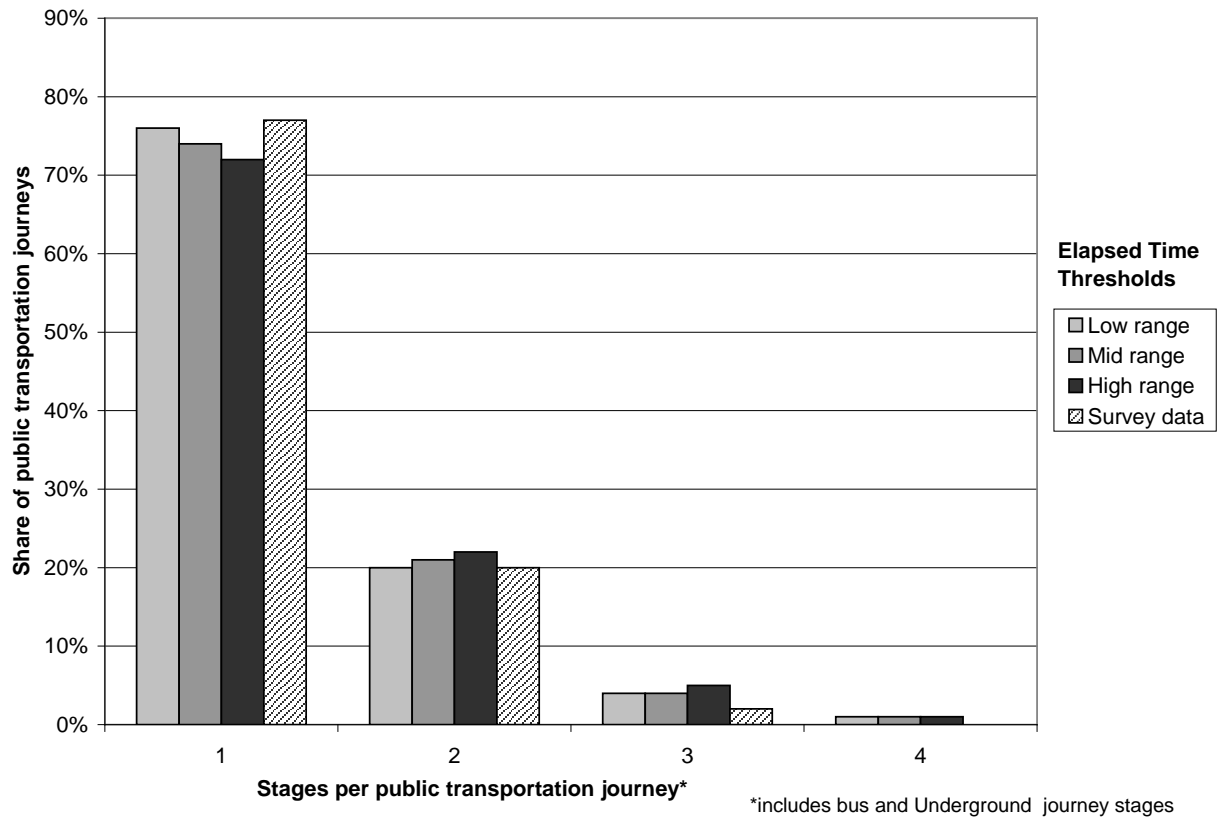


FIGURE 3 Stages per complete journey by time threshold range versus survey data.