

**Utilizing Automatically Collected Smart Card Data to  
Enhance Travel Demand Surveys**

by

Laura K. Riegel

Bachelor of Science in Industrial Engineering and Management Sciences  
Northwestern University, 2009

Submitted to the Department of Civil and Environmental Engineering  
in partial fulfillment of the requirements for the degree of

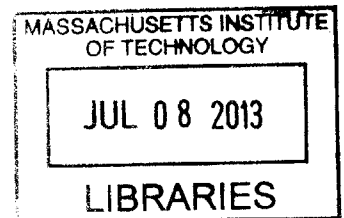
Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

**ARCHIVES**



© Massachusetts Institute of Technology 2013. All rights reserved.

Author .....  
Department of Civil and Environmental Engineering  
May 10, 2013

Certified by .....  
John P. Attanucci  
Research Associate of Civil and Environmental Engineering  
Thesis Supervisor

Certified by .....  
Mikel E. Murga  
Research Associate of Civil and Environmental Engineering  
Thesis Supervisor

Accepted by .....  
Heidi M. Nepf  
Chair, Departmental Committee for Graduate Students



# Utilizing Automatically Collected Smart Card Data to Enhance Travel Demand Surveys

by  
Laura K. Riegel

Submitted to the Department of Civil and Environmental Engineering  
on May 10, 2013, in partial fulfillment of the  
requirements for the degree of  
Master of Science in Transportation

## Abstract

Public transport agencies have used manual surveys to collect demographic and travel diary information in order to understand their customers' travel behavior for many years. Recently many agencies have also begun to use automated sources of data from fare collection, vehicle location, and passenger counting systems to improve the understanding of their customers' detailed geographic and temporal travel behavior as well as frequency of usage, and travel pattern variation at a much larger scale than is possible with manually collected survey data. Transport for London (TfL), the public body responsible for all transportation services in London, was chosen as a case study to determine how and to what extent automatic fare card (Oyster) data can be used to enhance and validate the London Travel Demand Survey (LTDS) single day travel diary responses.

This thesis found that combining survey responses with linked Oyster data for specific households could greatly enhance the validity of the single travel day and improve the understanding of the variability of weekly public transport (PT) use. However, it was difficult to match the survey diary responses and Oyster card records after the interview had taken place. This was evidenced by the fact that only 51.1% of Oyster journey stages had matching survey journey stages, only 45.6% of survey stages had matching Oyster stages, and only 44% of the sample had perfectly matching survey and Oyster stages. Even when there were matches, there were large differences in many journey start times and durations with an average start time difference of 61.2 minutes. This suggests that it would be advantageous to integrate the Oyster records earlier in the survey process, using some type of prompted recall methods with Oyster records in the near term, and new location tracking smart phone applications in the future. Analysis of the weekly variation in PT travel found that the single day survey overestimates typical PT use overall, but it underestimates the intensity of PT use on days when the survey sample chose to use the PT mode. Additionally, the reported frequency of PT use in the LTDS was significantly higher than the actual use as captured by the Oyster system, and therefore the LTDS is generally overestimating the PT use overall for London residents.

Thesis Supervisor: John P. Attanucci  
Title: Research Associate of Civil and Environmental Engineering

Thesis Supervisor: Mikel E. Murga  
Title: Research Associate of Civil and Environmental Engineering



## Acknowledgments

This thesis, and the research that preceeded it, would not have been possible without the help and support of many people.

- Thank you to John Attanucci and Mikel Murga, who guided this research, provided thoughtful perspectives, and spent many hours editing and improving this thesis. Thank you to Nigel Wilson, Haris Koutsopoulos, Marta Gonzalez, David Block-Schachter, and Fred Salvucci for your advice and feedback on this thesis.
- Thank you to Dale Campbell and Andrew Gaitskell and other employees of Transport for London for making this research possible. Thank you also for welcoming me in London and answering all of my many questions about the complexities of the TfL network.
- Thank you to the CEE staff for your dedication and support over the past two years.
- Thank you Cecilia for putting up with me as a roommate, officemate, exercise buddy, and costume enthusiast, but mostly thank you for being my friend.
- Thank you to Yonni for all of your beautiful singing and incredible dancing, but I'm still mad at you for graduating early. Thank you to Dom for always finding cool transit maps to show me and for your passion for improving the T. If anyone can fix it, it's you. Thank you to Kevin and Jay for teaching me what an Oyster Record is, and to Gabriel for explaining how to be a second year. Thank you to the many other members of the MST program and MIT community for their support, advice, and friendship.
- Finally, I would like to thank my parents, sister, and grandma for their unconditional support and love. I know that you didn't always know what I was doing, but you knew I could do it.



# Contents

<b>List of Figures</b>	<b>9</b>
<b>List of Tables</b>	<b>11</b>
<b>1 Introduction</b>	<b>13</b>
1.1 Research Motivation . . . . .	13
1.2 The Case Study . . . . .	14
1.3 Research Approach . . . . .	14
1.4 Thesis Organization . . . . .	15
<b>2 Background and Literature Review</b>	<b>17</b>
2.1 TfL's PT Network and Available Data . . . . .	18
2.1.1 Description of Available London Travel Demand Survey Data . . . . .	18
2.1.2 Description of Available Automatically Collected Fare Card Data . . . . .	21
2.2 Intrapersonal Variability in Travel Behavior . . . . .	22
2.3 Household Travel Surveys . . . . .	23
2.4 Combining Automatically Collected and Manual Survey Data . . . . .	24
2.4.1 GPS Studies . . . . .	24
2.4.2 Mobile Phone Studies . . . . .	26
2.4.3 Smart Card Studies . . . . .	28
2.5 Summary of Background and Literature Review . . . . .	29
<b>3 Results of Combining LTDS and Oyster Data</b>	<b>31</b>
3.1 Categorization of Person Types . . . . .	31
3.1.1 Type 1 - PT Users with Similar LTDS and OR Travel . . . . .	32
3.1.2 Type 2 - PT Users with Different LTDS and OR Travel . . . . .	32
3.1.3 Type 3 - PT Users with LTDS but No OR Travel . . . . .	33
3.1.4 Type 4 - PT Unreported on LTDS with PT OR on Travel Day . . . . .	33
3.1.5 Type 5 - PT Unreported on LTDS with Different OR Travel . . . . .	33
3.1.6 Type 6 - PT Unreported on LTDS with No OR Travel . . . . .	33
3.1.7 Type 7 - No Travel on LTDS with PT OR on Travel Day . . . . .	34
3.1.8 Type 8 - No Travel on LTDS with Different PT OR Travel . . . . .	34
3.1.9 Type 9 - No Travel in LTDS or OR . . . . .	34
3.2 Difference in Overall Magnitude of LTDS and OR Stages . . . . .	35
3.2.1 Difference in Overall Magnitude of LTDS and OR Stages by Mode . . . . .	36
3.2.2 Mode Share . . . . .	40
3.3 Difference between Specific LTDS and OR Stages . . . . .	40

3.4	Weekly Variation in PT Travel . . . . .	45
3.5	Summary of Combining LTDS and Oyster Data . . . . .	50
<b>4</b>	<b>Analyses of Oyster Card Users (Type 1 Respondents)</b>	<b>53</b>
4.1	Representativeness of Type 1 Respondents . . . . .	53
4.2	Overall Number of PT Stages . . . . .	59
4.2.1	Intensity of PT Use . . . . .	59
4.2.2	Reported vs. Actual Frequency of Use . . . . .	64
4.2.3	Difference in Oyster Card Types . . . . .	65
4.2.4	Difference in Stages per Person . . . . .	67
4.3	Unreported Oyster Records on Travel Day . . . . .	69
4.4	Reported Stages not Captured in Oyster Records . . . . .	70
4.5	Summary of Misreporting Specific PT Stages . . . . .	71
4.6	Trip Analysis . . . . .	72
4.7	Stage Analyses by Mode . . . . .	73
4.8	Stage Start Time Analyses . . . . .	76
4.9	Stage Durations . . . . .	83
4.10	Average Speeds used in LTDS . . . . .	85
4.11	People with Perfectly Matching Stages . . . . .	86
4.12	Summary of Type 1 Analyses . . . . .	87
<b>5</b>	<b>Analyses of Other Person Types</b>	<b>91</b>
5.1	Type 2 Analyses . . . . .	91
5.1.1	Type 2 Intensity of PT Use . . . . .	91
5.1.2	Type 2 Reported LTDS Stages found in OR within Previous Week . . . . .	95
5.2	Type 3 Analyses . . . . .	96
5.3	Types 6 and 9 Analyses . . . . .	96
5.4	Types 4 and 7 Analyses . . . . .	97
5.4.1	Types 4 and 7 Unreported Oyster Records on Travel Day . . . . .	97
5.4.2	Types 4 and 7 Intensity of PT Use . . . . .	98
5.5	Types 5 and 8 Analyses . . . . .	102
5.6	Summary of Other Person Types Analyses . . . . .	103
<b>6</b>	<b>Conclusions and Recommendations</b>	<b>105</b>
6.1	Conclusions from Combining LTDS and Oyster Data . . . . .	106
6.2	Variability of PT Travel . . . . .	107
6.3	Selection and Representativeness of the Type 1 Panel . . . . .	108
6.4	Recommended Next Steps . . . . .	109
6.4.1	Single Day Responses . . . . .	110
6.4.2	Travel Behavior Over Time . . . . .	110
<b>A</b>	<b>LTDS Trip Purpose Definitions</b>	<b>113</b>
<b>B</b>	<b>Demographic Characteristics</b>	<b>115</b>
	<b>Bibliography</b>	<b>119</b>



# List of Figures

3-1	Reported Frequency of PT Use (Max of All PT Modes)	47
3-2	OR Stage Frequency Two Weeks Before and After Travel Date	47
3-3	Percent of Days People who Reported PT Stage(s) on Travel Day have OR Stages	49
4-1	Comparison of PT Stages for Type 1 vs. All Others who Reported PT	55
4-2	Difference of % Type 1 People and % of Other People who Reported PT by Borough	57
4-3	LTDS vs. Oyster Stages for Type 1 People with Weekday Travel Days	62
4-4	LTDS vs. Oyster Stages for Type 1 People with Weekend Travel Days	63
4-5	Type 1 Reported vs. Actual Frequency of PT Use	64
4-6	Difference in PT Stages for Type 1 People by Oyster Card Type	66
4-7	Difference in Number of PT Stages (LTDS - OR) by Oyster Card Type	67
4-8	Difference in Number of PT Stages (LTDS - OR)	68
4-9	Number of Reported LTDS Stages vs. Difference in OR and LTDS Stages	69
4-10	Percent of Each Trip that has OR for each Stage by Number of Stages per Trip	72
4-11	Match by Mode	73
4-12	Matching Train Journeys	76
4-13	Stage Start Time, Type 1: 10 Min. Interval	79
4-14	Number of LTDS Start Times within Interval minus Number of Oyster Start Times within Interval	80
4-15	Difference in Stage Start Times for Matching Journeys with OR within two hours of Reported Start Time	81
4-16	Abs. Value of Diff. in Stage Start Times for Matching Journeys with OR (within three hours) by Trip Destination Purpose	82
4-17	Duration Difference for Matching Train Journeys	83
4-18	Absolute Value of Difference in LTDS and Oyster Durations for Matching Train Journeys, in Minutes	84
4-19	Percent Error of Speed (LTDS - Oyster/Oyster) for Matching One Stage Train Journeys	85
4-20	Distribution of PT Stages for Exact Journey People	86
4-21	Percent of People with Exact Journey Sequence by Modes Used	86
5-1	LTDS vs. Oyster Stages for Type 2 People with Weekday Travel Days	93
5-2	LTDS vs. Oyster Stages for Type 2 People with Weekend Travel Days	94
5-3	Type 2 Oyster Categories	95

5-4	Reported PT Stages of Type 3 People Compared to All People who Reported PT . . . . .	96
5-5	Type 4 OR the Week Before Travel Day . . . . .	98
5-6	Type 7 OR the Week Before Travel Day . . . . .	98
5-7	LTDS vs. Oyster Stages for Types 4 and 7 People with Weekday Travel Days	100
5-8	LTDS vs. Oyster Stages for Types 4 and 7 People with Weekend Travel Days	101
5-9	Types 5 and 8 Reported vs. Actual Frequency of PT Bus Use . . . . .	102
B-1	Demographic Characteristics of Person Types 2 through 9 . . . . .	117

# List of Tables

2.1	Key LTDS Data Items . . . . .	20
3.1	Person Types for People who Volunteered Oyster Card Number(s) . . . . .	32
3.2	Total Difference between Reported LTDS PT Stages and/or OR Stages . . . . .	35
3.3	Total Difference between Reported LTDS Bus Stages and/or OR Bus Stages . . . . .	37
3.4	Total Difference between Reported LTDS Train Stages and/or OR Train Stages . . . . .	39
3.5	Summary of Bus and Train Mode Share . . . . .	40
3.6	Summary of Over, Under, or Misreporting of Specific PT Stages . . . . .	43
3.7	Summary of Over and Underreporting of Specific PT Stages, Exclusive People . . . . .	45
4.1	Comparison of Average Number of PT Stages . . . . .	55
4.2	Mode Share Differences for People who Reported PT (not Type 1) and Type 1 . . . . .	56
4.3	Comparison of Demographic Characteristics . . . . .	58
4.4	Type 1 Overall Amount of PT Travel . . . . .	59
4.5	Categories of Reported Frequency Converted to Use per Week . . . . .	64
4.6	Average PT Stages by Oyster Card Type . . . . .	65
4.7	Summary of Over, Under or Misreporting of Specific PT Stages . . . . .	71
4.8	Matching Journeys by Mode . . . . .	74
4.9	Number of Bus Journeys per Person . . . . .	74
4.10	Number of Train Journeys per Person . . . . .	74
4.11	Matching Bus Routes and Taps per Person . . . . .	75
5.1	Summary of Over, Under or Misreporting of Types 2 - 9 . . . . .	103
B.1	Comparison of Demographic Characteristics . . . . .	116



# Chapter 1

## Introduction

Public transport agencies have used manual surveys to collect demographic and travel diary information in order to understand their customers' travel behavior for many years. Recently many agencies have also begun to use automated sources of data from fare collection, vehicle location, and passenger counting systems (generally referred to as automated data collection or ADC systems). These systems help improve the understanding of their customers' detailed geographic and temporal travel behavior as well as frequency of usage and travel pattern variation at a much larger scale than is possible with manually collected survey data.

Transport for London (TfL), the public body responsible for all transportation services in London, has conducted large-scale comprehensive household travel surveys since 1962, with continuous collection of the survey in its current form, the London Travel Demand Survey (LTDS), since 2005. This survey combines socio-demographic data with diary based records of Londoners' travel on a single day. TfL has also invested in ADC systems, including its automatic fare collection (Oyster) system, that provides a vast and detailed source of data that can be used to understand how its customers use the public transport (PT) network in London. With the voluntary collection of LTDS respondents' Oyster card numbers in April 2011 it was possible to combine specific travel diary records of PT travel with those respondents' automatically collected fare card records (Oyster Records or OR) for the first time. This thesis uses TfL as a case study to determine how and to what extent automatic fare card data can be used to enhance and validate the manually collected single day travel diary responses and improve the understanding of PT travel behavior over a span of time not available with current survey methods.

### 1.1 Research Motivation

For many years, manually collected surveys were the only way for public transport and other planning agencies to gather data and draw conclusions about travel behavior of individuals and households in an area. The collection of manual surveys can be quite burdensome and inefficient because they generally have a high marginal cost and a small sample size relative to overall population and PT ridership. They also may not capture much information

for infrequent PT users and are often limited to short time periods, like a single travel day. Additionally, the individual being surveyed may misinterpret a question or answer incorrectly for a variety of reasons. They may choose to report their typical day of travel instead of the trips they actually took on the survey day in question, they may forget short trips, or they may purposely exclude trips to shorten the length of the interview or questionnaire. This misreporting in manual surveys can result in incorrect estimations of travel behavior. However, many organizations now have access to ADC systems that provide a vast source of continuous information about the travel behavior on PT networks, at a negligible cost after the initial capital and programming investments, that can be used to some degree to validate and enhance manual survey responses.

## 1.2 The Case Study

London residents account for approximately three-quarters of all travel in London, and their travel behavior on all modes is estimated using the LTDS. The LTDS is used within TfL to inform travel demand models with trip purpose, travel frequency, and mode share distributions, to contribute to analysis of policy at various levels of geography, and to develop the sub-regional transport plans. The LTDS data is also used to monitor the implementation of policies and to understand their impact on Londoners' travel pattern (Transport for London, 2009). In April 2011, the LTDS started recording up to two of the most frequently used TfL smart card (Oyster card/Freedom pass) numbers of willing survey participants aged 16 or older. Respondents agreed that automatic fare collection records for the provided Oyster card numbers could be used by TfL to supplement their specific travel day diary responses. This thesis combined these data sets to enhance and validate the survey responses on the single travel day and contribute to the understanding of the variability in travel behavior over time.

## 1.3 Research Approach

The volunteered Oyster card numbers of LTDS respondents were encrypted and stored as unique 9-digit "Prestige ID's" that were matched (when possible) to Oyster data collected by TfL. This thesis analyzed nine months of LTDS survey responses from 1 July, 2011 through 31 March, 2012 including demographic and travel diary information for all households and individuals that were surveyed. Oyster records (OR) were analyzed for all individuals that volunteered their Oyster card number from eight weeks prior to the travel day starting on 19 June, 2011 (including weekends) through 18 June, 2012. Therefore the OR analysis period referred to throughout this thesis is from 19 June, 2011 through 18 June, 2012.

With the matching of the survey PT stage responses and OR stages, the accuracy of PT travel behavior was determined by comparing an individual's reported day of travel with their journeys captured by the Oyster system. This was only possible for the people who decided to volunteer their Oyster card number, reported using PT on their travel day, and/or had valid<sup>1</sup> OR stages on that travel day, i.e., the people categorized in this thesis

---

<sup>1</sup>A valid OR stage is any type of PT use and does not include purchasing Oyster credit or a period pass.

as the “PT Sample”. A subset of people in this sample, called Type 1 people, reported PT stages on their travel day *and* had valid OR stages on their travel day. Therefore their LTDS and OR PT stages were directly compared on the basis of the overall number of PT stages, train origin and destination (O-D) pairs or bus routes of PT stages, modes used, and temporal characteristics. The other people in the PT Sample either reported LTDS PT stages and had no OR stages on their travel day, or had OR stages on their travel day but did not report LTDS PT stages. Their LTDS and/or OR PT stages (on the travel day and other days around the travel day) were analyzed and assumptions were made to determine the reasons for these inconsistencies. By examining the specific reported LTDS and OR stages for all people and making assumptions about the reasons for any discrepancies, the misreporting of the fare media used, which was assumed to be of minor significance to TfL’s planning efforts, was separated from the misreporting of O-D pairs or bus routes, and the overreporting and underreporting of PT journeys to give an overall estimate of the accuracy of the single day diary responses of PT travel behavior reported in the LTDS. The database was also queried to determine if the reported travel day was representative of an “average day” by analyzing all PT stages that were captured in the Oyster system over the OR analysis period.

## 1.4 Thesis Organization

The remainder of this thesis is organized into five additional chapters. Chapter 2 provides further background on the LTDS, TfL’s transport network and ADC systems, summarizes previous research that integrates manually collected travel survey data with automatically collected sources of data, and summarizes prior work done examining the weekly variation in PT travel. Chapter 3 introduces the categorization of LTDS respondents depending on their reported PT travel behavior in the LTDS and their PT activity found in the Oyster card system on their travel day. This chapter also has a summary of the difference between the overall number of LTDS and OR stages and a comparison of the reported weekly variation in PT travel and the variation of PT use captured by the Oyster system. Chapter 4 discusses the selection and representativeness of the panel of Type 1 LTDS respondents and the detailed analyses of the Type 1 panel to determine how and to what extent Oyster data can be used to validate and enhance the LTDS survey responses. Chapter 5 describes the analyses of the other LTDS respondents who provided Oyster card numbers and the additional conclusions drawn from studying their reported LTDS and Oyster data. Finally, Chapter 6 summarizes the thesis research and results and suggests ideas for future research, including the best way to integrate the two sources of data in the future.

It is important to note that this thesis only applies to the analysis of the travel behavior of London residents, as the LTDS is primarily focused on the travel of London residents and does not survey any visitors or non-resident commuters unless they are present in a London household at the time of the interview. This should be kept in mind when extrapolating conclusions for the travel behavior of all people using the London transport network as visitors and non-resident commuters make up approximately one quarter of all travel in London and may have different travel patterns to that of resident commuters or other residents of London.





## Chapter 2

# Background and Literature Review

For many years, manually collected surveys of travel diary and socio-demographic information were the best way for public transport (PT) and other planning agencies to gather data and draw conclusions about the travel behavior of individuals and households in an area. However, there are many drawbacks to using manual surveys, including the high cost, small sample sizes, and limited response accuracy that is dependent on respondent's subjective recollection of travel. However, many organizations now have access to highly accurate automated data collection (ADC) systems that provide a vast source of information, at a very low marginal cost, about the travel behavior on PT networks that can be used to some degree to validate and enhance the travel diary survey responses and other manually collected travel behavior information.

This improved access to ADC systems has introduced the possibility of studying the integration of manually collected survey data and automatically collected fare card or GPS data over the last few years. This thesis goes further and analyzes a much larger and more detailed sample of households than the scale of most previous GPS studies, and unlike previous smart card studies, has household survey data linked with the specific household smart card data allowing for a more in-depth study of the accuracy and representativeness of a larger sample of survey responses than has previously been possible. Transport for London (TfL) was chosen as a case study for this thesis because it conducts an on-going comprehensive travel demand survey and also has a vast source of automatically collected data that can be used to understand how its customers use the PT network in London. This chapter provides background information about London's Travel Demand Survey (LTDS), TfL's PT transport network and ADC systems, summarizes prior work that examined the weekly variation in PT travel, and summarizes previous research that integrated manually collected travel survey data with automatically collected sources of data.

## 2.1 TfL's PT Network and Available Data

TfL is the integrated body responsible for London's transportation system. Its main role is to implement the Mayor of London's Transport Strategy and manage transportation services in London. This includes bus services (London buses), metro service (London Underground), regional rail (National Rail and London Overground), light rail (Docklands Light Rail (DLR) and Tramlink), London River Services, and the Victoria Coach Station, as well as managing the Congestion Charge, maintaining the main roads and London's traffic lights, regulating taxis, and promoting walking and cycling initiatives (Transport for London, 2013).

TfL has many sources of data that can be used to analyze its customers' travel behaviors. One of the primary sources is the LTDS, which is a paper-based manual household travel survey that collects household demographic data and detailed journey information for individuals on a single travel day. TfL also has a vast source of detailed and exact temporal and geographic records of individual journeys made on its network that are automatically collected from its fare collection (Oyster) system. In April 2011, the LTDS started recording up to two of the most frequently used fare payment smart cards (Oyster cards or discounted "Freedom Passes" for the elderly or people with disabilities are collectively referred to throughout this thesis as Oyster cards) of willing survey participants aged 16 or over. With the voluntary collection of LTDS respondents' Oyster card numbers in April 2011, it was possible to combine manual household travel survey data with automatically collected fare card data for the first time, allowing for the integration of the strengths of each data set.

### 2.1.1 Description of Available London Travel Demand Survey Data

The LTDS is a rolling sample of households and individuals residing within Greater London that over time is intended to build up to a comprehensive picture of the travel behavior of Londoners. The survey has been running since 2005 with a target annual sample size of 8,000 households. The LTDS captures comprehensive demographic information for each household overall and for its individual members during an in-home interview using a paper-based questionnaire. The interviewers also spend time asking each member of the household (aged five and over) about his or her individual travel on a single day, which is generally the day prior to the interview. The details of this travel day include temporal, geographic, modal, and journey purpose information. The interviewer records the information on a paper questionnaire and sends the completed questionnaire by mail to the agency contracted to enter and maintain the manually collected survey data in computer databases. Annual results can be used to compare inner and outer London, but three years of data should be compiled to compare travel for individual London boroughs (Transport for London, 2009).

This research analyzes three quarters of LTDS survey responses covering the nine months between July 2011 and March 2012. There were 14,325 people from 5,976 households (HH's) that responded to the LTDS between July 2011 and March 2012. Of those, 4,053 people (28%) volunteered one Oyster card number, and another 76 people (1%) volunteered two

Oyster card numbers for a total of 4,129 people (29%) who provided Oyster card numbers. These volunteered card numbers were encrypted and the resulting number is referred to by TfL as a “Prestige ID”. The participants who reported cards from the following three categories were not included in this analysis because there was no way to uniquely assign the Prestige ID number to that person:

- 13 households where two or more people shared an Oyster card (0.20% of HH’s)
- Six cards (0.15% of cards) shared in multiple HH’s
- Nine cards (0.22% of cards) where a 9-digit Prestige ID is not uniquely identifiable

This resulted in 4,102 unique 9-digit Prestige ID’s from 4,081 people (28% of total people). There were not very many households that reported multiple household members sharing cards (13 HH’s). This may be a reflection of reality because most people who use PT regularly carry their cards with them at all times and therefore do not share a card with someone else in their household. The cards that were shared between multiple households (six cards) must be a mistake in transcription of the Oyster card number as it is highly unlikely that people in different households that were selected to ensure demographic diversity throughout London would share an Oyster card. Also, the households and people with a travel day between 1 - 10 September, 2011 were excluded from the analysis due to an error in collecting Oyster records (OR) during this time. Therefore, there were 3,946 people (28% of people) from 2,830 households (47% of households) who volunteered their Oyster card number and whose reported PT travel was compared to the PT stages captured by the Oyster system.

Key data items captured in the LTDS survey are defined in Table 2.1.

**Table 2.1:** Key LTDS Data Items

Term	Definition
Household ID	Unique 8-digit code for each household. A household is defined by TNS, the surveying agency, as “a single person or group of people who normally live together and have the exact address as their only or main residence and who either are catered for by the same person or share cooking facilities and a living room, sitting room, kitchen, or dining room.”
Person ID	Unique 10 digit code for each person. Every person in the household aged five and over is surveyed regarding their journeys on the travel day.
Prestige ID	Unique 9-digit code for each Oyster card (encrypted Oyster card number)
Travel Day	The day travel took place, generally the day immediately prior to the interview day, and mutually agreed upon by the interviewer and the respondent before the interview. The travel day is the same for every person in the same household.
Trip	“A one way journey to accomplish a purpose.” A trip must always have an origin and a destination, and these must always be different. The trip must also always have a purpose and a method of travel used. A trip is made up of multiple stages may have multiple modes broken up into multiple stages. Trip data used in this analysis include start and end times, purpose, and number of stages.
Stage	A stage is a portion on a trip that is confined to one mode. Each change in mode or interchange on the same mode is a new stage. Walking transfers between modes are also stages. Stage data used in this analysis include: mode, ticket type, duration, origin, destination, and route number (if bus). Walk durations are given by the respondent. The remaining durations are calculated automatically using the origin and destination of the stage, the overall trip start and end time, and an average speed value for the mode(s) used. These durations are checked against the transportation network by TNS as soon as possible after the interview to ensure journey feasibility. OR stages do not capture interchanges in most instances. Therefore some LTDS train stages were combined in this analysis to make the LTDS data comparable to OR stages.
Ticket Type	The method of payment for each mode of transport that requires it. People are asked what method of payment they used for each PT stage. These payment methods are coded in order to determine which PT stages were paid for on an Oyster card and should therefore have corresponding OR stages.
Mode	The PT modes included in this analysis are: bus, Underground, Docklands Light Rail, Overground, National Rail, and Tram.
Demographic Information	The survey collects a vast amount of demographic data. Specific demographic information analyzed included gender, ethnicity, household income, size and vehicles owned, disability, and age.

### 2.1.2 Description of Available Automatically Collected Fare Card Data

TfL's Oyster card is accepted on all of the PT services within TfL's service area (excluding ferries), and over 80% of PT journeys in London are made using an Oyster Card (Muhs, 2012). Oyster card data provides a continuous collection of the PT travel behavior of millions of Londoners and visitors every day. These data allow TfL to monitor fare collection and revenue electronically, provide reliable estimates of the travel time flows on its network, among many other useful applications. However, the Oyster card database only includes basic demographic data, including gender, postcode, and age, for voluntarily registered Oyster customers. Many other cards have no associated demographic data associated with them making it difficult to disaggregate travel patterns among specific demographic groups. Additionally, Oyster data only provides information about PT travel and does not capture travel on other modes like automobiles, walking, or cycling.

OR stages for each volunteered Oyster card were generally stored for all days (including weekends) eight weeks prior to the individual's travel day starting on 19 June, 2011<sup>1</sup> and indefinitely, though this study only analyzes OR stages through 18 June, 2012. Therefore the OR analysis period referred to throughout this thesis is from 19 June, 2011 through 18 June, 2012. OR entries and exits at each gated rail station include the type of tap (completed or uncompleted entry, completed or uncompleted exit, or Tram), time of entry/exit, the station code, the ticket type, and the sequence number for that card, among other data including fare information that are not used for this analysis. Un-gated rail stations (at some National Rail (NR) and Docklands Light Rail (DLR) stations) also have card readers for people who are using Pay as You Go (PAYG) credit on Oyster cards. In these cases, entry and exit information is also collected. Interchange information is only collected at stations where people tap into a reader to show that they did not cross into a certain zone for fare calculation purposes. OR stages from buses only capture the entry record, including the route number, the time, the ticket type, and the sequence number for that card. However, a method was recently developed by Gordon (2012) using TfL's Oyster system and automatic vehicle location (AVL) systems to infer bus trip origins and destinations, and link the separate stages made by an Oyster card user into full, multi-modal journeys that include start and end time and location, trip durations and speeds, and mode shares for specific users (Gordon, 2012). This method is implemented as a Java program that was still under refinement at the time of this research, but could be used in the future to develop origin-destination (O-D) matrices for the panel of survey respondents who have provided Oyster card numbers, on a daily basis to validate and enhance the reported PT stages, provide exact details of start and end times and locations on the travel day, and produce detailed accounts of this panel's travel behavior over time.

A limitation of this research, and the analysis of OR stages in general, is that there are some un-gated train stations on the NR and DLR networks permitting people with a period pass on their Oyster cards to board/alight a train without tapping their card. This can result in occasions where a person may have actually completed a reported journey, but there is no corresponding OR. However, over two thirds of the reported LTDS train journeys were

---

<sup>1</sup>Note there are not eight weeks of Oyster data for the respondents with travel days before 14 August, 2011

made on the Underground where tapping in and out is required and many more were made using gated NR stations. Therefore any instances of higher numbers of reported train stages than train OR stages can only be minimally attributed to un-gated stations.

## 2.2 Intrapersonal Variability in Travel Behavior

As presented in Stopher and Zhang (2011), travel demand modeling has been based on the assumption that individual travel behavior is highly repetitive in the short run, i.e., intrapersonal variability - the variability in each person's day-to-day behavior - is non-existent or very minor. Data collection and modeling procedures have been based on the use of data from diary responses of one day of travel for each respondent assuming that the travel behavior of an individual is repeatable and predictable. It is assumed that any minor deviations from this will be compensated for across the entire sample and that choosing a random sample of households and a random weekday for the diary responses, the resulting data will be representative sample of undertaken journeys by the population. These assumptions were tested in the analysis of panels of data from a study done in Adelaide, South Australia over three years. The results showed that the repetitiveness that underlies all travel demand modeling is not present when one looks at evidence from data on people's travel patterns collected with a Global Positioning System (GPS) device. Some travel patterns were repeated with moderate frequency, like simple work or education tours, but repetition on a daily basis throughout the week did not happen. As a main conclusion they found that the underlying assumption that travel is repetitive from day to day was "highly suspect" (Stopher and Zhang, 2011).

This is because there are some parts of the weekly pattern of trips are dynamic and based on choices that cannot be captured on the day level because they are part of a series of choices made in that week. Three basic reasons for intrapersonal variability were outlined in Block-Schachter (2009):

1. People have weekly budget constraints and may vary their mode used based on their weekly budget.
2. Some behavior is day of the week specific
3. Some people have an innate preference for variability

Block-Schachter (2009) also summarized in his literature review that intrapersonal variability was found to exist, but it is difficult to effectively characterize that variability. He found that there was more variability in activity patterns in a two week sample than a one week sample, but after two weeks there was little additional variability gained by adding days to the sample. His thesis analyzed survey data from MIT employees and found that almost 1 in 5 employees commuting to the same location used different modes on different days in a given week in 2009. This does not imply that variability was random, but it was most likely caused by different activity patterns on different days of the week. These results suggest travel surveys that measure only one or two days of data for urban travel may not provide accurate estimates of the modes used, may overestimate the means and variances of travel such as person miles per day, number of trips per day, time spent traveling per day,

and average time and distance per trip (Stopher et al., 2006a), and that analyzing multiple days of PT travel would provide better estimates of the travel behavior in a region. This, however, may place too high a burden on respondents, so other data collection methods are needed, which are discussed in section 2.4.

## 2.3 Household Travel Surveys

Household travel surveys have been used all over the world for transportation planning purposes. Households are generally sampled across an area, and demographic and trip making by all modes is collected from each member of a household on a given day or set of days. The results are expanded statistically to represent the entire population in an area on an average day to analyze the travel behavior of the region on a large scale. Therefore these results are not always suitable for PT planning for a number of reasons outlined in (Chapleau et al., 2008). First, household surveys are expensive to perform and are usually carried out only once every few years<sup>2</sup>. Additionally, in areas where the private automobile is the predominant mode, the household survey may not capture PT trips. For example, in a diary survey individuals are required to complete a diary of all trips made for one day. If a person uses a bus service once a week, and a diary is being used, then there is only a one in seven chance that the trip will be picked up (Bagchi and White, 2005). Furthermore, many household surveys are not able to provide data that is detailed enough temporally or spatially for PT planning.

Additionally, gathering data that is representative of the target population is becoming more difficult with decreasing response rates, increasing costs due to non-response and phone systems structure, reduction in the completeness of most official lists of residents (number of homes without land lines and difficulty obtaining mobile phone numbers) and increasing number of important questions due to awareness of the complexity of urban life (Trepanier et al., 2009). Household travel surveys have been found to be particularly susceptible to non-response. The surveys are complex and require a two-stage interview process, plus the completion of a travel diary by all household members. Memory decay, failure to understand or to follow survey instructions, unwillingness to report full details of travel, and simple carelessness all contribute to the incomplete collection of travel data (PTV, 2011).

To accurately measure PT service consumption and monitor travel behavior change over time for in depth planning purposes, a large sample with the same respondents over a multi-day time period is desired. A multi-day travel survey would require significantly more resources for the same sample size than a one day sample and has been found to cause response fatigue (Chapleau et al., 2008). There is usually a drop-off in reporting travel on the second and subsequent days of multi-day travel surveys as the respondents become fatigued or find the diary task too burdensome. It has been observed that repetitive trips, such as journeys to and from work or school continue to be reported well on subsequent days, but other trips, especially short trips or walking trips are missing from subsequent travel days. Work in New South Wales, Australia found a drop off in the trip rate from 1.6

---

<sup>2</sup>The LTDS, which is the focus of this thesis project is collected continuously.

trips per household per day or 0.61 trips per person day. Another study in Adelaide found the non-mobility rate increased from 13.4 on the first day to 15.4 percent on the second day (Stopher et al., 2006a).

Household surveys rely on the ability of people to accurately recall the number of trips they make, the origin and destination addresses of their trips, the time at which each trip was made, and the duration and distance of the trip. Unfortunately, people are often unable to provide accurate reports of any of this information (Stopher et al., 2007). With self-reported surveys, there have been many studies that found that people round their estimates of the start and end times to an interviewer or through a self-administered survey. Rounding to the nearest 5 minutes at each end of a trip leads to inaccuracy in the travel time estimation. Rounding to the nearest 15 minutes or even 30 minutes, leads to sufficiently erroneous estimates of travel time that are not useful to the researchers. Stopher and Shen (2011) found that people are more easily able to recall the duration of trips than they were the start and end times. Additionally, people often do not know the addresses many of their destinations. They might know how to recognize the location, or how to get there, but cannot state the street name or number. As a result, that information collected by the household travel survey diaries is often deficient (Stopher et al., 2006b).

## 2.4 Combining Automatically Collected and Manual Survey Data

Travel surveys are an important source of information to describe the typical travel behaviors of a population, but there are difficulties gathering the required data that is desired for modeling purposes as modeling capabilities advance and more in-depth data as well as increased level of detail are desired. At the same time, the expanding availability of technology, such as smart card data, GPS data from many systems including mobile phones and other devices increases the data available for transportation modeling (Trepanier et al., 2009). Therefore the travel survey community has begun exploring new ways to enhance the regular travel datasets.

### 2.4.1 GPS Studies

Many studies have been done to validate the measurement of personal travel behavior, evaluate policies relating to behavioral change (Stopher et al., 2006b), and provide a far more reliable method to investigate daily variability in trip making (Stopher et al., 2006a) than household travel surveys. A review by Bricka and Bhat (2006b) of many telephone household travel surveys using GPS to validate diary responses found that the average size of GPS samples for studies comparing GPS and diary information was 5% of the total sample surveyed households. The levels of trip-underreporting estimates ranged from 10% to 81%. The greatest offenders in terms of the magnitude of trip under-reporting were the heaviest travelers, consistent with prior research on the impact of respondent burden on survey data completeness. Trip underreporting was most closely associated with the following demographic variables: households that own more vehicles (3+), households with incomes of less than \$50,000, and respondents under the age of 25, with the propensity



to underreport decreasing with age. The trip characteristics found to impact trip under-reporting are total trips, trips of short-duration, and trips of discretionary nature.

A study by Stopher et al. (2007) claims to be the first study using GPS to validate face-to-face interviews. It used the Sydney Continuous Household Travel Survey and validation with GPS devices in the first half of 2004 from 59 households and 81 persons generating 465 trips. After matching trip records from each data set, 86% of trips could be matched within 12.5 minutes of the starting times. 7.4% of trips were not recorded by the household survey. People making many trips were more likely to underreport their travel, as were those making trips after 17:00. Shorter trips in both time and distance were more likely to be underreported and those associated with activities of short duration. Males, those under 50, and those with lower incomes were more likely to underreport trips. Both travel time and trip distance were over-reported.

In 2010, the Atlanta Regional Commission (ARC) contracted PTV, NuStats, LLC to conduct the Regional Travel Survey to support the planning data needs of the Atlanta, Georgia region. This study deployed GPS data loggers to Atlanta households to collect detailed information about all trips made by the GPS subsample and estimate levels of trip underreporting in the subsample that could be applied to the larger, non-GPS sample. There were a total of 1,422 GPS vehicles and 649 GPS persons. A total of 9,967 GPS trips were collected compared to 8,711 reported trips for the same vehicles or persons. 45.2% of diary-reported vehicle trips had perfect matching. 5.3% of vehicle diary trips had no corresponding GPS trips. 16.2% of vehicle GPS trips were not reported in the diary. 32.1% of person-trips had perfect matching. 11.7% of person diary trips were not captured by GPS. 14.6% of person-trips captured by GPS were not reported in the diary (PTV, 2011).

Stopher and Shen (2011) compared the GPS and survey diary responses manually for 1,104 trips undertaken in Melbourne in one week. Approximately 53.2% of trips had GPS and diary data that could be compared and 46.8% that couldn't be compared. 6.4% had GPS only, 7.0% had diary only, 29.8% were matching, and 10.0% mismatching. The matched trips had mean difference between start times of 12.7 minutes with standard deviation of 20.3 minutes. They also discussed the reasons for misreporting.

Trips that only appeared in the travel diary (and not the GPS) were thought to be due to:

- Trips may be of short duration or distance, or the stop time between trips may be short where the GPS failed to locate the position before the next trip started,
- The person may have forgotten to take the GPS device with them on a trip, and
- Telescoping – a trip reported on one day may have actually occurred on a different day.

Trips that only appeared in the GPS records (and not the travel diary) may be due to:

- Memory problems of the respondent,

- The respondent may be reluctant to report the location or purpose of some trips,
- GPS may misreport some trips if the GPS device is at rest but is still recording position,
- Respondents may misunderstand the definition of a trip, and
- There may also be delays in the GPS reporting; the satellite may be obstructed.

There are some problems with using GPS as a substitute for more conventional interview or self-administered surveys including biases between those who are willing to undertake the GPS survey and those who are not. Bricka and Bhat (2006a) found that people who self-select to participate in GPS studies were different from those who do not elect to participate. GPS participants tended to report higher incomes and own their own homes. However, there are biases in all household travel surveys that arise from coverage problems (recruiting by telephone which excludes households without phones, or sampling from other lists that exclude renters). Typically standard household surveys under represent one-person and large households, those who travel very little and those who travel a great deal, public transport users, non-car owners, and renters (Stopher et al., 2006b). A study by Stopher et al. (2006b) found that the biases that were potentially present in GPS studies appeared to be similar to the biases in conventional household travel surveys. There were no consistent data that suggested that the samples of households that will undertake GPS surveys are essentially different from the underlying population, nor that they differ from conventional surveys (Stopher et al., 2006b). Additionally, GPS devices do not provide any information directly on purpose or mode, however both of these can be inferred with a very high accuracy from the information provided by the GPS device. These inferences are detailed in Stopher et al. (2006b) and are improving all the time.

Travel diaries and GPS studies both have drawbacks. Being required to enter data before and after each trip is considered a burden, even if it is assisted by an intelligent device. With the analysis of passively obtained trajectories, important attributes can only be statistically estimated. A synthesis of these approaches is the prompted recall survey, which can be done with Internet applications and various forms of mobile phone GPS devices or smart cards which will be described in section 2.4.2.

#### **2.4.2 Mobile Phone Studies**

As described in Cottrill et al. (2012), GPS devices have greatly expanded transportation data collection options much farther than was possible with the household travel surveys alone. Location enabled technologies can reduce the number of erroneous no travel days and missed trips, improve the accuracy of reported trip times, locations and paths, and reduce respondent burden. GPS studies have been widely implemented worldwide, but these studies can only record accurate time and geographic information; trip attributes like purpose and mode must be added or enhanced by the participants. Prompted recall (PR) methods incorporating interaction between the surveyor and survey participant have been found to be more successful because there is less reporting bias when the user does not have to recollect the travel diary from memory, but is presented with an initial suggestion

to enable the reporting of travel behavior.

PR surveys provide respondents who have earlier carried a GPS device with them for a day or more with the information about the travel recorded on that GPS device. They are then asked to provide additional information, such as the mode, journey purpose, and the size of their travel party, as well as correct any errors. Early forms of the PR survey had maps of each day of travel printed and incorporated within a paper survey, but that was found to be rather clumsy (Stopher et al., 2010). The transition from paper to the Internet provides an interactive environment which allows respondents to correct the GPS processed record. The PR survey requires that respondents are familiar with maps and map reading, and they require access to and familiarity with the Internet, which could reduce the proportion of households and household members who could respond to a PR survey over the Internet. However, there is not necessarily a need for the PR survey to be undertaken by a representative sample because the purposes of the PR portion of the survey are generally not to expand the PR results to the entire population, but to check the processing of the GPS data and provide a data source for improving the processing software that can use statistical and geographical information to make inferences about mode and journey purpose (Stopher et al., 2010).

There are some problems with providing GPS devices to the survey subjects for PR or other surveys. Namely, the devices may be subject to loss or damage or the survey participant may forget to carry the GPS device every day. Smart phones help with these problems as they belong to the survey subject and are generally not forgotten by the participant as people are accustomed to carrying their mobile phone with them. Also, smart phones are often equipped with GPS receivers, but can also locate the user by the alternative means of WiFi network signatures and the mobile network, which also works inside buildings (Zilske and Nagel, 2012) and throughout the subways in many places. Due to privacy concerns, location based applications on mobile phones have been limited to personal users and the data has been kept private. Recently, however, individuals have been more willing to share their personal information and locations with others. Many people, especially the young, value their accessibility as a much more important aspect to their social lives than privacy. Voluntary location sharing platforms are also becoming more common. However, it is likely that only certain personalities will participate and may only select data from certain days, but at present this was not found to be a sufficient deterrent to abandon this line of research (Xu, 2011).

One example of a smart phone based PR survey is the Future Mobility Study (FMS). It is an effort undertaken in Singapore as a subset of the nationwide Singaporean Household Interview Travel Survey which is conducted every four to five years. This project aims to support data collection initiatives for transport modeling purposes. The study was conducted in four stages: registration of the household with basic demographic information, a pre-survey with more detailed information about the household including socioeconomic information and vehicle ownership, an activity diary where the participants visited the FMS website to validate activity and mode information recorded and detected from use of the FMS mobile phone application, and an exit survey where participants provided feedback on the survey experience and additional household and preference information.

The application and study are described in detail in Cottrill et al. (2012). The survey required an extensive period of testing and evaluation in order to develop a simple, usable system that was understood by participants, frugal with mobile phone battery use, and useful to practitioners.

The FMS project faced specific concerns related to online surveys including mandatory versus optional question responses, limited question responses, question ordering, and privacy and confidentiality. They provided a “prefer not to answer” option to address the question of mandatory versus optional questions, and provided a privacy policy that clearly stated how collected location data was shared. Cottrill et al. (2012) found that their online instrument must compensate for the lack of trained interviewers to administer it because Internet survey implementation will affect accessibility, compatibility, consistency across respondents, and respondent motivation and experience. In terms of validation details, only highly engaged users added new locations, but users were comfortable deleting wrongly detected locations, leading the researchers to prefer false positives over false negatives (within reason). The pilot study has not resulted in a significant sample yet, but has provided valuable insight into user needs regarding the interface, as well as training data for the background processing intelligence needed for stop and mode detection. The researchers found a need for clear survey workflow and simple user interaction in order to maintain participation rates. Their study demonstrates the capability of smartphone-based travel surveys and the effort needed for successful development. This research area is still developing, but presents an exciting new area of travel demand data collection opportunities that should be considered in the future by transportation planning agencies.

### 2.4.3 Smart Card Studies

Some on-board surveys can fill the gaps of household travel surveys for PT travel analysis, but as smart card fare payment technologies improve and expand, many agencies can use the data from these systems to study the behavior of their passengers at almost no additional cost after the initial capital investment of the system (Chapleau et al., 2008). Smart card fare collection systems are gaining popularity in many public transport systems around the world to manage fare revenues, control fare policies, and reduce fraud. They have also been found to be useful data collectors about the day-to-day variability of user behaviors at a very detailed spatial and temporal resolution (Trepanier et al., 2009). Compared to existing sources of data, including manual surveys of travel behavior, smart card data allows for much larger samples and analysis of behavior over much longer periods of time that can be linked to individual cards and travelers (Bagchi and White, 2005). Smart card technologies allow for this in a way that is passive on the part of the traveler. There is no need for the traveler to recall or record any information, which avoids the possibility of response fatigue (Chapleau et al., 2008). Pelletier et al. (2011) found that the continuously flowing data on passenger behavior can serve to enhance the strategic, tactical, and operational performance of transit authorities. If privacy concerns are overcome and adequate security measures are put in place, planners and researchers will finally have a continuous source of data to enable them better understand transit user behavior, improve the PT system, and increase its role in sustainable transportation (Pelletier et al., 2011).

Some research has been done using data from a small Canadian transit authority to compare various indicators estimated with data from household travel surveys and data from the smart card fare collection systems for the same time period, without matching specific household responses to their respective smart card data. A study, summarized in Trepanier et al. (2009), found that data fusion techniques can allow the linking of travel surveys with smart card data to increase the information available from each source. It also found that results from large household surveys, based on the concept of the average weekday were often diverging from the figures obtained with smart card data. For example, studying individual route information using travel surveys alone seemed to be insufficient to inform on load profiles on all services. For large routes, the survey gave estimations close to those obtained with smart card data. Smart card data provided more details on small movements (for empty cells of household survey based matrix). The Trepanier et al. (2009) study also found that smart card data could be used to improve the quality of the survey expansion process. This thesis goes further than previous studies using smart card data because it includes a larger sample and has household survey data linked with the specific household smart card data, allowing for a more in-depth study of the accuracy and representativeness of a larger sample of survey responses than has previously been possible.

Smart card data cannot be seen as a complete replacement of existing transportation data sources, however. There are some limitations, like the absence of demographic characteristics<sup>3</sup> of the traveler, journey purpose, the ultimate origin and destination, information about non-PT modes of travel, and attitudes towards transport roles that suggest a complementary role may be more appropriate (Bagchi and White, 2005). As described in section 2.1.2 a method was recently developed by Gordon (2012) using TfL's Oyster system and automatic vehicle location (AVL) systems to infer bus trip origins and destinations, and link the separate stages made by an Oyster card user into full, multi-modal journeys that include start and end time and location, trip durations and speeds, and mode shares for specific users (Gordon, 2012). This method is implemented as a Java program that was still under refinement at the time of this research, but could be used in the future to develop origin-destination (O-D) matrices for the panel of survey respondents who have provided Oyster card numbers on daily, monthly or yearly basis (or other time period) to validate and enhance the reported PT stages, provide exact details of start and end times and locations on the travel day, and produce detailed accounts of this panel's travel behavior over time.

## 2.5 Summary of Background and Literature Review

Travel surveys are an important source of information to describe the typical travel behaviors of a population, but there are difficulties gathering the required data that is desired for modeling purposes as modeling capabilities advance and more in-depth data as well as increased level of detail are desired. At the same time, the expanding availability of technology, such as smart card data, GPS data from many systems including mobile phones and other devices increases the data available for transportation modeling (Trepanier et al.,

---

<sup>3</sup>The fare paid might give clues to the age of the passenger if there are discounts for students or the elderly.

2009). Therefore the travel survey community has begun exploring new ways to enhance the regular travel datasets. This chapter provided an overview of the previous work done on this topic. The remaining chapters will describe the work done comparing the LTDS and Oyster card data that analyzes a much larger sample of households (almost 3,000 households with reported public transport travel and/or smart card records over a nine month period) than the scale of most GPS studies, and has household survey data linked with specific smart card data for that household, allowing for a more in-depth study of the accuracy and representativeness of household survey responses than similar studies done in the past.

## Chapter 3

# Results of Combining LTDS and Oyster Data

The main purpose of this thesis is to combine the London Travel Demand Survey (LTDS) with Oyster smart card data to understand the accuracy and representativeness of the survey and determine how to use Oyster data to study the variability in public transport (PT) travel behavior in great detail over a span of time not available with current survey methods.

A total of 3,946 people volunteered to provide up to two of their most frequently used Oyster card numbers (and the associated travel details of those cards) between July 2011 and March 2012 (28% of total respondents with travel days between July 2011 and March 2012). Their LTDS survey responses and Oyster Record (OR) stages were analyzed in varying degrees of detail depending on their reported LTDS PT travel behavior and their PT activity captured by the Oyster system on their travel day and other days within the OR analysis period (which, as described in Chapter 2, is from 19 June, 2011 through 18 June, 2012). The categorization of respondents and a description of the analysis that was conducted for each category is described in this chapter. This is followed by a summary of the difference in magnitude between the overall number of reported LTDS PT stages and OR stages for the sample of people who were interviewed between July 2011 and March 2012, volunteered their Oyster card number(s), reported LTDS PT stages on their travel day using their Oyster card(s), and/or had valid OR stages on their Oyster card(s) on their travel day. This chapter concludes with a comparison of the LTDS reported weekly variation in PT travel and the variation of PT travel captured by the Oyster system.

### 3.1 Categorization of Person Types

This research analyzed the 3,946 (28% of total respondents) people who volunteered their Oyster card number(s) between July 2011 and March 2012. Table 3.1 shows the categorization of people organized into “Person Types” by whether or not they reported LTDS PT stages on their travel day using their Oyster card(s), reported travel on their travel day on non-PT modes or on PT modes using a form of payment other than their Oyster card(s), or

reported no travel on their travel day, and whether or not any valid OR stages were found on the respondent’s travel day or other days within the OR analysis period. At this level of analysis, a valid OR stage was a tap on or off of any mode of PT and does not include adding value to an Oyster card or purchasing any type of period pass. Each “Person Type” is described in more detail below.

**Table 3.1: Person Types for People who Volunteered Oyster Card Number(s)**

Type Number (% of Total)	Have OR on Travel Day	Have No OR on Travel Day, but have OR on Other Days	Have no OR	Total volunteered Oyster (% of Total)
People with PT on Oyster on Travel Day	Type 1 1,148 (29%)	Type 2 292 (7%)	Type 3 117 (3%)	1,557 (39%)
People with Travel, but No PT on Oyster on Travel Day	Type 4 127 (3%)	Type 5 1,474 (37%)	Type 6 224 (6%)	1,825 (46%)
People with No Travel on Travel Day	Type 7 24 (1%)	Type 8 466 (12%)	Type 9 74 (2%)	564 (14%)
Total volunteered Oyster (% of Total)	1,299 (33%)	2,232 (57%)	415 (11%)	3,946

### 3.1.1 Type 1 - PT Users with Similar LTDS and OR Travel

Type 1 people reported LTDS PT stages using their volunteered Oyster card(s) and had valid OR stages on their card(s) on their travel day. For most types of travel, it is unlikely that a person used an Oyster card other than the reported one or two most frequently used cards in a single day. Therefore any discrepancies between reported LTDS PT stages and OR stages of a Type 1 person were assumed to be due to the person not making the same journeys that he or she reported or traveling between two un-gated National Rail (NR) or Docklands Light Rail (DLR) stations using a period pass where tapping in and out was not required. Type 1 people are analyzed in detail in Chapter 4.

### 3.1.2 Type 2 - PT Users with Different LTDS and OR Travel

Type 2 people reported LTDS PT stages using their volunteered Oyster card(s), but did not have valid OR stages on their card(s) on their travel day. They did, however, have OR stages on their card(s) on other days within the OR analysis period, showing that they sometimes used their reported Oyster card(s). Given the length of the interview process, it is unlikely that a person took the time to describe a journey he or she has never taken, so Type 2 people likely used a different Oyster card on their travel day or reported a day other than their travel day that they may have believed was more representative of their



normal travel behavior. It is also possible that they only traveled between two un-gated NR or DLR stations using a period pass where they were not required to tap their card. Type 2 people are analyzed in detail in section 5.1.

### **3.1.3 Type 3 - PT Users with LTDS but No OR Travel**

Type 3 people reported LTDS PT stages using their volunteered Oyster card(s), but did not have valid OR stages on their card(s) on their travel day or any other days within the OR analysis period. It is unlikely that a person took the time to describe a journey he or she has never taken, so Type 3 people likely used a different card or fare media on their travel day, or the Oyster card number was transcribed incorrectly. Analysis of this type of person is limited to demographic and travel behavior information. Type 3 people are discussed in section 5.2.

### **3.1.4 Type 4 - PT Unreported on LTDS with PT OR on Travel Day**

Type 4 people reported other travel besides using their volunteered Oyster card(s) on PT, but had valid OR stages on their card(s) on their travel day. It is possible Type 4 people reported a PT stage, but said they used a magnetic ticket or cash when they actually used their Oyster card(s), or that someone else in their household used their card(s) without reporting sharing Oyster card(s). However, there is an advantage for people to underreport journeys to shorten the length of their interviews, so the OR stages found for these people are most likely journeys they made but did not report. Type 4 people are analyzed in detail in section 5.4.

### **3.1.5 Type 5 - PT Unreported on LTDS with Different OR Travel**

Type 5 people reported other travel besides using their volunteered Oyster card(s) on PT, and did not have any valid OR on their card(s) on their travel day. They had OR stages on their card(s) on other days within the OR analysis period. Analysis of these people's OR stages can show the frequency of PT use captured by the Oyster system over the OR analysis period compared to the reported frequency of PT use in the LTDS. Type 5 people are analyzed in detail in section 5.5.

### **3.1.6 Type 6 - PT Unreported on LTDS with No OR Travel**

Type 6 people reported other travel besides using their volunteered Oyster card(s) on PT, and did not have valid OR stages on their card(s) on their travel day or any other days within the OR analysis period. Analysis of this type of person is limited to demographic and travel behavior information. Type 6 people are discussed in section 5.3.

### **3.1.7 Type 7 - No Travel on LTDS with PT OR on Travel Day**

Type 7 people did not report any travel on their travel day, but had valid OR stages on their volunteered card(s) on their travel day. Someone else in their household could have used their Oyster card(s) that day without reporting sharing Oyster card(s), but there is an advantage for people to underreport journeys to shorten the length of their interviews, so the OR stages found for these people are most likely journeys they made but did not report. Type 7 people are analyzed in detail in section 5.4.

### **3.1.8 Type 8 - No Travel on LTDS with Different PT OR Travel**

Type 8 people did not report any travel on their travel day and did not have any valid OR stages on their volunteered card(s) on their travel day. They had OR stages on their card(s) on other days within the OR analysis period. Analysis of these people's OR stages can show the frequency of PT use captured by the Oyster system over the OR analysis period compared to the frequency of PT use reported in the LTDS. Type 8 people are analyzed in detail in section 5.5.

### **3.1.9 Type 9 - No Travel in LTDS or OR**

Type 9 people did not report any travel on their travel day and did not have any valid OR stages on their volunteered card(s) on their travel day or any other days within the OR analysis period. Analysis of this type of person is limited to demographic and travel behavior information. Type 9 people are discussed in section 5.3.

A limitation of this research is that there is no guarantee that people actually used their volunteered Oyster card(s) on their travel day or that Oyster card numbers were transcribed correctly during the interview process. This made it difficult to determine if differences in reported LTDS PT stages and OR stages were because the person did not make that journey on that day, or because the person used a different card or fare media than reported. However, almost three quarters of the people who volunteered their Oyster card(s) and reported PT travel using their Oyster card(s) on their travel day had OR stages on their volunteered card(s) on that day (Type 1). This makes it unlikely that they also used another Oyster card other than the one or two most frequently used cards that they reported in the survey.

For these reasons, Type 1 people were assumed to have used their volunteered Oyster card(s) on their travel day and any discrepancies of these people can be assumed to be due to not making the same journeys that they reported, or not reporting journeys that they made. Some discrepancies may also be due to the person traveling between two un-gated NR or DLR stations using a period pass where they were not required to tap their card. With the other person types (Types 2 - 9), assumptions were made about whether the discrepancies were due to the individual misreporting travel as opposed to misreporting the card number or fare media used. The majority of analysis in this report is therefore focused on Type 1 people, which will be discussed briefly in the following sections and described in more detail

in Chapter 4. There were some conclusions that were drawn by examining the specific reported LTDS stages and OR stages (in terms of bus routes or origin-destination (O-D) pairs) for the other types of people, which are also summarized in the following sections and described in detail in Chapter 5.

### 3.2 Difference in Overall Magnitude of LTDS and OR Stages

A total of 1,708 LTDS respondents (interviewed between July 2011 and March 2012) from Person Types 1, 2, 3, 4, and 7 volunteered their Oyster card number(s), reported LTDS PT stages on their travel day using their Oyster card(s) and/or had OR stages on their card(s) on their travel day. These respondents are referred to as the PT Sample. Table 3.2 shows the comparison of total reported LTDS PT stages and OR stages by “Person Type” by whether or not the person reported LTDS PT stages *and* had OR stages (Type 1), had reported LTDS PT stages but no OR stages (Types 2 and 3), or had OR stages but no reported LTDS PT stages (Types 4 and 7).

**Table 3.2:** Total Difference between Reported LTDS PT Stages and/or OR Stages

Type	People	% of PT Sample	LTDS Stages <sup>1</sup>	OR Stages	LTDS - OR	% Difference: $\frac{LTDS - OR}{LTDS}$
Total PT Sample	1,708	100%	4,454	4,067	387	8.7%
1 Have LTDS and OR Stages	1,148	67%	3,338	3,634	-296	-8.9%
1a People with same number of LTDS and OR Stages	666	39%	1,719	1,719	0	0%
1b People with more LTDS than OR Stages (overreported)	191	11%	712	426	286	40.2%
1c People with fewer LTDS than OR Stages (underreported)	291	17%	907	1,489	-582	-64.2%
2 Have LTDS Stages, but no OR Stages (overreported)	292	17%	765	0	765	
3 Have LTDS Stages, but no OR Stages (overreported)	117	7%	351	0	351	
4+7 Have OR Stages, but no LTDS Stages (underreported)	151	9%	0	433	-433	

<sup>1</sup>Includes only PT stages

In the PT Sample there were approximately 8.7% more LTDS PT stages than OR stages. Only 39% of the PT Sample had the same number of OR stages captured by the Oyster system on their volunteered Oyster card(s) as reported in the LTDS (Type 1.a). The remaining 61% either reported more PT stages than stages captured by the Oyster system, or had more OR stages than reported in the LTDS. Approximately 35% of the PT Sample had more LTDS PT stages than stages captured on their volunteered Oyster card(s) (overreported stages from Types 1.b, 2, and 3), and approximately 26% of the PT Sample had fewer LTDS PT stages than stages captured on their volunteered Oyster card(s) (underreported from Types 1.c, 4, and 7).

As described above, Type 1 people were assumed to have used their volunteered Oyster card(s) on their travel day and any discrepancies of these people were assumed to be due to not making the same journeys that they reported, or not reporting journeys that they made, as opposed to misreporting the fare media used. Therefore, analyzing the differences in reported LTDS PT stages and OR stages for Type 1 people was assumed to be the most accurate way to validate the PT reporting in the LTDS. Table 3.2 shows that 58% of Type 1 people reported the same number of PT stages as OR stages, while the other 42% either reported more LTDS stages than found in their OR stages or had more OR stages than reported LTDS stages. About 25% of Type 1 people underreported LTDS PT stages, and 17% overreported LTDS PT stages, which at an aggregate level, averages out to 8.9% fewer LTDS PT stages than OR stages. Therefore, the people who can be assumed to have used their volunteered Oyster card(s) on the day they reported PT use underreported their single day PT journeys by at least 8.9% in overall magnitude, which is calculated by averaging the difference in trips reported by people who have both overreported and underreported.

The other types of people misreported something about their travel behavior. By examining the specific reported LTDS stages, in terms of bus routes or train O-D pairs without corresponding OR stages (Types 2 and 3 well as Type 1 OR stages without corresponding LTDS stages) and the OR stages without reported LTDS stages (Types 4 and 7 OR stages as well as Type 1 OR stages without corresponding LTDS stages), and making assumptions about the reasons for these discrepancies, the misreporting of the fare media used (which was assumed to be of minor significance for TfL's planning efforts) was separated from the misreporting of O-D pairs or bus routes, and the overreporting and underreporting of PT journeys. This analysis of specific reported PT stages and OR stages for each person is described in Section 3.3. Sections 3.2.1 and 3.2.2 describe the difference in the overall magnitude of LTDS and OR stages by mode and the mode share differences, respectively.

### **3.2.1 Difference in Overall Magnitude of LTDS and OR Stages by Mode**

In order to understand the modal differences between the LTDS PT stages and the OR stages captured by the Oyster system, the analysis was separated by bus and train. Journeys on the Underground, NR, DLR, Tram, and Overground were examined together because they are coded the same way in the Oyster system, and many of these modes share stations on the network making it very difficult to distinguish between train modes in the records captured by the Oyster system. The results are described in the following two sections,

followed by a summary of the overall difference of reported LTDS PT mode share and OR mode share.

### Bus Reporting Differences

A total of 1,265 people from Person Types 1, 2, 3, 4, and 7 reported bus stages using their Oyster card(s) and/or had bus OR stages on their Oyster card(s) on their travel day (Bus Sample). Table 3.3 summarizes the comparison of total reported LTDS bus stages and OR bus stages by “Person Type” by whether or not the person reported LTDS bus stages and/or had OR bus stages.

**Table 3.3:** Total Difference between Reported LTDS Bus Stages and/or OR Bus Stages

Type	People	% of Bus Sample	LTDS stages	OR stages	LTDS - OR	% Difference: $\frac{LTDS - OR}{LTDS}$
Bus Sample	1,265	100%	2,574	2,492	82	3.2%
1 Have Reported Stages and/or OR	859	68%	1,905	2,203	-298	-15.6%
1a People with same number of reported Stages and OR	453	36%	955	955	0	0%
1b.1 People with more reported Stages than OR (overreported)	111	9%	368	222	146	39.7%
1b.2 Have Reported Stages, but no OR (overreported)	34	3%	52	0	52	
1c.1 People with fewer reported Stages than OR (underreported)	202	16%	530	922	-392	-74.0%
1c.2 Have OR, but no Reported Stages (underreported)	59	5%	0	104	-104	
2 Have Reported Stages, but no OR (overreported)	204	16%	465	0	465	
3 Have Reported Stages, but no OR (overreported)	85	7%	204	0	204	
4+7 Have OR, but no Reported Stages (underreported)	117	9%	0	289	-289	

In the Bus Sample there were approximately 3.2% more LTDS bus stages than OR bus stages. Only 36% of the Bus Sample had the same number of bus stages captured by the Oyster system on their volunteered Oyster card(s) as reported in the LTDS (Type 1.a). The other 64% either reported more LTDS bus stages than bus OR stages, or had more bus OR stages than bus stages reported in the LTDS. Approximately 34% of the Bus Sample had more LTDS bus stages than stages captured on their volunteered Oyster card(s) on bus (overreported stages from Types 1.b, 2, and 3), and approximately 30% of the Bus Sample had fewer LTDS bus stages than stages captured on their volunteered Oyster card(s) on bus (underreported from Types 1.c, 4, and 7).

As described above, Type 1 people were assumed to have used their volunteered Oyster card(s) on their travel day and any discrepancies of these people were assumed to be due to not making the same journeys that they reported, or not reporting journeys that they made, as opposed to misreporting of the fare media used. Therefore, analyzing the differences in reported LTDS bus stages and OR bus stages for Type 1 people was assumed to be the most accurate way to validate the bus reporting in the LTDS. Table 3.3 shows that 53% of Type 1 people reported the same number of LTDS bus stages as OR bus stages, while the other 47% reported more LTDS bus stages than found in their OR bus stages or had more OR bus stages than reported LTDS bus stages. About 30% of Type 1 people underreported LTDS bus stages, and 17% overreported LTDS bus stages. Therefore, the people who can be confidently assumed to have used their volunteered Oyster card(s) on the day they reported bus use underreported their single day bus journeys by at least 15.6% in overall magnitude, which is calculated by averaging the difference in trips reported by people who have both overreported and underreported.

### **Train Differences**

A total of 911 people from Person Types 1, 2, 3, 4, and 7 reported train stages using their Oyster card(s) and/or had train OR stages on their Oyster card(s) on their travel day (Train Sample). In the Train Sample there were approximately 16.2% more LTDS train stages than OR train stages. Table 3.4 summarizes the comparison of total reported LTDS train stages and OR train stages by "Person Type" by whether or not the person reported LTDS train stages and/or had OR train stages. Approximately 49% of the Train Sample had the same number of train stages captured by the Oyster system on their volunteered Oyster card(s) as train stages reported in the LTDS (Type 1.a). The other 51% either reported more train LTDS stages than train stages captured by the Oyster system, or had more train OR stages than train stages reported in the LTDS. Approximately 33% of the Train Sample had more LTDS train stages than stages captured on their volunteered Oyster card(s) on train (overreported stages from Types 1.b, 2, and 3), and approximately 18% of the Train Sample had fewer LTDS train stages than stages captured on their volunteered Oyster card(s) on train (underreported from Types 1.c, 4, and 7).

**Table 3.4:** Total Difference between Reported LTDS Train Stages and/or OR Train Stages

Type	People	% of Train Sample	LTDS stages	OR stages	LTDS - OR	% Difference: $\frac{LTDS - OR}{LTDS}$
Train Sample	911	100%	1,880	1,575	305	16.2%
1 Have Reported Stages and/or OR	656	72%	1,433	1,431	2	0.1%
1a People with same number of reported Stages and OR	446	49%	967	967	0	0.0%
1b.1 People with more reported Stages than OR (overreported)	77	8%	249	134	115	46.2%
1b.2 Have Reported Stages, but no OR (overreported)	33	4%	54	0	54	
1c.1 People with fewer reported Stages than OR (underreported)	70	8%	163	283	-120	-73.6%
1c.2 Have OR, but no Reported Stages (underreported)	30	3%	0	47	-47	
2 Have Reported Stages, but no OR (overreported)	128	14%	300	0	300	
3 Have Reported Stages, but no OR (overreported)	62	7%	147	0	147	
4+7 Have OR, but no Reported Stages (underreported)	65	7%	0	144	-144	

As described above, Type 1 people were assumed to have used their volunteered Oyster card(s) on their travel day and any discrepancies of these people can be assumed to be due to not making the same journeys that they reported, or not reporting journeys that they made, as opposed to misreporting of the fare media used. Therefore, analyzing the differences in reported LTDS train stages and OR train stages for Type 1 people is assumed to be the most accurate way to validate the train reporting in the LTDS. Table 3.4 shows that 68% of Type 1 people reported the same number of LTDS train stages as OR train stages, with the other 32% reporting more LTDS train stages than found in their OR train stages or having more OR train stages than reported LTDS train stages. About 15% of Type 1 people underreported LTDS train stages, and 17% overreported LTDS PT stages

for a difference of only two more LTDS train stages than OR train stages. Therefore, of the Type 1 people, the magnitude of overreporting and underreporting is almost the same (only 0.1% more LTDS than OR train stages).

This shows that overall people reported train journeys more accurately than bus journeys because more Type 1 people had the same number of reported LTDS and OR train stages than bus stages, and the overall magnitude of LTDS and OR train stages were almost exactly the same while the bus stages were underreported by 15.6%.

### 3.2.2 Mode Share

The previous two sections describe the differences between the LTDS and OR stages for bus and train modes. However, many people had both bus and train stages on their travel day. In order to determine the effects of the overall bus underreporting per person, the average reported LTDS mode share between bus and train<sup>2</sup> was compared to the average mode share per person found in OR stages with the results shown in Table 3.5. The average LTDS reported bus share per person for the entire PT sample was 1.5% lower than the OR bus share and the average reported bus share for the Type 1 people was 1.7% lower than the OR bus share. These results combined with the analysis above shows that overall bus stages were slightly underreported compared to train stages.

**Table 3.5:** Summary of Bus and Train Mode Share

	Reported Bus Mode Share	Reported Train Mode Share	OR Bus Mode Share	OR Train Mode Share
PT Sample	58.2%	41.8%	59.7%	40.3%
Type 1	57.0%	43.0%	58.7%	41.3%

### 3.3 Difference between Specific LTDS and OR Stages

Type 1 people were assumed to have used their volunteered Oyster card(s) on their travel day and any discrepancies of these people were generally assumed to be due to not making the same journeys that they reported, or not reporting journeys that they made, as opposed to misreporting the fare media used. There were, however, some cases where Type 1 people misreported the fare media, which will be described below. The other types of people misreported something about their travel behavior that could be misreporting entire journeys or just the card or fare media used. By examining the specific reported LTDS stages, in terms of bus routes or train O-D pairs without corresponding OR stages<sup>3</sup> and the OR stages without reported LTDS stages,<sup>4</sup> and making assumptions about the reasons for these discrepancies, the misreporting of the fare media used (which was assumed to be of

<sup>2</sup>Train includes all journeys on the Underground, NR, DLR, Tram, and Overground

<sup>3</sup>Types 2 and 3 as well as Type 1 people had OR stages without corresponding LTDS stages

<sup>4</sup>Types 4 and 7 as well as Type 1 people had OR stages without corresponding LTDS stages



minor significance to TfL's planning efforts) was separated from the misreporting of O-D pairs or bus routes, and the overreporting and underreporting of PT journeys.

There were 376 people (33% of Type 1 people) with perfectly matching reported LTDS PT stages and OR stages, that is, all of the PT journey stages reported in the LTDS had matching OR stages in terms of bus routes and/or train O-D pairs in the same chronological order, and one additional person that had the same LTDS PT and OR stages, but out of sequence. The remaining 1,331 people in the PT sample had LTDS PT stages without corresponding OR stages (possible overreporting) or OR stages without corresponding LTDS PT stages (possible underreporting). It was possible for a person to under, over, and/or misreport some or all of their LTDS PT stages if they underreported a bus stage and overreported a train stage, etc. The possible categories of mis, under, and overreporting are listed and explained with examples below (where matching refers to matching a specific bus route or train O-D pair):

- **Perfectly Matching:** The person had a matching OR stage for each LTDS reported PT stage, and had an LTDS PT stage for each captured OR stage. For example, a person only reported one bus journey on route 62 followed by one bus journey on route 5 in the LTDS and had one bus OR on route 62 followed by one bus OR on route 5. The person had no LTDS stages without corresponding OR stages and no OR stages without LTDS stages. This category is a subset of Type 1 people.
- **Misreporting Fare Media:** The person misreported the fare media used on his travel day. There were three different situations where this was observed from Person Types 1, 2, 3, 4, and 7.
  - The person had an OR stage without a corresponding LTDS PT stage that was reported using an Oyster card, but had the same reported LTDS PT stage that was reported using a magnetic ticket or cash. For example, the person had an OR stage on bus route 4 and reported paying cash for an LTDS bus stage on route 4, and had no LTDS stages on bus route 4 where he reported paying with his Oyster card. This type of misreporting was observed with Person Types 1 and 4.
  - The person's Oyster card had unreported OR stages that matched reported LTDS PT stages of another member of the same household with missing OR stages (without reporting a shared card). For example, in a household consisting of two people where both reported Oyster card numbers, one person reported taking bus route 33 on his travel day using his Oyster card, but had no OR stages on that day on that card. The other person reported no LTDS PT stages, but had an OR stage on bus route 33 on her card on the household travel day, showing that he likely used her card on the travel day. This type of misreporting was observed with Person Types 4 and 7.
  - The person reported LTDS PT stages that had no corresponding OR stages on his travel day or any days within the previous week of his travel day. For example, the person reported taking bus route 4 on his travel day, but had no OR stages on buses on his travel day (that could have been a misreported bus route) and no OR stages on bus route 4 within the week prior to his travel day. It is unlikely that he never takes bus route 4, so he most likely used a different card on his

travel day. This type of misreporting was observed with Person Types 1, 2, and 3.

- **Misreporting Portion of Origin-Destination:** The person misreported a stage as a different bus route or train O-D pair. For example, a person only reported an LTDS stage on bus route 4, but did not have an OR stage on bus route 4 on his travel day, but had an OR stage on bus route 56 instead. This type of misreporting was only observed with Type 1 people.
- **Overreporting:** The person reported LTDS PT stages that had no corresponding OR stages and likely reported a day other than his travel day that he believed was more representative of his normal travel behavior. This was determined if there were some or all of his reported LTDS PT stages without corresponding OR stages on the travel day in their OR stages within the previous week of his travel day. For example, a person reported an LTDS stage on bus route 33, had no OR stages on bus route 33 on his travel day, had no other OR bus stages without corresponding LTDS stages on his travel day (that could have been misreported bus routes), but had a bus stage(s) on bus route 33 within the previous week of his travel day. Overreporting was observed with Person Types 1 and 2.
- **Underreporting:** A person had OR stages without corresponding LTDS stages who has likely underreported stages because he had no other LTDS stages without corresponding OR stages that could have been misreported. For example, if a person had an OR stage on a bus route that was not reported in the LTDS, and reported no bus stages without corresponding OR stages, he is considered to have underreported the bus stage. Underreporting was observed with Person Types 1 and 4.

Table 3.6 shows the results of the categorization of specific PT stages. Approximately 59% of Type 1 people and all Types 2 and 3 people (or approximately 63% of the PT Sample) reported LTDS PT stages on their travel day that were not found in the Oyster system in terms of the specific bus route or train O-D pair. When combining the specific journeys from Types 1, 2, and 3 that were most likely overreported, there were approximately 11.6% of the PT Sample reported LTDS stages that were overreported, an additional 21.1% that were misreported in terms of the specific bus route or train O-D pair, and 21.9% stages that were misreported in terms of the fare media or payment used or were between two un-gated stations that did not require a tap. This is a lower bound on the overreported stages because all Type 3 people are assumed to only have misreported the fare media, while it is likely given the results of the other person types, that they also overreported some of their PT stages as well. This analysis is described in more detail in section 4.4 for Type 1 people, section 5.1 for Type 2 people, and section 5.2 for Type 3 people.

Approximately 60% of Type 1 people and all Types 4 and 7 people (approximately 49% of the PT Sample) had OR stages on their travel day that were not reported in the LTDS in terms of the specific bus route or train O-D pair. When combining the specific journeys from Types 1, 4, and 7 that were most likely underreported there were approximately 20.9% of the PT Sample reported LTDS stages that were underreported on the travel day, an additional 21.1% that were misreported in terms of the bus route or train O-D pair, and 4.2% that were misreported in terms of the fare media or payment used. This analysis is

described in more detail in section 4.3 for Type 1 people and section 5.4 for Types 4 and 7 people. Table 3.6 summarizes the over, under, and misreporting of specific PT stages for Types 1, 2, 3, 4, and 7 people.

**Table 3.6:** Summary of Over, Under, or Misreporting of Specific PT Stages

	Sample Description	Type 1	Type 2	Type 3	Types 4 and 7	Total
Perfectly Matching	People (% PT Sample)	377 (22%)				377 (22%)
	Stages (% LTDS Stages)	818 (18.4%)				818 (18.4%)
Misreporting Fare Media		138 (8%)	160 (9%)	117 (7%)	47 (3%)	462 (27%)
		210 (4.7%)	427 (9.6%)	351 (7.9%)	168 (3.8%)	1,156 (26.0%)
Misreporting Portion of O-D		562 (33%)				562 (33%)
		942 (21.1%)				942 (21.1%)
Overreporting (likely misreporting travel day)		111 (6%)	132 (8%)			243 (14%)
		177 (4.0%)	338 (7.6%)			515 (11.6%)
Underreporting		335 (20%)			104 (6%)	439 (26%)
		666 (15.0%)			265 (5.9%)	931 (20.9%)

Only 22% of the PT Sample had perfectly matching reported LTDS PT stages and OR stages, which only accounted for 18.4% of the PT Sample's reported LTDS stages. The remaining 78% of the PT Sample either underreported, overreported, or misreported (or some combination of under, over, and/or misreporting) their PT stages. As shown in Table 3.6, approximately 27% of the PT Sample have likely misreported the fare media they used on their travel day, which made up 26.0% of the PT Sample's reported LTDS stages. This is not considered a serious problem, as the reporting of the type of fare media used may have specific implications at TfL, but does not affect the overall level of PT use reported in the LTDS. About 33% of the PT Sample have likely misreported some portion(s) of the train O-D pair or bus route(s) used, which made up approximately 21.2% of the PT Sample's reported LTDS stages. This will have implications for the LTDS survey, if the specific origins and destinations or bus routes reported are being used for planning purposes. About 14% of the PT Sample overreported their PT stages, most likely in terms of the travel day, that is they reported a day other than their agreed upon travel day they may have thought was more representative of their regular travel behavior, which made up approximately 11.6%

of the PT Sample's reported LTDS stages. An additional 26% underreported their PT stages, which made up approximately 20.9% of the PT Sample's reported LTDS stages. Only 51.1% of OR stages had matching LTDS stages, and only 45.6% of LTDS stages had matching OR stages, in terms of the train O-D pairs or bus routes used. The over and underreporting of PT stages impacts the overall level of PT use reported in the LTDS. When taking into account the under and overreported PT stages, an additional 416 stages (931 Underreported - 515 Overreported) or 9.3%, should be added to the PT Sample reported LTDS PT stages.

Table 3.6 shows the over, under, or misreporting of PT stages overall and is not separated into categories by the people who have only misreported the fare media, only misreported the O-D pair, only overreported, or only underreported. The perfectly matching people are obviously mutually exclusive of any other category. Table 3.7 has people separated into specific mutually exclusive categories. As stated above, the misreporting of fare media is not considered a major concern, so people who have only misreported the fare media used on their travel day are combined here with the perfectly matching people. People who have some combination of over, under, or misreporting portions of their train O-D's or bus routes are combined into a category that is mutually exclusive from other categories.

When separating people into mutually exclusive categories, the people who had perfectly matching LTDS and OR stages or only misreported the fare media used on their travel day made up only 44% of the PT Sample and 41.3% of the PT Sample's reported LTDS stages. The other 56% of people over, under, or misreported O-D pairs or bus routes of their 58.7% of stages, which has significant implications about the accuracy of the LTDS if more than half of the people who reported PT travel in the survey<sup>5</sup> are incorrectly reporting the number of PT stages or the origins, destinations, or bus routes used on their travel day.

---

<sup>5</sup>Only Types 1, 2, and 3 reported PT stages in the survey and 54% of them misreported their trips (332 + 258 + 96 + 31 + 132/1,557 Total Types 1, 2, and 3 people)

**Table 3.7:** Summary of Over and Underreporting of Specific PT Stages, Exclusive People

	Sample Description	Type 1	Type 2	Type 3	Types 4 and 7	Total
Perfectly Matching or Only Misreporting Fare Media	People (% PT Sample)	431 (25%)	160 (9%)	117 (7%)	47 (3%)	755 (44%)
	Stages (% LTDS Stages)	895 (20.1%)	427 (9.6%)	351 (7.9%)	168 (3.8%)	1,841 (41.3%)
Combo of Misreporting Portion of O-D, Under or Overreporting		332 (19%)			332	332 (19%)
		1,279 (28.7%)				1,279 (28.7%)
Only Misreporting Portion of O-D		258 (15%)				258 (15%)
		407 (9.1%)			407	407 (9.1%)
Only Underreporting		96 (6%)			104 (6%)	200 (12%)
		164 (3.7%)			265 (5.7%)	429 (9.6%)
Only Overreporting (likely misreporting travel day)		31 (2%)	132 (8%)			163 (10%)
		48 (1.1%)	338 (7.6%)			386 (8.7%)

### 3.4 Weekly Variation in PT Travel

The majority of analyses in this thesis compared the reported LTDS PT stages and OR stages for respondents on their travel day. However, the continuous collection of OR stages for all people who volunteered an Oyster card number allowed for the analysis of OR stages on other days in the OR analysis period to show the weekly variation in personal travel for LTDS respondents. OR stages for each volunteered Oyster card number were collected and stored for all days (including weekends) eight weeks prior to the individual's travel day and indefinitely, though this thesis only analyzed OR stages through 18 June, 2012. Respondents ranged from having no days of OR stages to having OR stages on almost 100% of possible days of OR stage collection. To account for the differences in the respondents and normalize the number of OR stages analyzed for each person, OR stages for each person who volunteered an Oyster card number and had OR stages were analyzed for two weeks prior and two weeks after his or her travel day. Approximately 88% of respondents who volunteered an Oyster card and had OR stages on any PT mode within the OR analysis period (19 June, 2011 - 18 June, 2012) had OR stages within the two weeks prior and two weeks after the travel day.

The frequency of use of each PT mode is reported in the LTDS. These responses are shown in Figure 3-1 for each Person Type. The maximum frequency of all PT modes<sup>6</sup> was used in this analysis. For example, if a person reported using the bus five times per week but the Underground three times per week and all other modes less than once a month, their response was categorized here as using PT five times per week. This figure shows that people who reported LTDS PT stages on their travel day (Types 1, 2, or 3) or had OR stages on their travel day (Types 4 and 7) overwhelmingly reported taking PT five days per week.

Figure 3-2 shows the OR frequency of use for two weeks prior and two weeks after the travel day for each person separated by the Person Types described in section 3.1. The frequency categories shown are the same categories that the LTDS respondents use to answer how frequently they use each mode of transportation. This figure shows that a higher percentage of the people who had OR stages on their travel day (Types 1, 4, and 7) had OR stages three or more days per week in the two weeks prior and two weeks after their travel day. It also shows that Type 1 people (reported LTDS PT stages *and* had OR stages on their travel day) are the most frequent users of PT with over 72% of Type 1 people having OR stages three or more days per week. Types 4 and 7 people did not report PT stages on their travel day, but had OR stages on their travel day and have most likely underreported their PT travel. This figure shows that they were also frequent PT users with 61% and 54% of Type 4 and 7 people, respectively, having OR stages three or more days per week. Type 2 people reported LTDS PT stages on their travel day, but had no OR stages on that day. Their weekly frequency of use had a much different distribution than Type 1 people (who also reported LTDS PT stages on their travel day) with only about 29% of Type 2 people having OR stages three or more days per week for the two weeks prior and two weeks after their travel day. This shows that having OR stages on the travel day is a better predictor of actually being a frequent PT user than reporting PT use on the travel day in the LTDS or, in other words, the Types 2 or 3 people not only overreported PT trips on their travel day, but generally overestimate their actual frequency of PT use by a substantial amount. It is clear that the reported frequency of PT use is much higher than the actual use as captured by the Oyster system, and therefore the LTDS is overestimating the overall PT use for London residents.

---

<sup>6</sup>PT modes are bus, Underground, NR, DLR, Overground, and Tram.

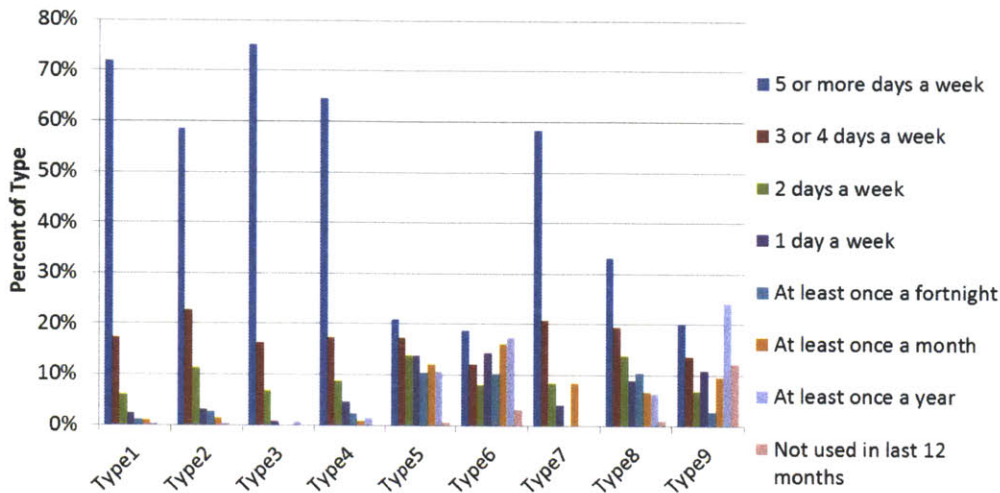


Figure 3-1: Reported Frequency of PT Use (Max of All PT Modes)

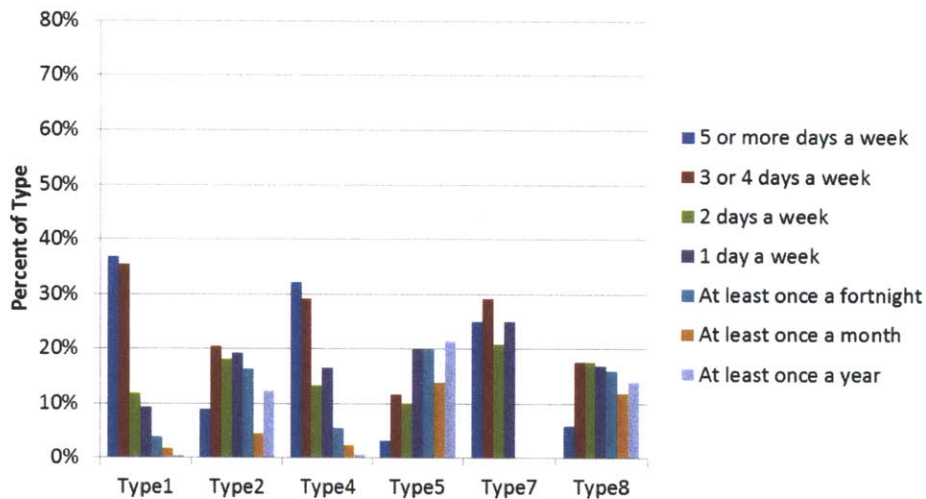


Figure 3-2: OR Stage Frequency Two Weeks Before and After Travel Date

Another benefit of collecting OR stages over time is the ability to analyze PT behavior of people who did not report PT stages on their travel day. For example, Types 5 and 8 people did not report PT stages on their travel days, but had OR stages on other days in the OR analysis period. Figure 3-2 shows that these people were infrequent users of PT with 55% and 42% of Type 5 and 8 people, respectively, having OR stages less than weekly. Some OR stages could be analyzed to get an idea of typical journeys these people took that were not on their travel days. Even without analyzing specific journeys, the OR stages can show a more accurate picture of how frequently these person types, as well as all other types, used the PT system.

Respondents were also categorized into the number of PT stages they reported on their travel day, which ranged from zero to twelve PT stages. For each person, every day of OR stage collection two weeks prior and two weeks after the travel day was categorized by number of OR stages on that day. Figure 3-3 shows the percent of person-days (over the period two weeks prior and two weeks after the travel day) that people who reported a certain number of PT stages on their travel day had each number of OR stages<sup>7</sup>. For example, as shown in Figure 3-3.a, for the people who reported zero PT stages on their travel day, almost 80% of their person-days of OR collection had zero OR stages. These distributions generally show that as the number of PT stages on the travel day increases, the percent of person-days that had zero OR stages decreases.

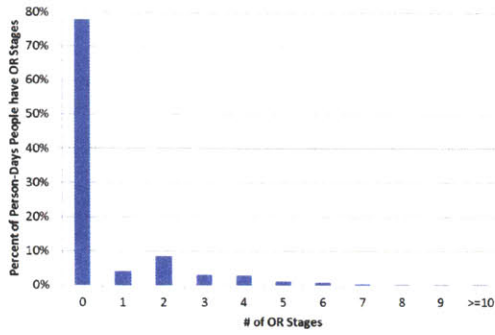
After zero OR stages, two OR stages had the highest percentage of person-days for each category of person that reported four or fewer PT stages on their travel day. The people who reported five or six PT stages on their travel day had slightly higher percentage of person-days with four OR stages. This suggests that the people who reported higher numbers of PT stages on their travel day also used PT more intensely than people who reported fewer PT stages on their travel day on days other than their travel day.

One of the most important findings from analyzing the weekly variation in PT travel is that people who reported one or more PT stages on their travel day had from 38-57% of their remaining person-days (over the period two weeks prior and two weeks after the travel day) with zero PT stages. This suggests that the people who are considered PT users in the survey are only using PT about half of the time, and if the one-day survey just so happens to assign a travel day on a day where the person used PT, it would overestimate their typical PT use. A two or three day survey would better capture the variability in PT use for survey respondents. Alternatively, the use of OR's and Gordon's recently developed method to link the separate stages made by an Oyster card user into full, multi-modal journeys could supplement the one-day survey and provide a much better estimate of the variability of PT travel than the frequency questions asked in the survey or fixed trip rates used in travel demand models.

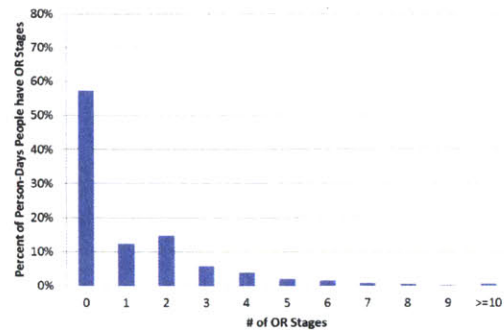
---

<sup>7</sup>Only people who reported six or fewer PT stages on their travel day are shown here because they made up approximately 99% of the people who reported LTDS PT stages on their travel day.

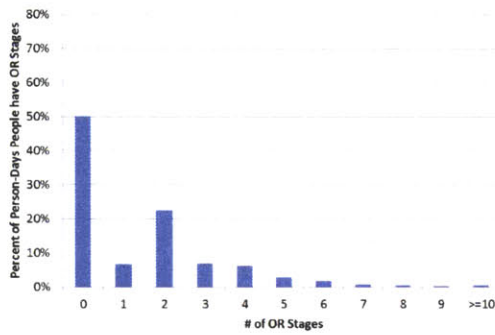




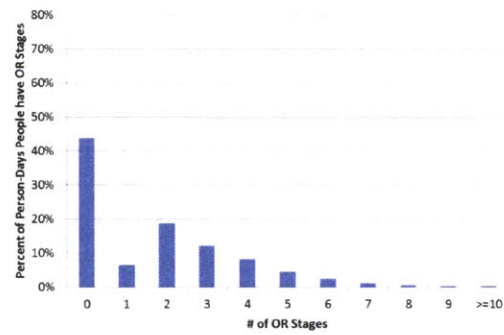
(a) Reported 0 PT Stages



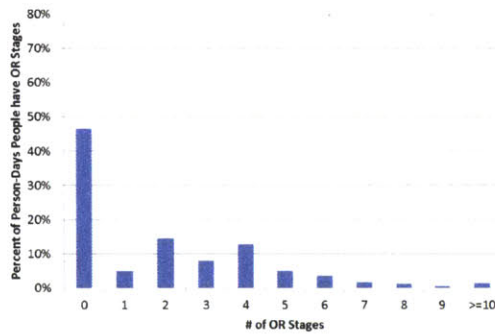
(b) Reported 1 PT Stage



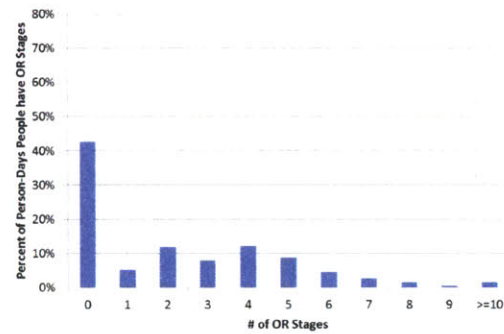
(c) Reported 2 PT Stages



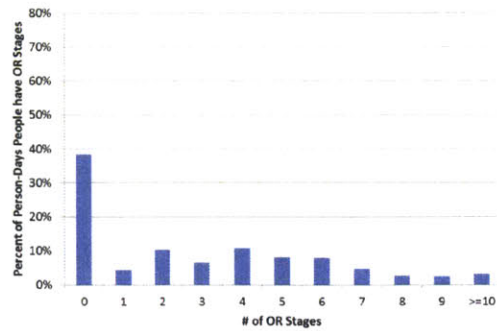
(d) Reported 3 PT Stages



(e) Reported 4 PT Stages



(f) Reported 5 PT Stages



(g) Reported 6 PT Stages

Figure 3-3: Percent of Person-Days People who Reported PT on Travel Day had OR Stages

### 3.5 Summary of Combining LTDS and Oyster Data

The main purpose of this research is to assess how one can combine the LTDS with Oyster card data to understand the accuracy and representativeness of the survey and determine how to use Oyster data to study the variability in PT travel behavior in great detail over a span of time not available with current survey methods. This was done most effectively for the sample of people who volunteered their Oyster card number(s), reported taking PT stages on their travel day, and had valid OR stages on that day (Type 1 people). With the other person types, assumptions were made about whether the discrepancies were due to the individual misreporting travel as opposed to misreporting the card number or fare media used.

In overall magnitude, the people who can be confidently assumed to have used their volunteered Oyster card(s) on the day they reported PT use (Type 1) underreported their single day PT journeys by at least 9%. Type 1 people underreported bus stages by 16% overall, and reported the same number of LTDS and OR train stages.

Only 22% of the PT Sample had perfectly matching reported LTDS PT stages and OR stages, which only accounted for 18% of the PT Sample's reported LTDS stages. The remaining 78% of the PT Sample either underreported, overreported, or misreported (or some combination of under, over, and/or misreporting) their PT stages. Approximately 27% of the PT Sample likely misreported the fare media they used on their travel day, which made up 26% of the PT Sample's reported LTDS stages. This is not considered a serious problem, as the reporting of the type of fare media used may have specific implications at Tfl, but will not affect the overall level of PT use reported in the LTDS. About 33% of the PT Sample likely misreported some portion(s) of the train O-D pair or bus route(s) used, which made up approximately 21% of the PT Sample's reported LTDS stages. This will have implications for the LTDS survey, if the specific origins and destinations or bus routes reported are being used for planning purposes. About 14% of the PT Sample overreported their PT stages, most likely in terms of the travel day, that is they reported a day other than their agreed upon travel day they may have thought was more representative of their regular travel behavior, which made up approximately 12% of the PT Sample's reported LTDS stages. An additional 26% underreported their PT stages, which made up approximately 21% of the PT Sample's reported LTDS stages. Only 51% of OR stages had matching LTDS stages, and only 46% of LTDS stages had matching OR stages, in terms of the train O-D pairs or bus routes used. The overall over and underreporting of PT stages will have an impact of the overall level of PT use reported in the LTDS. When taking into account the under and overreported PT stages, an additional 416 stages (931 Underreported - 515 Overreported) should be added to the PT Sample reported LTDS PT stages.

When looking at mutually exclusive categories of people who have mis, over, or underreported specific LTDS and OR stages in terms of the reported bus routes and train O-D pairs, the people who had perfectly matching LTDS and OR stages or only misreported the fare media used on their travel day made up only 44% of the PT Sample and 41% of the PT Sample's reported LTDS stages. The other 56% of the PT Sample over, under, or misreported the O-D pair or bus route of their 59% of stages, which has significant implica-

tions about the accuracy of the LTDS if more than half of the people who had reported PT travel in the survey are incorrectly reporting the number of PT stages and/or the origins, destinations, or bus routes used on their travel day.

The continuous collection of OR stages for all people who volunteered an Oyster card number allowed for the analysis of OR stages on other days in the OR analysis period to show the weekly variation in personal travel for LTDS respondents compared to the reported frequency of use of each PT mode in the LTDS. The reported frequency of PT use in the LTDS was much higher than the actual use as captured by the Oyster system, and therefore the LTDS is overestimating the PT use overall for London residents. The people who had OR stages on their travel days (Types 1, 4, and 7) were also the most frequent PT users in the two weeks prior and two weeks after their travel days. The presence of OR stages on the travel day was a much better predictor of actually being a frequent PT user than simply reporting PT use on the travel day in the LTDS.

One of the most significant findings from analyzing the weekly variation in PT travel is that people who reported one or more PT stages on their travel day had from 38-57% of their remaining person-days with zero PT stages. This suggests that the people who are considered PT users in the survey are only using PT about half of the time, and if the one day survey just so happens to assign a travel day on a day where the person used PT, it would overestimate their typical PT use. A two or three day survey would better capture the variability in PT use for survey respondents. Alternatively, the use of OR's and Gordon's method to link the separate stages made by an Oyster card user into full, multi-modal journeys could supplement the one-day survey and provide a much better estimate of the frequency of use and variability of PT travel than the frequency questions asked in the survey or fixed trip rates used in travel demand models.



## Chapter 4

# Analyses of Oyster Card Users (Type 1 Respondents)

As described in the previous chapters, the main purpose of this research is to examine how to combine Oyster card data with the London Travel Demand Survey (LTDS) to enhance and validate the survey responses on the single travel day and contribute to the understanding of the variability in travel behavior over a span of time not available with current survey methods. With the voluntary collection of LTDS respondents' Oyster card numbers, it was possible to combine these two data sets for the first time, allowing for the integration of the strengths of each data set. This was done most effectively for the sample of people who volunteered their Oyster card number, reported taking public transport (PT) journeys on their travel day, and had valid Oyster record (OR) stages on that day, the category of survey respondents defined as the Type 1 panel.

Of the 1,557 people who reported using their Oyster card on a PT stage on their travel day, 1,148 (74%) had valid OR stages on that day and are classified as Type 1 people. Valid stages are any type of PT use (not topping up or purchasing a period pass) and may or may not be the exact stage as reported in the LTDS. The most detailed analysis was done for Type 1 people because they had the most directly comparable information available about their travel behavior and can be assumed to have used the Oyster card they volunteered on their travel day. Type 1 people made up approximately 29% of all people who volunteered Oyster card numbers. The representativeness of Type 1 people along with detailed analyses of the accuracy of their survey responses are described in this chapter.

### 4.1 Representativeness of Type 1 Respondents

With the voluntary inclusion of Oyster card numbers in the LTDS, the accuracy of PT stages was determined by comparing an individual's reported PT stages with their PT stages captured by the Oyster system. However, this was only possible for the people who decided to volunteer their Oyster card number, reported using PT on their travel day, and

had valid<sup>1</sup> OR stages on that travel day - the people in the Type 1 panel. Of the 14,325 who took part in the LTDS between July 2011 and March 2012, 4,027 (28%) of them reported taking PT stages on their travel day using some form of fare media. Of these, 1,148 (8% of total) volunteered their Oyster card number and had valid OR stages on that card on their travel day (Type 1). The other 2,879 (20% of total) either did not volunteer an Oyster card number (2,470 people or 17% of total), or reported using their volunteered Oyster card, but did not have valid OR stages on that card on their travel day (409 Type 2 or 3 people, or 3% of total). In order to determine if the sample of Type 1 people was a representative sample of all people who reported taking PT on their travel day (regardless of fare media used) and draw conclusions about the overall accuracy of the reported PT use in the LTDS, statistical tests were done to compare the reported number of PT stages, mode share distributions, and demographic characteristics between Type 1 people and all other people who reported taking PT on their travel day.

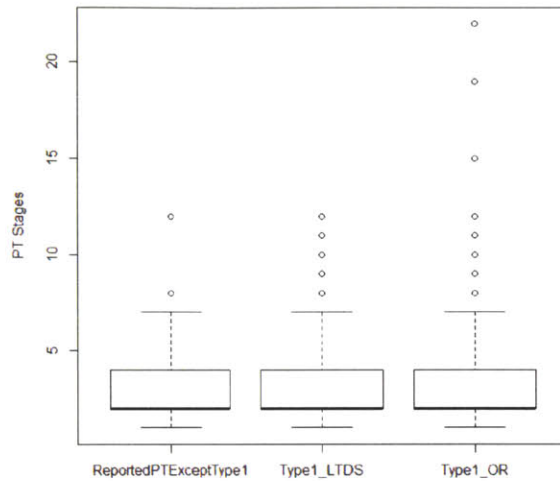
In general, Type 1 people were representative of all people who reported taking PT stages on their travel day in terms of the distribution of PT stages reported and OR stages on the travel day, as shown in Figure 4-1. The boxplot<sup>2</sup> on the left shows the distribution of reported PT stages for all people, except Type 1, who reported taking PT on their travel day. The middle boxplot shows the distribution of reported PT stages using an Oyster card on the travel day for Type 1 people, and the boxplot on the right shows the distribution of OR stages on the travel day for Type 1 people. The distributions are almost identical, with slightly more outliers found in the OR stages of Type 1 people, but overall Type 1 people were representative of all people who reported PT stages in terms of the distribution of reported PT stages.

---

<sup>1</sup>A valid Oyster record stage is any type of PT use (not topping up) on the Oyster card and may or may not be the exact stage as reported in the LTDS

<sup>2</sup>All boxplots have the following features from top to bottom:

- 1) Circles represent outliers which are more than 1.5 times the upper or lower quartile
- 2) The top line represents the Maximum which is the greatest value, excluding outliers
- 3) The top line of the box is the Upper Quartile meaning 25% of the data is greater than this value
- 4) The middle dark line in the box is the Median meaning 50% of the data is greater than this value
- 5) The bottom line of the box is the Lower Quartile meaning 25% of the data is less than this value
- 6) The lowest line is the Minimum or the least value excluding outliers



**Figure 4-1:** Comparison of PT Stages for Type 1 vs. All Others who Reported PT

The average number of reported PT stages was also analyzed for Type 1 people and all others who reported PT stages on their travel day. The results are shown in Table 4.1. The P-values were calculated for a two-tailed hypothesis test that the average number of reported LTDS PT stages was exactly the same for Type 1 people as for all other people who reported PT and the average number of Type 1 LTDS PT stages was exactly the same as the average OR stages for Type 1 people. Large P-values (more than 0.05) mean that it is likely that the average number of PT stages was exactly the same for each sample. Overall, Type 1 people reported approximately the same average number of LTDS PT stages using their Oyster card(s) as all others who reported PT stages, but had a higher number of average OR stages on their travel day. This shows that Type 1 people were not different than the other people who reported PT stages in terms of the number of PT stages that were reported in the LTDS, but they had more OR stages on their travel day than they reported. This result suggests that the other people who reported PT, but did not volunteer an Oyster card or have valid OR stages on their travel day may also have traveled more than they reported on their travel day, and therefore the single day survey may be underestimating the intensity of PT use on days where the survey sample chose to use the PT mode.

**Table 4.1:** Comparison of Average Number of PT Stages

	Reported PT, not Type 1	Type 1 Reported Stages on Oyster	Type 1 OR Stages
Average # Stages	2.82	2.91	3.17
Std Dev # Stages	1.51	1.54	2.08
P-Value (Reported PT, not Type 1 = Type 1 LTDS )	<b>0.09</b>	P-Value (Type 1 LTDS = Type 1 OR)	0.01

Table 4.2 summarizes the differences in the reported mode shares for the Type 1 people and all other people who reported PT stages. The share of trips for each mode was calculated by dividing the total reported trips in the sample for each mode (using the distance based main mode of transport for trips with multiple stages) by the total number of reported trips in the sample. The P-values were calculated for a two-tailed hypothesis test that the proportions of trips by each mode are exactly the same for Type 1 people and all others who reported PT stages. Large P-values (more than 0.05, and shown in bold in the table) mean that it is likely that the mode shares are exactly the same for each sample. This table shows that the mode share distributions for the two samples were statistically the same for Auto, Underground, and Taxi/Other modes. Type 1 people had a higher share of Walk and Other Train trips, and a lower share of Bus and National Rail trips. While these populations are not statistically exactly the same with respect to their use of all modes, their differences are less than 4% for all modes except Walk, and the approximate relative use of all modes was similar. This may mean that Type 1 people walked slightly more than people who did not volunteer an Oyster card number or did not use their volunteered card on their travel day.

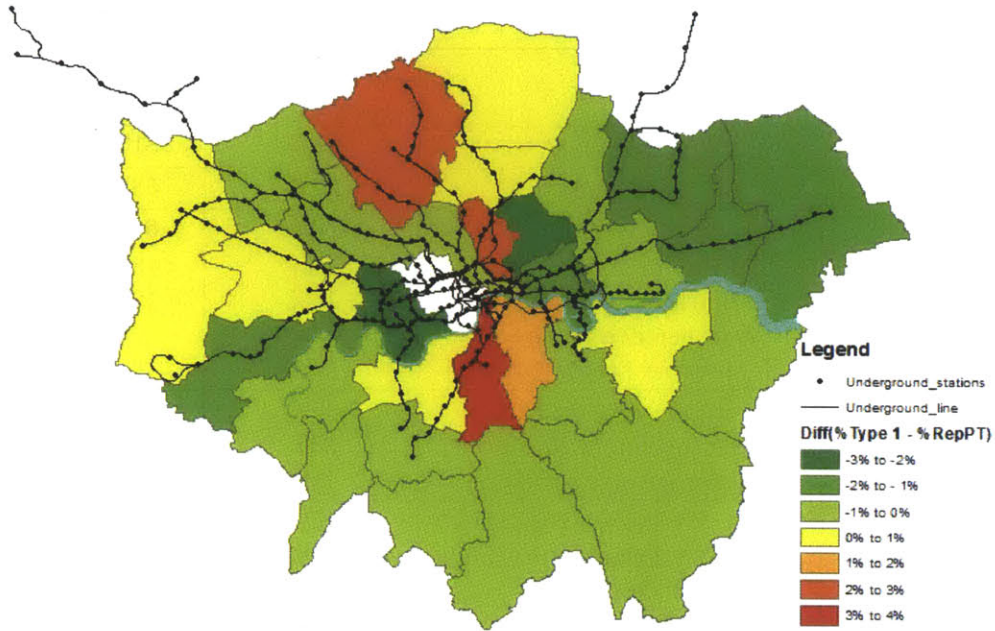
**Table 4.2:** Mode Share Differences for People who Reported PT (not Type 1) and Type 1

LTDS Reported Mode <sup>3</sup>	Reported PT (not Type 1)	Type 1	P-Value
Walk or Cycle	21.7%	27.8%	0.00
Auto	8.5%	8.2%	<b>0.61</b>
Bus	35.4%	31.3%	0.00
Underground	17.7%	18.9%	<b>0.13</b>
National Rail	12.4%	9.1%	0.00
Other Train	2.1%	2.8%	0.02
Taxi/Other	2.1%	1.9%	<b>0.30</b>
Number of Samples	8,614	3,940	

Figure 4-2 shows the difference of the percentage of Type 1 people minus the percentage of all other people who reported PT stages on their travel day in each borough.

<sup>3</sup>Walk and cycle are combined in Walk or Cycle, Car and motorcycle driver and passenger are combined in Auto, DLR, Tramlink, and Overground are combined in Other Train, all other modes not listed are combined in Taxi/Other





**Figure 4-2:** Difference of % Type 1 People and % of Other People who Reported PT by Borough

Boroughs with negative values (in green) had fewer Type 1 people than all other people who reported PT in that borough. Boroughs with positive values (in yellow, orange, and red), had more Type 1 people than all other people who reported PT in that borough. Boroughs along the center of London from east to west had fewer Type 1 people, and boroughs in the center just south of the River Thames and in the north of London had more Type 1 people. However over 70% of boroughs had a difference smaller than 1%, showing that overall, the spread of Type 1 people geographically was relatively similar to the spread of all other people that reported taking PT on their travel day.

Table 4.3 shows the demographic differences between Type 1 people and all other people who reported PT stages on their travel day. The table shows the percentage of people in each sample with each demographic characteristic.

**Table 4.3: Comparison of Demographic Characteristics**

Demographic Characteristic		Reported PT (not Type 1)	Type 1	P-Value
Travel Day	% Weekday	81%	81%	<b>0.78</b>
Gender	% Female	52%	58%	0.00
Ethnicity	% White	60%	68%	0.00
	% Asian or Asian British	18%	13%	0.00
	% Black or Black British	16%	12%	0.01
	% Other Ethnic Group	4%	3%	<b>0.43</b>
	% Multiple Ethnic Groups	3%	3%	<b>0.68</b>
HH Income	Lower Income (<£25k)	47%	46%	<b>0.88</b>
	Middle Income (£25 - £75k)	38%	37%	<b>0.61</b>
	Higher Income (>£75k)	15%	16%	<b>0.83</b>
Household Size	%1 Person HH	17%	24%	0.00
	%2 Person HH	27%	35%	0.00
	%3 Person HH	20%	17%	<b>0.09</b>
	%4 Person HH	20%	15%	0.00
	%5 Person HH	9%	6%	0.00
	% 6 or more Person HH	8%	3%	0.00
Zero Vehicle HH	% Zero Vehicle HH	47%	52%	0.00
Disability	% No Disability	93%	88%	0.00
Age	% 5-10	4%	0%	0.00
	% 11-15	10%	0%	0.00
	% 16-24	16%	17%	<b>0.95</b>
	% 25-44	39%	45%	0.00
	% 45-59	16%	17%	<b>0.64</b>
	% 60+	14%	22%	0.00
Total		2,879	1,148	

The travel day for each sample was analyzed to see if more Type 1 people had travel days that were weekdays than others who reported PT. The percentages of each sample with a weekday travel day were statistically the same, showing there was no bias of weekday or weekend travel days within the Type 1 sample. There were also the same proportions of household incomes in each sample.

Type 1 people were more likely to be white females from the age groups of 25-44 and over 60 than the others that have reported PT stages, but did not volunteer or did not use their volunteered Oyster card on their travel day. They were also more likely to be from smaller households with less access to a vehicle and to have reported a disability that impacts their mobility. There were no Type 1 people under the age of 16 because the survey only collected Oyster card information for individuals aged 16 and older. These differences should be kept in mind when using a panel of Type 1 people to represent all those who reported PT

stages on their travel day. Over time, as the panel grows the differences should be reduced. Additionally, the surveys can be administered in a way to encourage more men from other ethnic groups and with other demographic characteristics lacking in the Type 1 sample to volunteer the Oyster card number they used on their travel day through incentives such as Oyster credits.

Demographic characteristics for all other Person Types compared to the entire sample of people who volunteered their Oyster card number(s) are shown in Appendix B.

## 4.2 Overall Number of PT Stages

The total number of PT stages reported in the LTDS by Type 1 people was compared to the total number of stages captured by the Oyster system for Type 1 people on their travel day. Reported PT stages were combined, when necessary, to ensure stages were counted consistently in both sets of data. For example, two consecutive PT stages on the Underground would appear as two reported PT stages in the LTDS, but one OR stage because interchanges on the Underground are not captured by the Oyster system. In cases like this, the LTDS stages were combined and counted as one stage to ensure equivalent stage comparisons. The LTDS interviews respondents on all days of the week and captures travel day information for weekdays as well as weekends. The results were separated to show the differences of reported PT stages and OR stages for people with weekday or weekend travel days. The total number of reported LTDS PT stages and OR stages for Type 1 people are summarized in Table 4.4. This shows that overall Type 1 people reported 8.9% fewer PT stages than OR stages. Approximately 81% of Type 1 people had a weekday travel day, with their reported LTDS stages and OR stages making up approximately 81.6% and 81.8% of all reported or OR stages respectively. Weekday stages were underreported by about 9.2% and weekend stages were underreported by about 7.5%.

**Table 4.4:** Type 1 Overall Amount of PT Travel

Travel Day Type	People	% of Type 1	LTDS Stages <sup>4</sup>	OR Stages	LTDS - OR	% Difference: $\frac{LTDS - OR}{LTDS}$
Weekday	930	81%	2,723	2,973	-250	-9.2%
Weekend	218	19%	615	661	-46	-7.5%
Total Type 1	1,148	100%	3,338	3,634	-296	-8.9%

### 4.2.1 Intensity of PT Use

The OR stages for Type 1 people were collected eight weeks prior to the travel day<sup>5</sup> and indefinitely after the interviews, though this thesis only analyzes OR stages through 18

<sup>4</sup>Includes only PT stages

<sup>5</sup>It is possible to retrieve eight weeks of Oyster data retroactively.

June, 2012. These records were analyzed to find the average number of OR stages for each Type 1 person for the entire OR analysis period. Figures 4-3 and 4-4 show side by side comparisons of the reported PT stages and OR stages for all Type 1 people with travel days on weekdays or weekend days, respectively.

In each figure, the boxplot on the far left (Avg WDayOR Cap WDays) is the distribution of the number of weekday OR stages captured by the Oyster system over all weekdays where the person had OR stages. This represents the distribution of people's average OR stages per weekday that they took PT, as captured by the Oyster system. The boxplot second from the left (Avg WDayOR Pos WDays) shows the distribution of the average number of weekday OR stages over all weekdays where OR stage capture was possible (eight weeks prior to the individual's travel day through 18 June, 2012). The third boxplot from the left (Avg WEndOR Cap WEnds) shows the distribution of the number of weekend OR stages captured by the Oyster system over all weekend days where the person had OR stages. This represents the distribution of people's average OR stages per weekend that they took PT, as captured by the Oyster system. The boxplot fourth from the left (Avg WEndOR Pos WEnds) shows the distribution of the average number of weekend OR stages over all weekend days where OR stage capture was possible (eight weeks prior to the individual's travel day through 18 June, 2012). The boxplot second to the right shows the distribution of reported LTDS stages for people on their travel day, and the boxplot on the far right (Oyster on Travel Day) shows the distribution of OR stages for people on their travel day.

The distributions of reported PT stages in the LTDS and captured by the Oyster system on the travel day for people with weekday travel days (shown in Figure 4-3) were very similar. The medians for these two distributions are the same, but the average number of OR stages on the travel day was 9.2% higher than the average number of reported LTDS PT stages. The distribution of average weekday OR stages over all days where weekday travel day people had OR stages had a higher median and a narrower range than the reported LTDS or captured OR stages on the travel day, and the average weekday OR stages over all possible weekdays of OR collection had a lower median and a narrower range than the travel day. The weekend results were similar with slightly lower medians and narrower distributions than the weekday results. This shows that overall, people with weekday travel days did not use PT every weekday throughout the OR analysis period, but on the weekdays they did use PT, they used it more on average than on their travel day. Therefore, weekday travel days were underreported by 9.2% in terms of the number of PT stages, and this weekday was not necessarily representative of all weekdays the person took PT or of all possible travel weekdays for this sample. Their travel patterns were also different for weekend days, with less use on the weekend days overall, and slightly less intensity of use than weekday travel.

The distributions of reported PT stages in the LTDS and captured by the Oyster system on the travel day for people with weekend travel days (shown in Figure 4-4) had similar ranges, but the median of the OR stages on the travel day was higher than the reported PT stages on the travel day, and the average number of OR stages on the travel day was 7.5% higher than the average number of reported PT stages. The distribution of average weekend OR stages over all days where weekend travel day people had OR stages had a higher median

and a narrower range than the reported LTDS or captured OR stages on the travel day, and the average weekend OR stages over all possible weekend days of OR collection had a much lower median and a narrower range than the travel day. The average weekend OR stages over all possible weekend days of OR collection was much lower than the reported travel day for the weekend sample compared to the respective comparison of the sample of Type 1 people with weekday travel days. This shows that overall, people with weekend travel days did not use PT every weekend throughout the OR analysis period to a greater extent than the weekday people did not use PT every possible weekday throughout the OR analysis period, and on the weekend days they used PT, they used it more on average than on their travel day. Therefore, weekend travel days were underreported by 7.5% in terms of the number of PT stages, and this weekend was not necessarily representative of all weekend days the person took PT or of all possible travel weekend days for this sample.

Figures 4-3 and 4-4 raise some interesting and very important questions about the use of LTDS survey results. The figures suggest that at least for PT travel, the single day survey overestimates typical PT use overall and especially on weekend days, but it underestimates the intensity of PT use on days when the survey sample chose to use the PT mode.

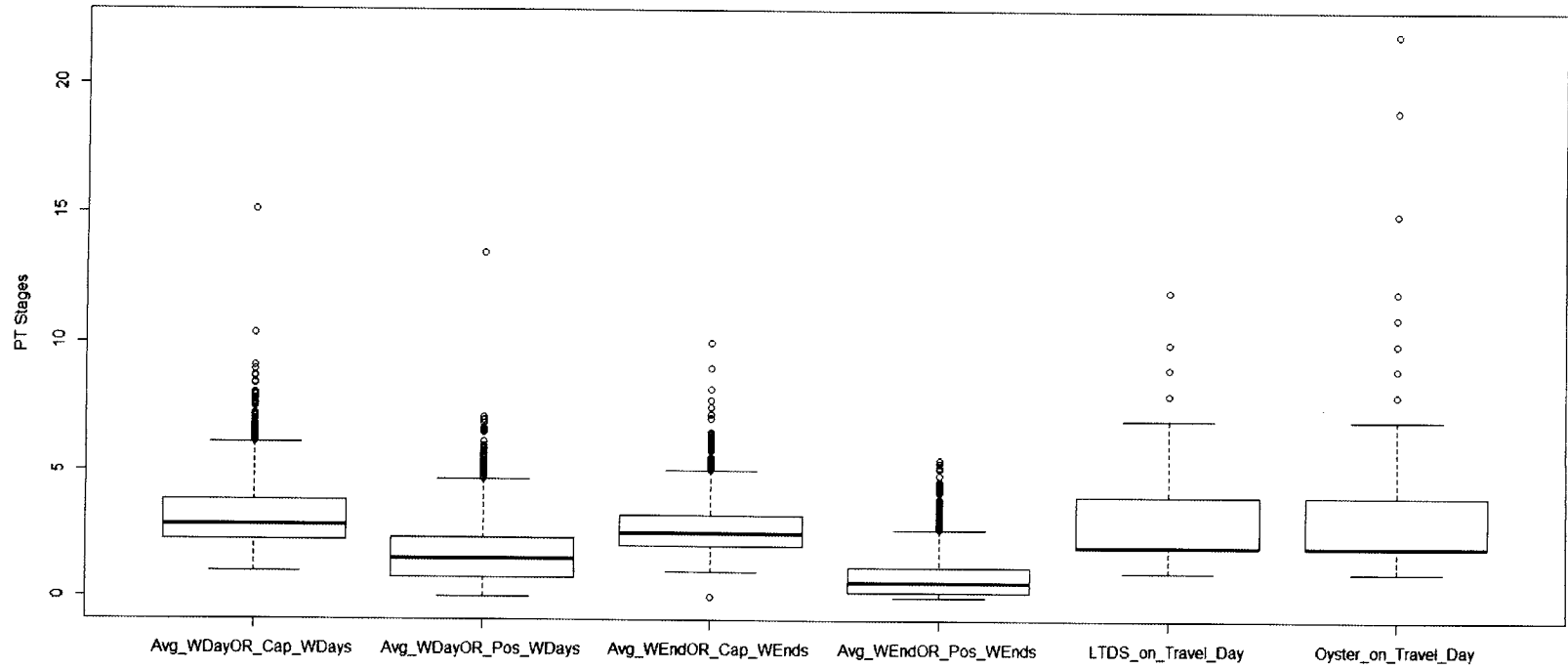


Figure 4-3: LTDS vs. Oyster Stages for Type 1 People with Weekday Travel Days

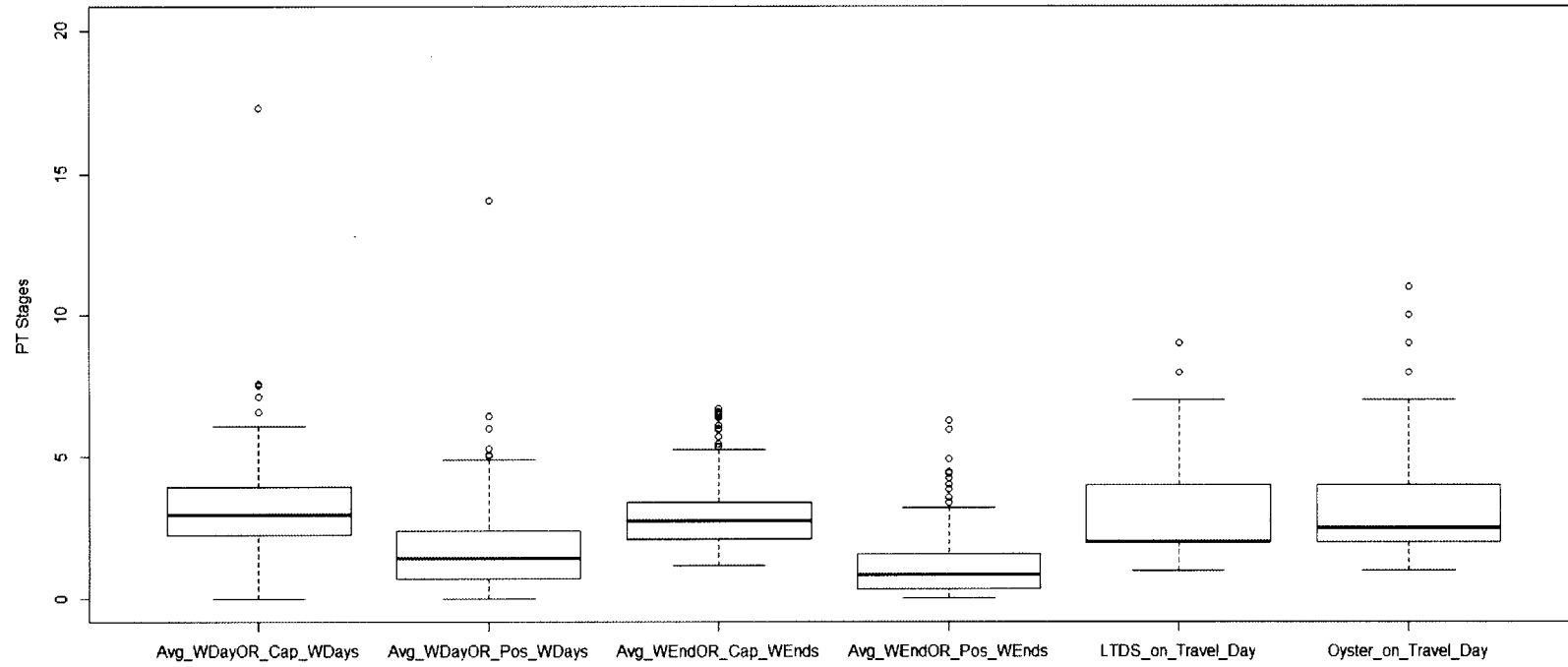


Figure 4-4: LTDS vs. Oyster Stages for Type 1 People with Weekend Travel Days

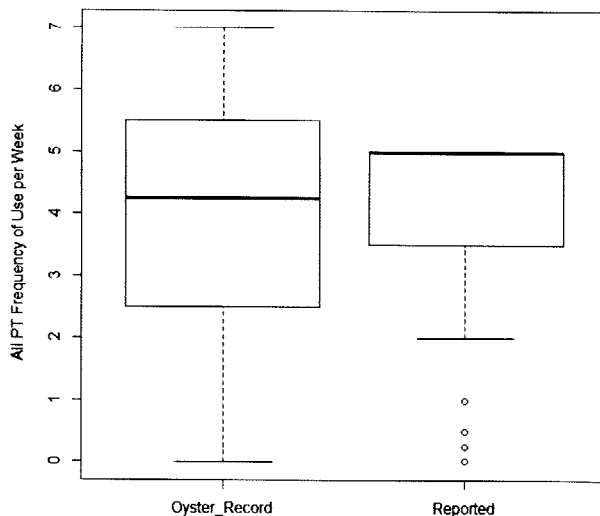
## 4.2.2 Reported vs. Actual Frequency of Use

The survey asks people questions about their frequency of use of all modes of the PT system. In order to compare this LTDS reported frequency of use with the actual frequency of use as captured by the Oyster system, the LTDS reported frequency of PT use was converted to a use per week and is shown below in Table 4.5. The maximum frequency of all PT modes<sup>6</sup> is used in this analysis. For example, if a person reported using the bus five times per week but the Underground three times per week and all other modes less than once a month, their response is categorized here as using PT five times per week.

**Table 4.5:** Categories of Reported Frequency Converted to Use per Week

LTDS Category	Use per Week
Never used	0
Not used in last 12 months	0
At least once a year	0.02
At least once a month	0.25
At least once a fortnight	0.5
1 day a week	1
2 days a week	2
3 or 4 days a week	3.5
5 or more days a week	5

In addition, the actual PT frequency of use was calculated by counting the number of days the person had valid OR stages two weeks prior and two weeks after their travel day to normalize the counting method for all people. Analysis of Type 1 peoples' OR stages compared with the LTDS reported frequency of PT use is shown in Figure 4-5.



**Figure 4-5:** Type 1 Reported vs. Actual Frequency of PT Use

<sup>6</sup>PT modes are bus, Underground, NR, DLR, Overground, and Tram.



The LTDS reported frequency of PT use was higher and had a tighter distribution than the actual PT use found in OR stages for Type 1 people. The median response for Type 1 people's reported frequency of use was five or more days per which was the most frequent categorical response possible. The actual variability captured by the Oyster system shows a wider range of variability with a median of 4.3, showing that the LTDS is not capturing the actual variability of PT use for the survey sample. It is not clear how the frequency of use question is used in the TfL modeling process, but if it is used to scale up the frequency of PT use for all London residents, it may be overestimating the number of PT journeys people are making.

### 4.2.3 Difference in Oyster Card Types

The difference between the number of reported PT stages and OR stages was compared for Type 1 people depending on if they reported having a period pass or pass for free travel on their Oyster card, or if they used Pay as You Go (PAYG) credit only. Approximately 48% of Type 1 people did not report having a pass on their Oyster card<sup>7</sup>.

It was assumed that people who had a period pass or pass for free travel would travel more than people who used PAYG only. This was tested by calculating the average number of PT stages for Type 1 people on their travel day (OR stages and reported LTDS stages) with any kind of period pass or pass for free travel compared to the average number of PT stages for Type 1 people with PAYG only as shown in Table 4.6. This table shows that people who had any kind of pass had a higher average number of bus and total PT stages than people who used PAYG only, and the differences are statistically significant. The average LTDS train stages and OR train stages are not statistically different between the people who had a pass or PAYG only. This shows that overall, people who had a pass used the bus more intensely on their travel days than people with PAYG only.

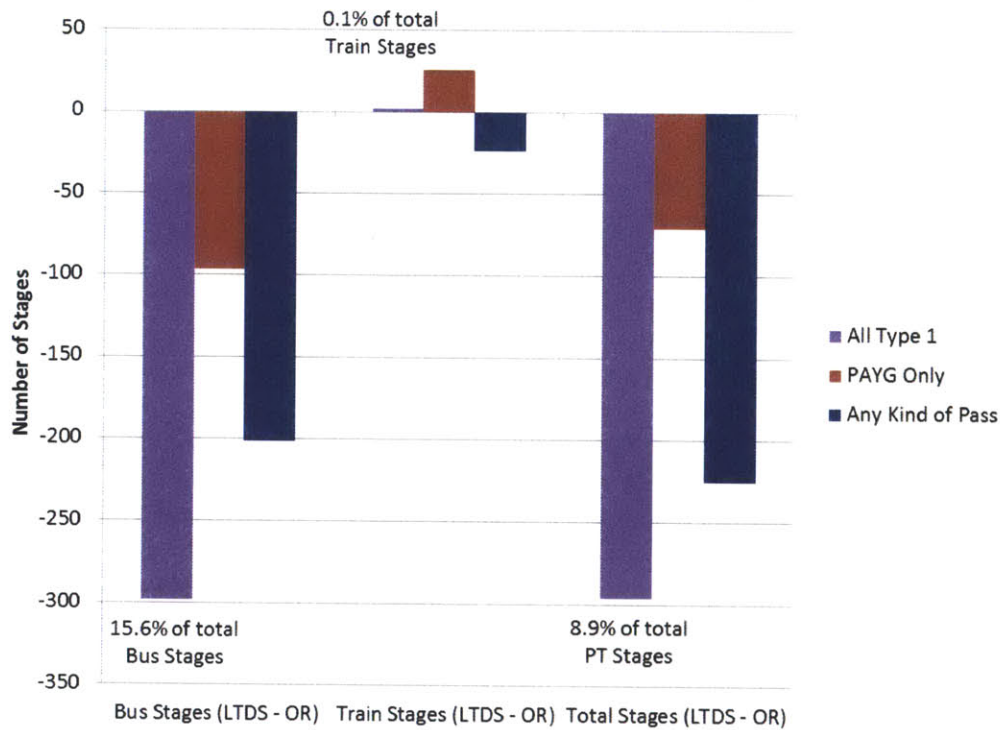
**Table 4.6:** Average PT Stages by Oyster Card Type

	Average Stages					
	LTDS Bus	OR Bus	LTDS Train	OR Train	LTDS Total	OR Total
Any Kind of Pass	1.88	2.22	1.19	1.23	3.07	3.45
PAYG Only	1.42	1.60	1.31	1.26	2.73	2.86
P-Value (Diff = 0)	0.00	0.00	<b>0.15</b>	<b>0.75</b>	0.00	0.00

It was also assumed that people who only use PAYG credit would more accurately report the number of PT stages they took on their travel day because they are more conscious of the money they spend each time they travel. The results of this comparison are shown in Figure 4-6. This figure shows the difference in the number of PT stages (LTDS - OR) for all

<sup>7</sup>This proportion of PAYG only for Type 1 people was statistically the same at the 95% confidence level as the other people who reported having an Oyster card and using it for PT travel on their travel day (Types 2 and 3 people).

Type 1 people, combined, by whether or not they had a period pass or pass for free travel or used PAYG credit only. The purple bars show the difference in the number of PT stages for all Type 1 people, which is made up of the addition of the red (PAYG only) and dark blue (any kind of pass) bars. There were 296 more OR stages than reported LTDS stages for all Type 1 people. Approximately 76% of these extra OR stages were captured by the Oyster system on the Oyster cards of Type 1 people with a period pass or pass for free travel. When looking at bus stages only, there were 298 fewer reported bus stages than OR stages, with 67% of these extra OR stages captured by the Oyster system on the Oyster cards of Type 1 people with a period pass or pass for free travel.

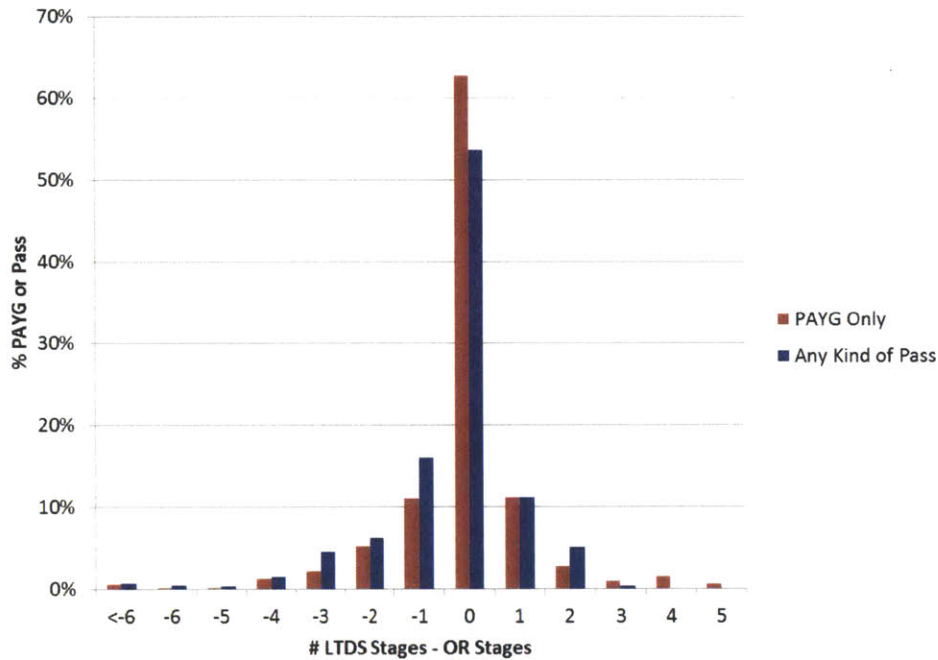


**Figure 4-6:** Difference in PT Stages for Type 1 People by Oyster Card Type

These results confirm the assumption that people who used PAYG credit for PT travel more accurately reported the total number of PT bus stages than people with period passes or passes for free travel. For train stages, there was only an overall difference of two stages with 24 fewer LTDS than OR train stages for people with passes, and 26 more LTDS than OR stages for people with PAYG only. This was most likely due to the people with PAYG only overreporting these 26 train stages and the people with a period pass underreporting these train stages.

The difference in the number of reported PT stages and OR stages per person was also analyzed by whether or not the person had a period pass or pass for free travel or used PAYG only. The results are shown in Figure 4-7. Negative differences between the number of reported LTDS stages and OR stages were people who underreported the number of PT stages on their travel day. This figure shows that the people who more accurately reported their total PT stages were people who used PAYG only because more PAYG only

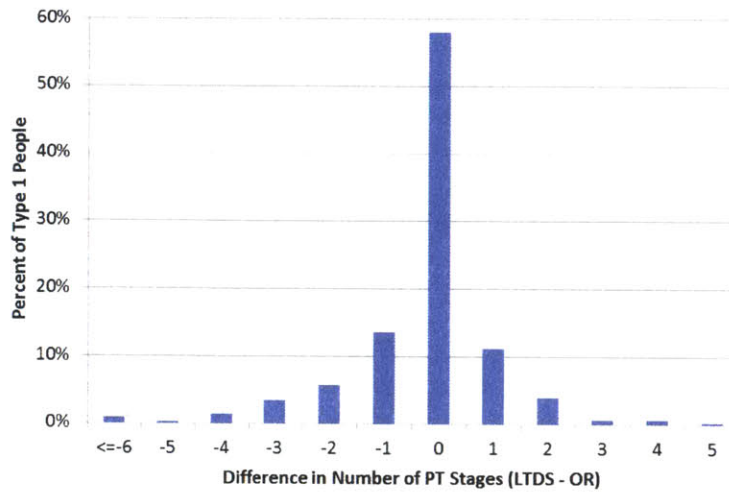
people had a difference of zero PT stages. More people with a period pass or pass for free travel underreported their PT stages, with approximately 30% of people with a pass underreporting and only 21% of PAYG only people underreporting. The number of people with more LTDS stages than OR stages was approximately the same for people with passes or PAYG only at approximately 17%.



**Figure 4-7:** Difference in Number of PT Stages (LTDS - OR) by Oyster Card Type

#### 4.2.4 Difference in Stages per Person

Next, individual Type 1 people’s stages were analyzed to see the difference in reported PT stages and OR stages per person. Figure 4-8 shows the difference in number of stages between the reported PT and OR stages by the percentage of Type 1 people with each difference. The distribution is relatively normal with an average difference 0.26 fewer reported stages per person than OR stages. Approximately 58% of Type 1 people reported the same number of PT stages as captured by the Oyster system; however, 25% of people underreported their PT journeys and 17% of people overreported PT stages on their Oyster card. Overall, Type 1 people underreported PT stages by 8.9%.



**Figure 4-8:** Difference in Number of PT Stages (LTDS - OR)

It is also important to know the difference in stages reported in the LTDS and found in the Oyster system by the total number of stages reported. For example, if most of the differences in stages were from people who had a large number of stages, it would not have as big an impact as if the differences were for people with only a few PT stages. Figure 4-9 shows the number of reported LTDS stages on the x-axis and the percentage of Type 1 people on the y-axis. The colors represent the difference between the number of LTDS stages and OR stages on the travel day. Negative values for the colors represent people who reported fewer stages in the LTDS than were found in the OR stages for that person (underreporting), and positive values represent people who reported more LTDS stages than OR stages (overreporting). This figure shows that over 63% of the people underreporting PT stages were people who reported three or fewer PT stages on their travel day, which is the average number of PT stages reported by Type 1 people. Approximately 56% of the people overreporting PT stages reported four or more PT stages on their travel day. This shows that people who underreported PT stages reported fewer stages overall and were missing a higher percentage of their stages on their travel day than people who overreported PT stages. More details about the under and overreporting of stages will be discussed in sections 4.3 and 4.4, respectively.

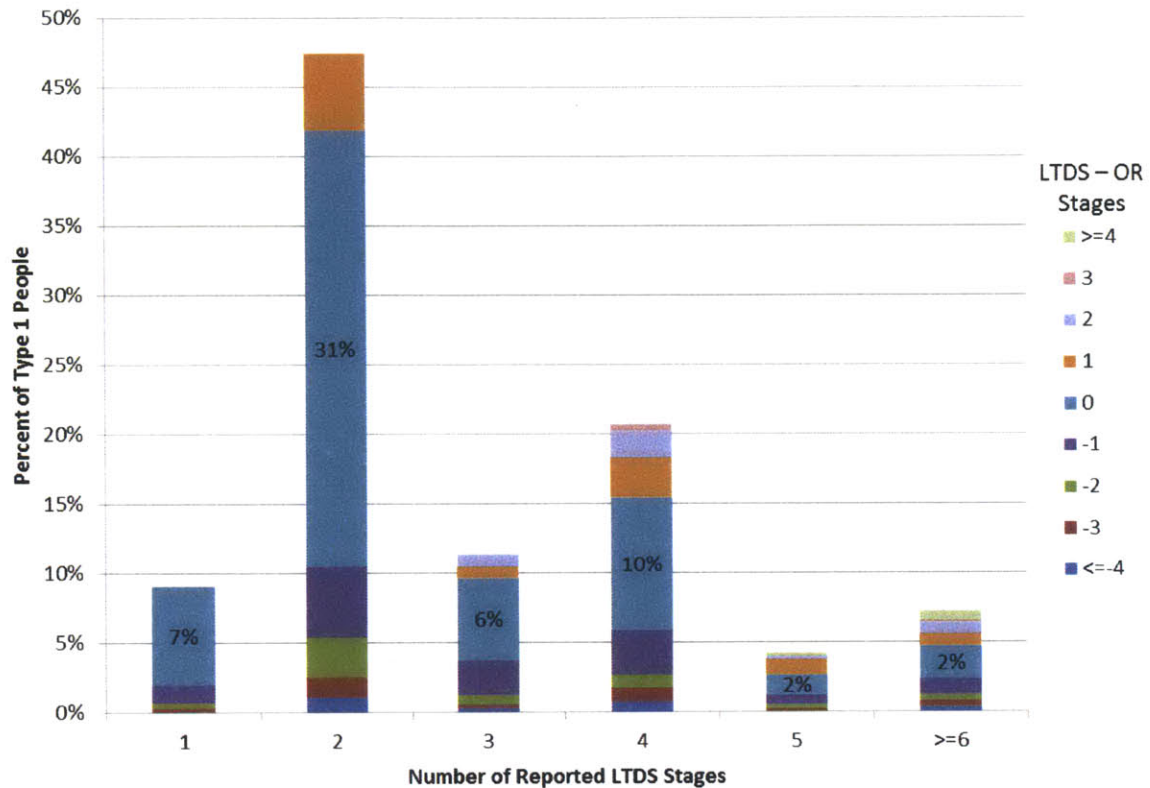


Figure 4-9: Number of Reported LTDS Stages vs. Difference in OR and LTDS Stages

### 4.3 Unreported Oyster Records on Travel Day

Approximately 60% of Type 1 people had OR stages on their travel day that were not reported in the LTDS. These records could be from journeys that were completely forgotten and therefore underreported, or journeys that were misreported as a different bus route or train origin-destination (O-D) pair. In order to determine which people had unreported or misreported OR stages, the OR stages with no corresponding LTDS stages and the LTDS stages with no corresponding OR stages were compared for each person by mode. For example, if a person had an OR stage on a bus route that was not reported in the LTDS, but reported a bus stage on a route that was also not found in his OR stages, this particular OR stage was categorized as a misreported journey in terms of the bus route or train O-D pair. However, if a person had an OR stage on a bus route not reported in the LTDS, and reported no bus stages without corresponding OR stages, it was categorized as an underreported journey.

Type 1 people had 1,608 OR stages that were not reported in the LTDS. There were 335 people who underreported 666 LTDS PT stages, and 562 people who misreported 942 LTDS PT stages in terms of the bus route or train O-D pair. There were also nine people who had 20 OR stages matching a reported LTDS PT stage they said they paid for with something other than their Oyster card(s). These nine people's 20 OR stages were considered misreported in terms of the fare media reported or used.

## 4.4 Reported Stages not Captured in Oyster Records

Approximately 59% of Type 1 people reported LTDS stages on their travel day that were not found in the Oyster system. These records could be from journeys that were misreported as a different bus route or train O-D pair, journeys that were reported and not taken on the travel day, and therefore overreported, or journeys that were reported and where the person actually made the journey but not using the volunteered Oyster card or where the person was able to travel without having to tap his Oyster card on a card reader (between two DLR or NR stations that do not require tapping in and out with a period pass). In order to determine which people had extra LTDS stages due to misreporting the bus route or train O-D pair, overreporting, or misreporting the fare media used, the LTDS stages with no corresponding OR stages and the OR stages with no corresponding LTDS stages were compared for each person by mode. For example, if a person had an LTDS stage on a bus route that was not found in the Oyster system, but had an OR stage on a bus that was also not reported in the LTDS, this particular LTDS stage was categorized as a misreported journey in terms of the bus route or train O-D pair. However, if a person had an LTDS stage on a bus route not found in the Oyster system, and had no OR stages without corresponding LTDS stages, it was categorized as an overreported journey (that could be a result of journeys that were reported and not taken on the travel day or with the volunteered Oyster card, or journeys that were taken, but where the person did not tap an Oyster card reader for the reasons described previously). These journeys were examined further to determine if the person misreported the fare media used or reported a day other than their travel day that they may have believed was more representative of their normal travel behavior, which was considered overreporting.

Type 1 people had 1,309 LTDS stages that were not captured by the OR system. There were 240 Type 1 people who overreported 367 stages in the LTDS, and 562 Type 1 people<sup>8</sup> who misreported 942 LTDS stages in terms of the bus route or train O-D pair.

The LTDS stages of the 240 people who overreported were compared to their OR stages within the week prior to their travel day to determine if they may have reported a day other than their travel day that they felt was more representative of their normal travel behavior. Approximately 46% of these overreporters (111 people) had some or all of their 177 reported LTDS PT stages in their OR stages within the previous week of their travel day. This suggests that these Type 1 people may have reported a different day they felt was more representative of their travel behavior. However, some of these people also misreported stages on their travel day, so it was hard to tell which specific stage was misreported and which was overreported without examining the PT network to determine likely alternate paths for each person. The other 54% (129 people) of Type 1 overreporters had none of their 190 reported LTDS PT stages in OR stages within the previous week of their travel day, so it is likely these people used a different Oyster card or did not tap their card for these stages on their travel day.

---

<sup>8</sup>Note there were many people who had both overreported and misreported PT stages, as well as people who had underreported bus stages, and overreported train stages, etc.

## 4.5 Summary of Misreporting Specific PT Stages

Table 4.7 shows the summary of the over, under, and misreporting of specific PT stages for Type 1 people.

**Table 4.7:** Summary of Over, Under or Misreporting of Specific PT Stages

	Type 1 People	Type 1 Bus	Type 1 Train	Total Type 1	Sample Description
Perfectly Matching	377 (33%)	443 (23.3%)	375 (30.9%)	818 (24.5%)	(% Type 1 People   % Total Stages)
Misreporting Fare Media	138 (12%)	93 (4.9%)	117 (8.2%)	210 (6.3%)	
Misreporting Portion of O-D	562 (49%)	497 (26.1%)	445 (31.1%)	942 (28.2%)	
Overreporting (likely misreporting travel day)	11 (10%)	110 (5.8%)	767 (4.7%)	177 (5.3%)	
Underreporting	335 (29%)	449 (26.2%)	167 (11.7%)	666 (20.0%)	

There were only 376 people Type 1 people (22% of the PT Sample and 33% of Type 1 people) with perfectly matching reported LTDS PT stages and OR stages, that is, all of the PT journey stages reported in the LTDS had matching OR stages in terms of bus routes and/or train O-D pairs in the same chronological order. One additional person had the same LTDS PT and OR stages, but out of sequence. Their stages made up 18.4% of the PT Sample's reported LTDS stages or 24.5% of Type 1 people's PT stages. The remaining 67% of Type 1 people either underreported, overreported, or misreported (or some combination of under, over, and/or misreporting) their PT stages. As shown in Table 4.7, approximately 12% of Type 1 people have likely misreported the fare media they used on their travel day, which made up 6.3% of Type 1 people's reported LTDS stages. This is not considered a serious problem, as the reporting of the type of fare media used may have specific implications at TfL, but will not affect the overall level of PT use reported in the LTDS.

About 49% of Type 1 people have likely misreported some portion(s) of the train O-D pair or bus route(s) used, which made up approximately 28.2% of Type 1 people's reported LTDS stages. This will have implications for the LTDS survey, if the specific origins and destinations or bus routes reported are being used for planning purposes. About 10% of Type 1 people overreported their PT stages, most likely in terms of the travel day, that is they may have reported a day other than their agreed upon travel day because they thought was more representative of their regular travel behavior or for other reasons, which made up approximately 5.3% of Type 1 people's reported LTDS stages. Approximately 29% of Type 1 people underreported their PT stages, which made up approximately 20.0% of the Type 1 people's reported LTDS stages. The overall over and underreporting of PT stages will have

an impact of the overall level of PT use reported in the LTDS. When taking into account the under and overreported PT stages, 489 stages (666 Underreported - 177 Overreported) should be added to the PT Sample reported LTDS PT stages.

## 4.6 Trip Analysis

Next, an individual's reported PT trips were compared to the captured OR stages on that day in terms of the bus route for bus stages, and the O-D pair for train stages. A PT trip is made up of multiple stages<sup>9</sup> and the percentage of each trip with an OR for each stage can be understood using the following example. If a person has a trip made up of one bus and one train stage, but there was only an OR stage for the bus route, the trip would be considered to have OR stages for 50% of its reported stages. Figure 4-10 shows the distribution of stages per trip for Type 1 people and how many reported stages in the trip had OR stages. This graph shows the distribution of all trips and is not separated for each person. 70% of trips made by Type 1 people only had one stage, and approximately 65% of those trips had corresponding OR stages. A quarter of the reported trips had two stages, but only about 35% of those two-stage trips had corresponding OR stages for both stages. The percentage of trips that had corresponding OR stages for each reported PT stage decreases with the increase in number of stages per trip. Therefore, the accuracy of reporting decreases as the trips become more complicated. Use of Gordon's recently developed method to link the separate stages made by an Oyster card user into full, multi-modal journeys (described in detail in Chapter 2) would be a better way to capture accurate records of complicated trips.

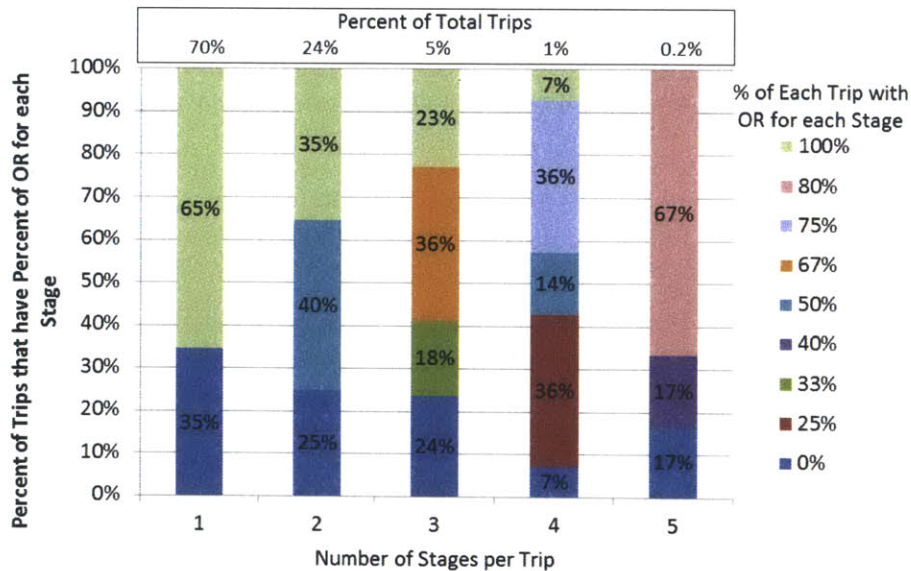


Figure 4-10: Percent of Each Trip that has OR for each Stage by Number of Stages per Trip

<sup>9</sup>This analysis ignored the walk or other non-PT stages reported in each trip.



## 4.7 Stage Analyses by Mode

Next individual's reported stages were analyzed to determine if there were differences in the number of reported stages with OR stages depending on the mode. Figure 4-11 shows the number of people with reported journeys on each mode that match a mode found in that person's OR stages. For example, 800 Type 1 people reported using a bus for one of their stages with 766 (96% of people reporting bus) having OR stages on a bus on their travel day. Underground and NR journeys are similar with 469 out of 486 (97%) people who reported taking a Underground journey having Underground OR stages on their travel day and 194 out of 214 (91%) people who reported taking a NR journey having NR OR on their travel day. The lower percentage of NR journeys found in OR could be because some NR stations are un-gated and it is impossible to tell from OR stages if people actually took a trip between two un-gated stations if they use a period pass and are not required to tap their Oyster card. There were far fewer people reporting journeys on DLR, Tram, and Overground, but most of these people had OR stages on their reported modes of travel. This figure shows that overall, most people who reported using PT on their Oyster card actually used their Oyster card on the same mode as reported on their travel day, and there were no significant differences between bus and train journeys.

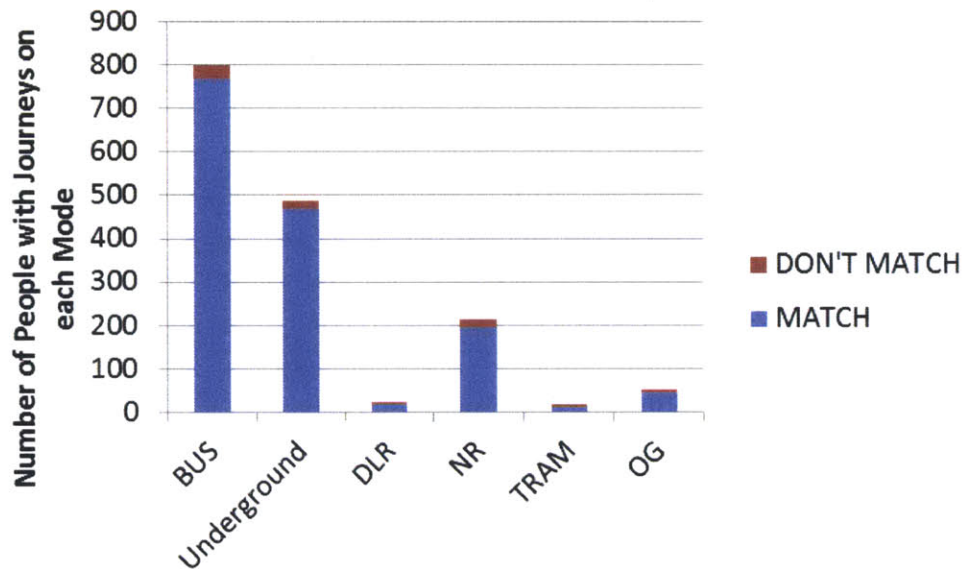


Figure 4-11: Match by Mode

Many people used multiple modes of PT on their travel day. Looking at each person's travel day individually, Table 4.8 shows that 93% of Type 1 people had OR stages on the reported modes for all of their reported journeys. This does not necessarily mean that they had OR on the specific bus route, or O-D pair, but they had OR for every instance of a mode they reported using.

**Table 4.8:** Matching Journeys by Mode

People	All Journeys	Some Journeys	No Journeys	Total Type 1 People
Match by Mode	1,064 (93%)	67 (6%)	17 (1%)	1,148

Table 4.9 shows the level of under or overreporting for Type 1 people with reported bus stages and/or bus stages in their OR stages (Type 1 Bus Sample). Over 53% of the Type 1 Bus Sample had the same number of bus journeys in their OR stages as reported in the LTDS. However, about 17% of the Type 1 Bus Sample overreported bus journeys, and about 31% underreported bus journeys.

**Table 4.9:** Number of Bus Journeys per Person

Reported same number of bus stages as found in OR	Reported more bus stages than OR bus stages (overreported)	Reported bus stages, but no bus OR (overreported)	Reported fewer bus stages than OR bus stages (underreported)	Has bus OR, but no reported bus stages (underreported)	Type 1 Bus Sample
453 (53%)	111 (13%)	34 (4%)	202 (24%)	59 (7%)	859 (100%)

Table 4.10 shows the level of under or overreporting for Type 1 people with reported train stages and/or train stages in their OR stages (Type 1 Train Sample). Approximately 71% of the Type 1 Train Sample had the same number of train journeys in their OR stages as reported in the LTDS, which is about 18% higher than bus journeys. Additionally, about 17% of the Type 1 Train Sample overreported train journeys, and about 16% underreported train journeys, showing that bus journeys are underreported more than train journeys.

**Table 4.10:** Number of Train Journeys per Person

Reported same number of train stages as found in OR	Reported more train stages than OR train stages (overreported)	Reported train stages, but no train OR (overreported)	Reported fewer train stages than OR train stages (underreported)	Has train OR, but no reported train stages (underreported)	Type 1 Train Sample
466 (71%)	77 (12%)	33 (5%)	70 (11%)	30 (5%)	656 (100%)

Bus and train records were then examined to determine if Type 1 people had OR stages on

the specific buses or at the same train stations as they reported in the LTDS. Table 4.11 shows the breakdown of Type 1 people with reported bus stages on their travel day by route and taps on each route. The first row shows how many people had all, some, or none of their reported bus routes in their OR stages. Some people reported taking the same bus route multiple times on their travel day, so the second row shows if all, some, or none of the bus taps on all routes were found in the person's OR stages. This shows that over 60% of Type 1 people reported all of their bus routes relatively accurately, but may have over-reported taking a bus multiple times a day. This could possibly be because they reported a trip to work in the morning and then reported making the opposite journey in the evening, when they actually took a different bus or mode. Also, due to the dense PT network in London, people have many bus route options to reach their destination, so people may forget the specific bus they used that day because they used whichever bus came first. Therefore, if the use of specific bus routes is important, fully linked Oyster journeys are a more accurate account of the bus routes used for each person than survey responses.

**Table 4.11:** Matching Bus Routes and Taps per Person

	All Reported in OR	Some Reported in OR	No Reported in OR	Type 1 People with Reported Bus Stages
People with X Route (s) in OR	493 (62%)	151 (19%)	156 (20%)	800
People with X Taps (s) in OR	366 (46%)	278 (35%)	156 (20%)	800

Figure 4-12 shows the percentage of people with matching OR stages for each reported train mode by whether or not the full journey, just the origin, just the destination, or some combination of the journey is found in the OR stages. The people who had some combination of a journey are people who may have had the origin of one journey and the destination of another journey, but not the full O-D pair of a reported stage. Depending on the type of train used, 35-58% of people did not report the same O-D pairs captured by their OR stages. O-D pairs for Underground journeys were matched most accurately at 65% of people with the full reported Underground journey found in the OR stages. If measuring travel between specific O-D pairs for the LTDS sample is important, fully linked Oyster journeys are a more accurate record of the O-D pairs for each person than the survey responses.

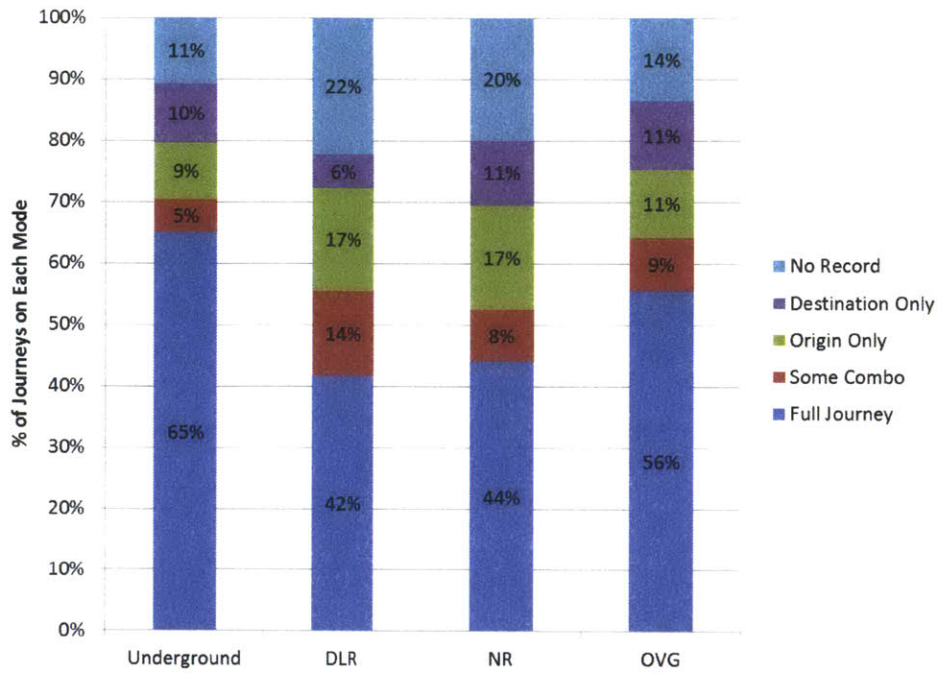


Figure 4-12: Matching Train Journeys

## 4.8 Stage Start Time Analyses

The LTDS asks respondents for the start and end time of each trip, the estimated duration of each walk stage, and the specific start and end location of each stage within the trip. All trips are made up of at least one walk mode and most trips are made up of one other mode of travel with walk modes before and after. For example many PT journeys are made up of three stages:

1. A walk stage from home to the origin bus stop or Underground station
2. A bus or Underground stage
3. A walk stage from the destination bus stop or station to the ultimate trip destination

The agency contracted to enter the manually collected survey data into a computer database also estimates the stage durations for each trip using a transportation modeling program. Inputs to the program are the reported trip start and end times, specific stage start and end locations<sup>10</sup>, reported walk durations<sup>11</sup>, and average travel speeds for each mode. The program has routing capabilities to calculate distances between the start and end location via the existing road network and estimates stage durations using the average speed for each mode. Improbable journeys are flagged and analyzed manually to determine if there were any reporting or transcription errors. Follow-up interviews are sometimes necessary

<sup>10</sup>The program has detailed geographic data that is used to locate specific start and end locations.

<sup>11</sup>Walk durations are sometimes adjusted if it is unlikely the walk distance could be covered in the reported time.

to clarify trip characteristics, which adds more time and cost to the already high marginal cost of the survey.

The wait time for PT modes is included in the journey travel time in the LTDS. For example, if a person reported leaving home at 8:00 a.m., walking to a bus stop for five minutes, and traveling on the bus to reach his destination at 9:00 a.m., the time the bus stage is assumed to have started is 8:05 a.m. and the waiting time for the bus is included in the 55 minute journey time. The same assumptions apply for train modes in the LTDS. The start time captured by the Oyster system is the time a person boards a bus or enters a train station. Therefore the start times for train modes are directly comparable across the two data sets because the start time for the LTDS and Oyster system is when the person enters (taps into) a train station. The wait time for the train is counted in both data sets after the inferred (LTDS) or captured (Oyster) stage start time. However, the Oyster start time for bus modes is captured after the person has already waited for the bus, while the LTDS bus stage start time is when the person arrived at the bus stop before he has waited for the bus to arrive. Therefore it was expected that Oyster start times for bus stages would be slightly later than inferred bus stage start times in the LTDS.

All reported stage start times in LTDS were compared with all OR stage start times for Type 1 people to see if there were any differences in the reported peaks of PT travel and the PT travel peak times as captured by the Oyster system. The analysis was done on all LTDS PT stage start times and OR stage start times for Type 1 people regardless of whether or not their journeys matched by mode, route, or O-D pair. Figure 4-13 shows the number of people with reported stage start times in the LTDS and the OR stage start times for all Type 1 people in 10 minute intervals.

All Type 1 people had journeys that started throughout the day with slightly higher peaks in the morning and evening. The reported stage start times were more peaked than the actual OR stage start times because many of the reported start times were right on the hour, while the OR start times were spread more evenly. To show this more clearly, the number of reported start times were compared to the number of OR start times in ten minute intervals (chosen to span five minutes before and five minutes after the hour and increasing by ten minutes) throughout the day. Figure 4-14 shows that there were many more reported start times on the hour or half hour than actual OR start times, especially for 8:00, 10:00, 11:00, 13:00, 15:00, 17:30, 18:00, 20:00, and 24:00.

The difference in start times for each PT stage that had a corresponding OR stage was then calculated. A corresponding OR stage for a bus stage is a tap on the same bus route as reported in the LTDS. A corresponding OR stage for a train stage is an entry tap at the same reported origin and an exit tap at the same reported destination station and therefore corresponds to the entire O-D pair. Start times for partial segments were not analyzed because they cannot be assumed to be the same journey as reported in the LTDS. Start time differences ranged from OR stages that were approximately 13 hours earlier than the reported stage, to OR stages that were approximately 15 hours later than the reported stage with an average difference of 61.2 minutes and standard deviation of 151.7 minutes. It is unlikely that a person would misreport the start time of a journey by more than two

hours, so for this part of the analysis, any stage with a difference of more than two hours was excluded, but this only represented approximately 10% of stages. Figure 4-15 shows the distribution of start time differences on each mode for journeys with OR stages within two hours of the reported starting time. This shows that train journeys were reported within 10 minutes of the OR stage start time about 50% of the time and bus journeys were reported within 10 minutes of the OR stage start time about 38% of the time. One possible explanation for this difference is that the bus start time in the LTDS does not take into account waiting time and is just the time the person arrived at the bus stop, so even if the person reported their trip start time accurately, the OR stage time should be later than the LTDS start time for bus stages. Overall, start times were reported within 10 minute accuracy less than 40% of the time. Therefore, one could conclude that start times from fully linked OR stages are a more exact way to determine the actual PT start times for people with matching journeys.

Figure 4-16 shows the distribution of the start time differences for each reported LTDS trip (stages from multi-stage trip all are categorized as having the same purpose) destination purpose for journeys with OR stages within three hours<sup>12</sup> of the reported starting time. A description of each destination trip purpose is included in Appendix A. The median start time differences ranged from about nine minutes for Work purposes to 24 minutes for Social purposes. Many purposes had a wide distribution of start time differences, especially for the Shopping, Personal, Social, and Home purposes. The destination purposes with the smallest distribution of start time differences were Medical, Work, and Worship. Some interesting results were the start time differences between Work (usual place of work) and Work Other (work related, but to location other than usual place of work) and Work and Home. The difference of start times were much smaller for stages going to the usual place of work than for going to places other than the usual place of work or home. This shows that people were generally good estimators of their start times for work, which is often a fixed start time throughout the course of the week, but were not as good at estimating start times for journeys for work to places other than their usual place of work, and also not as good at estimating when they leave work or other locations to go home. People were also relatively good at estimating their start times for medical appointments and journeys to worship because these also generally have a fixed starting time that a person is likely to remember.

---

<sup>12</sup>Three hours was chosen to show a large range of start time differences by trip purpose.

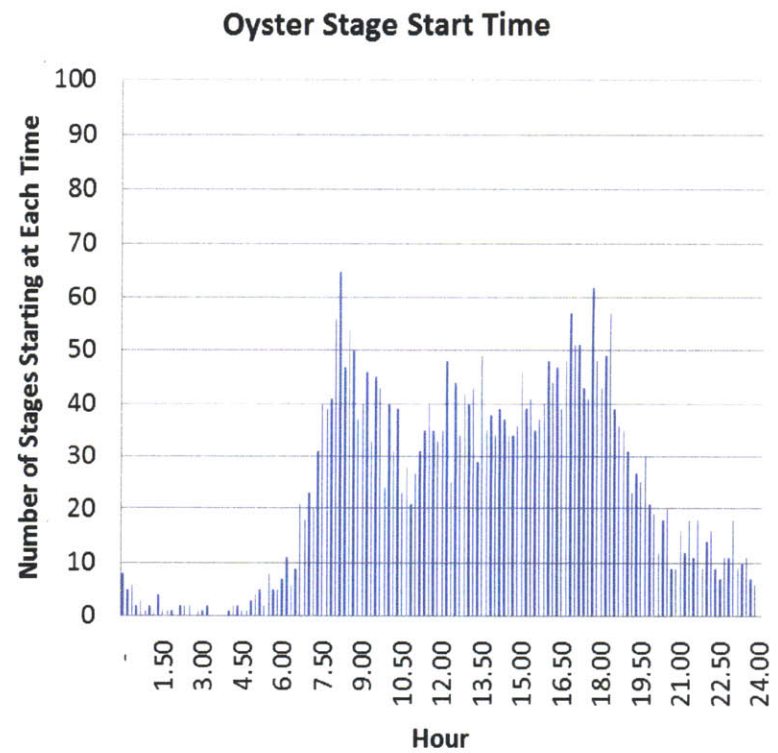
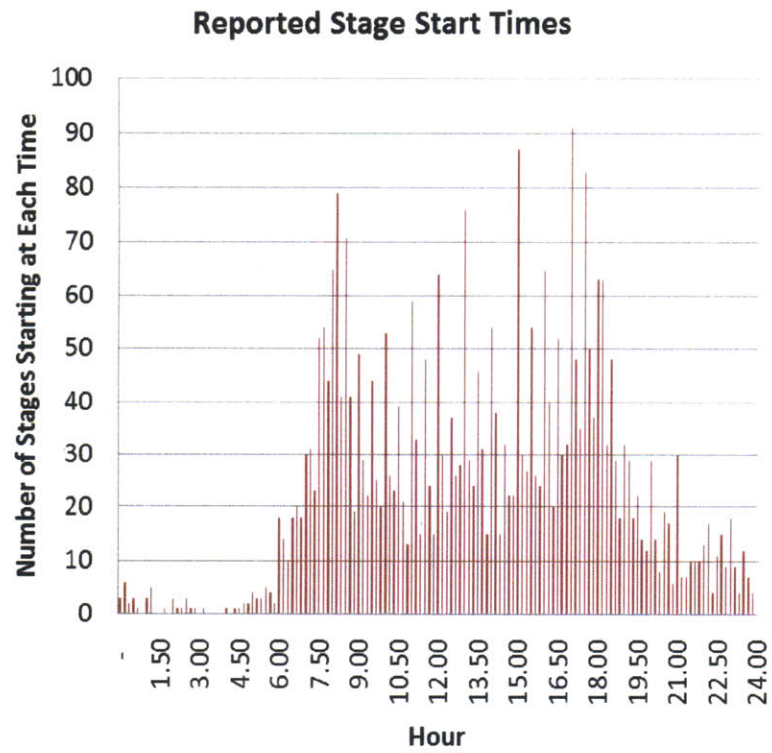


Figure 4-13: Stage Start Time, Type 1: 10 Min. Interval

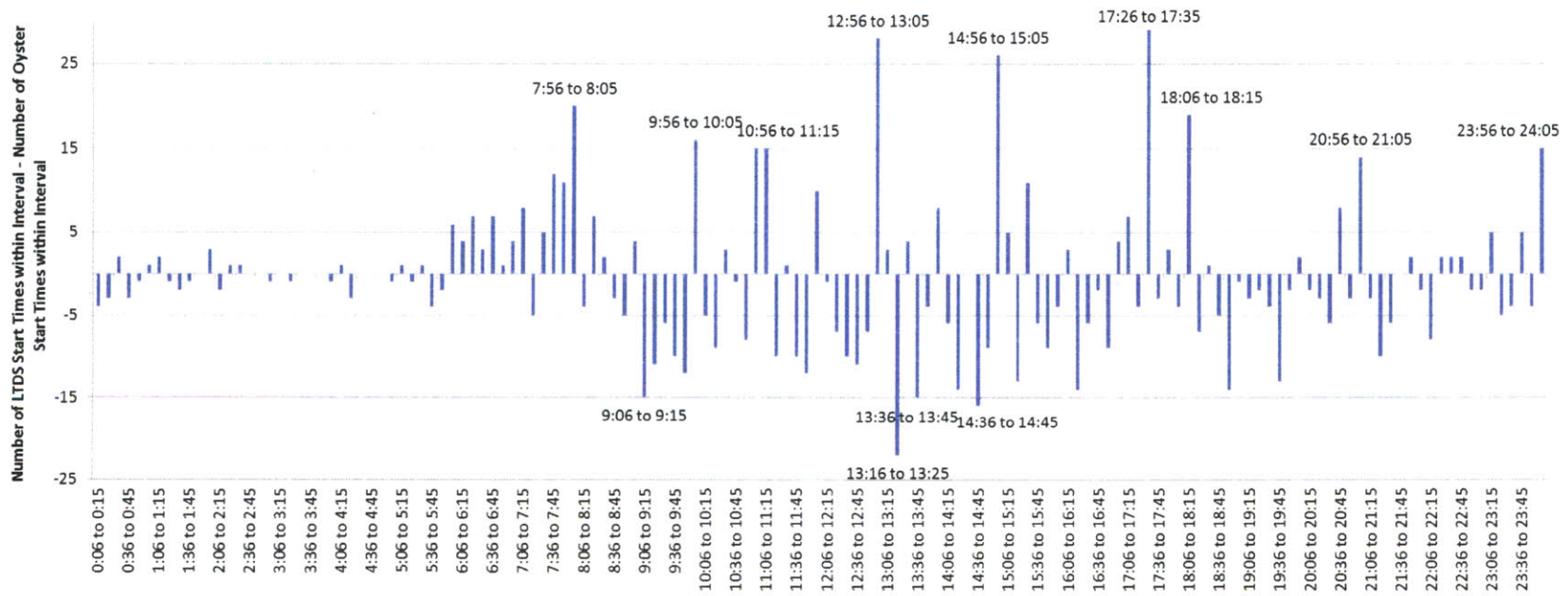


Figure 4-14: Number of LTDS Start Times within Interval minus Number of Oyster Start Times within Interval



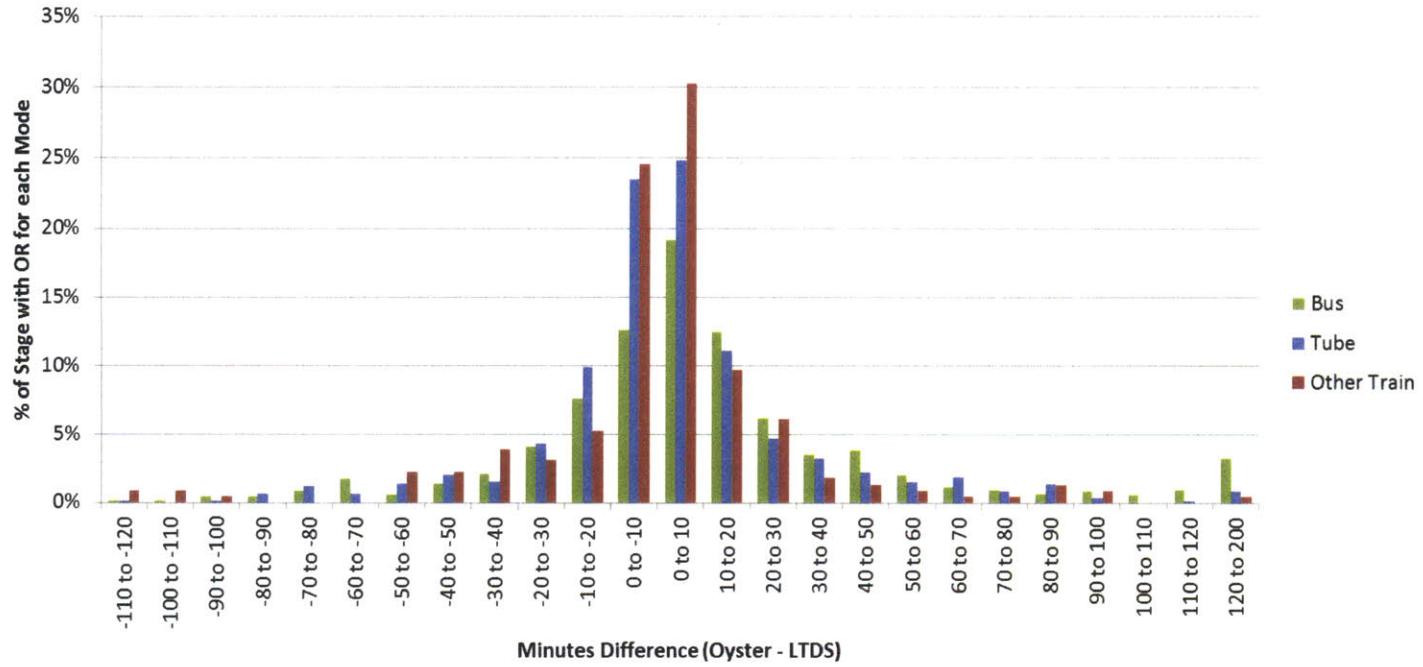
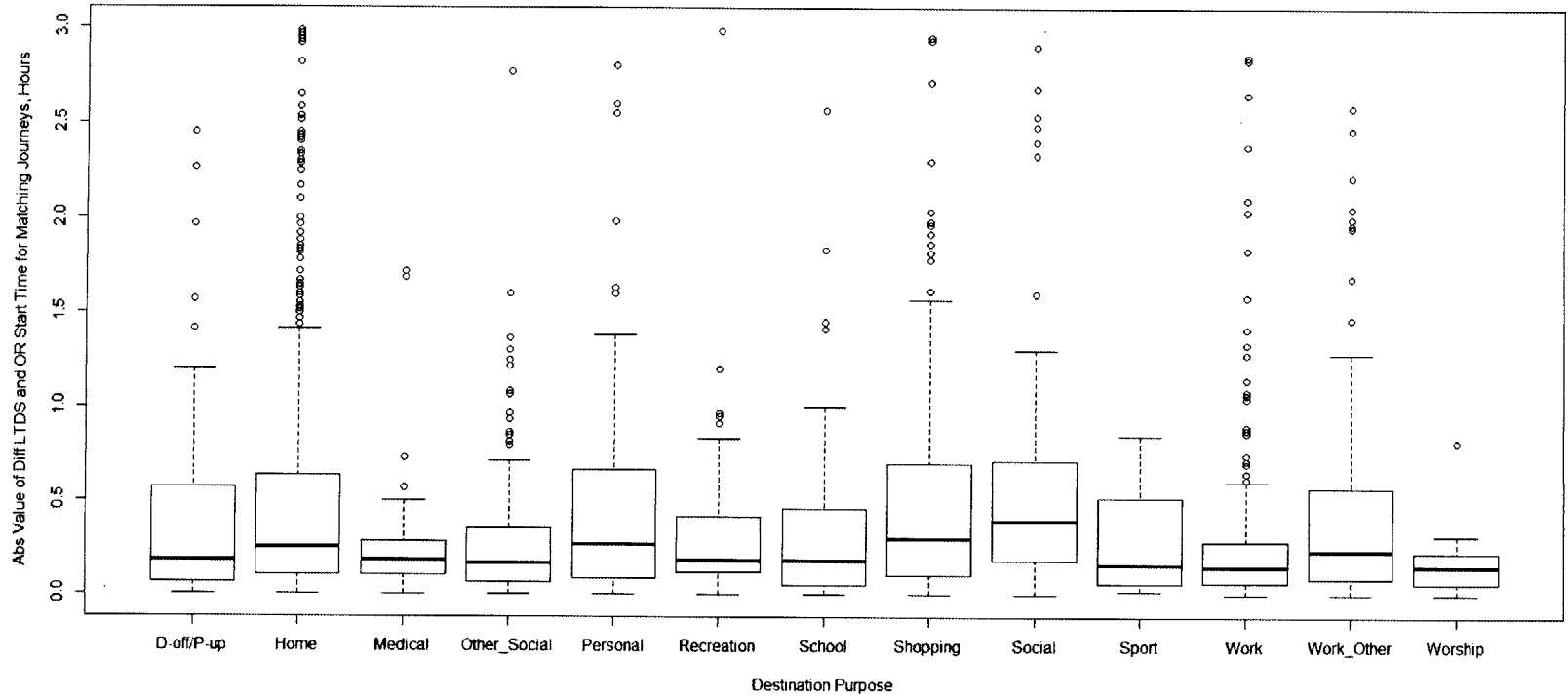


Figure 4-15: Difference in Stage Start Times for Matching Journeys with OR within two hours of Reported Start Time



**Figure 4-16:** Abs. Value of Diff. in Stage Start Times for Matching Journeys with OR (within three hours) by Trip Destination Purpose

## 4.9 Stage Durations

As described in section 4.8, stage durations for each trip are estimated in the LTDS using a transportation modeling program with reported trip start and end times, specific stage start and end locations and estimated travel distances, reported walk durations, and average travel speeds for each mode. Durations are validated by the third party survey agency for feasibility on the transportation network. Improbable journeys are flagged and analyzed manually to determine if there were any reporting or transcription errors. Follow-up interviews are sometimes necessary to clarify trip characteristics.

Durations for journeys captured by the Oyster system are calculated in the raw Oyster data as the difference of the completed exit time and completed entry time for train journeys, which includes the walk time from the entry turnstile to the platform, the wait time for the train, the train running time, and the walk time from the destination platform to the exit turnstile. The wait time for train OR stages is included in the journey duration because the Oyster system captures the time a person tapped into the entry station and then tapped out of the exit station and the entire duration includes wait time on the platform. This is consistent with the way wait time is included in the LTDS stage durations, as described in section 4.8.

The range of inferred durations from the LTDS survey were compared to the actual durations found in OR stages for matching train journeys and is shown in Figure 4-17. The median durations were very close, but the inferred durations varied more than the Oyster durations.

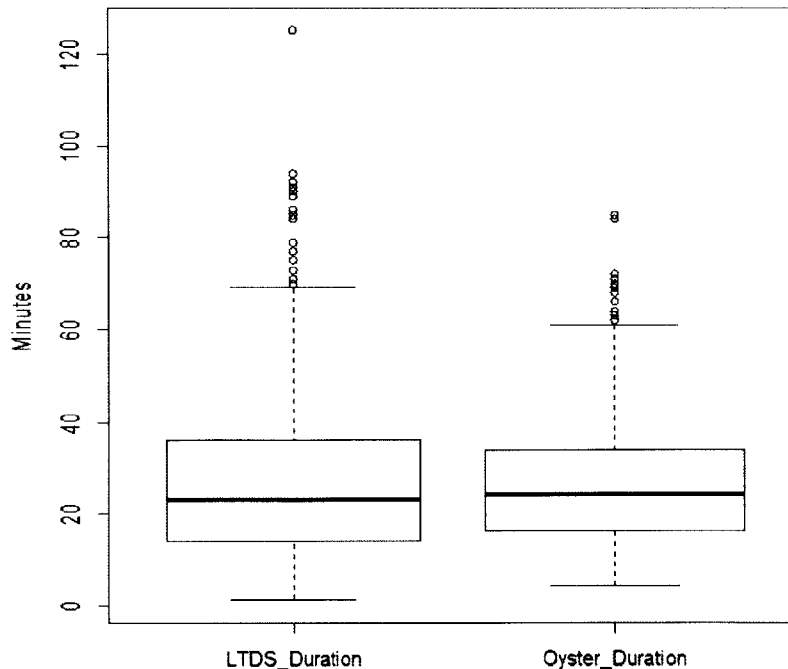
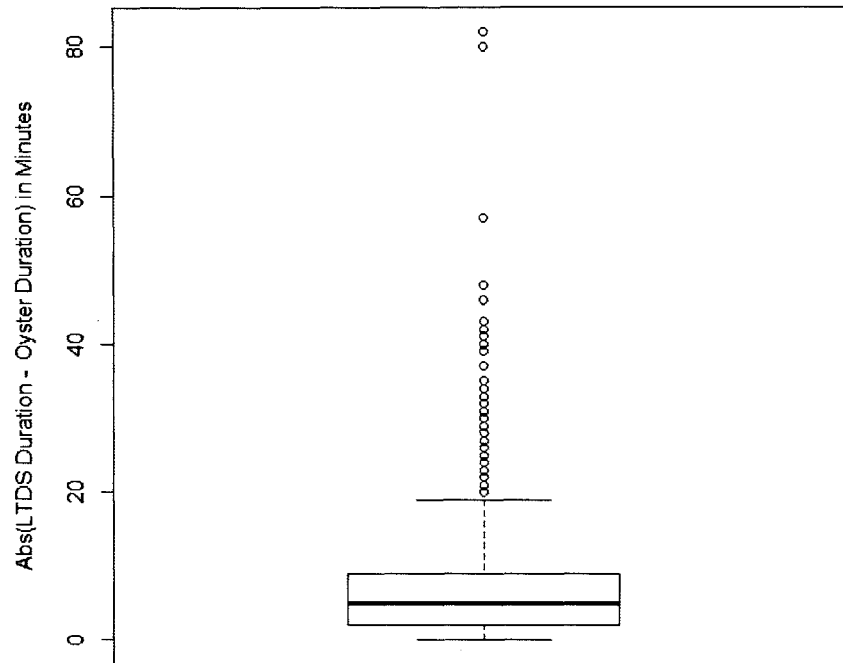


Figure 4-17: Duration Difference for Matching Train Journeys

Additionally, the absolute value of the difference in LTDS calculated duration and duration captured by the Oyster system for matching train stages is shown in Figure 4-18. The average and median differences were about seven and five minutes respectively, with differences ranging from zero to 80 minutes. There are many outliers in this distribution, showing that for many people, the durations calculated by the LTDS process varied greatly from the durations captured by the Oyster system, which is a much more direct measure of journey duration.



**Figure 4-18:** Absolute Value of Difference in LTDS and Oyster Durations for Matching Train Journeys, in Minutes

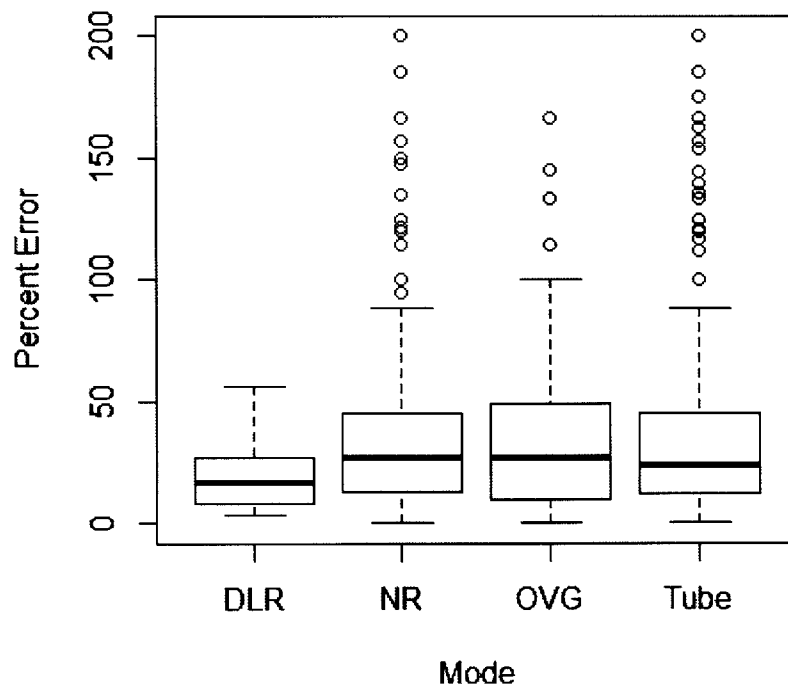
Alighting a bus is not captured in the Oyster system because people only tap their Oyster cards upon boarding the bus and there is no distance-based fare for bus travel in London. Therefore, analysis of Oyster bus journey durations was not possible in this thesis. However, the method recently developed by Gordon using TfL's Oyster system and automatic vehicle location (AVL) systems to infer bus trip origins and destinations, and link the separate stages made by an Oyster card user into full, multi-modal journeys could be used in the future. Bus alighting locations are assumed in this program by looking at the next trip of a user and assigning the alighting location to the stop on the bus route nearest to the next boarding stop or station entry. Durations for bus journeys for each person are estimated in this program by matching the time stamp of the fare transaction record with time stamps of events in automatic vehicle location data.

For PT journeys, durations calculated using Gordon's method to infer fully linked journeys captured by the Oyster system for each user are much more reliable estimates of actual durations than average speeds and survey respondents' perceptions of start times and walk durations. This program could be used in the future to calculate exact journey start and

end times, durations, and speeds for Type 1 people to validate and enhance the reported PT stages in the LTDS survey.

#### 4.10 Average Speeds used in LTDS

As described earlier, there is only one average speed used for each mode to help infer the journey stage durations in the LTDS regardless of the type of service. For example, NR journeys all have the same speed for both local and express services. Buses have the same speed regardless of the congestion in the network. Figure 4-19 shows the percent error of LTDS average speeds compared to Oyster speeds for each type of train mode for matching one stage train journeys (including journeys where the person had an interchange would likely confound the speed calculation). This figure shows that the inferred speeds from LTDS reports varied greatly from the speeds calculated by the Oyster system, which uses same distance as the LTDS inference, but has exact entry and exit times and is a much more accurate estimation of modal speeds. The average speeds used by the LTDS system are a primary input to the duration calculation for each stage, which could also be more accurately estimated using fully linked Oyster journeys for each person who reported using an Oyster card on their travel day and took PT journeys.



**Figure 4-19:** Percent Error of Speed (LTDS - Oyster/Oyster) for Matching One Stage Train Journeys

## 4.11 People with Perfectly Matching Stages

There were only 376 people (22% of the PT Sample) with captured OR stages perfectly matching reported PT stages. A perfect matching means they had matching OR stages in the same chronological order for each reported PT stage<sup>13</sup>. These people had an average of 2.59 PT stages with the distribution of number of stages shown in Figure 4-20 using the modes shown in Figure 4-21.

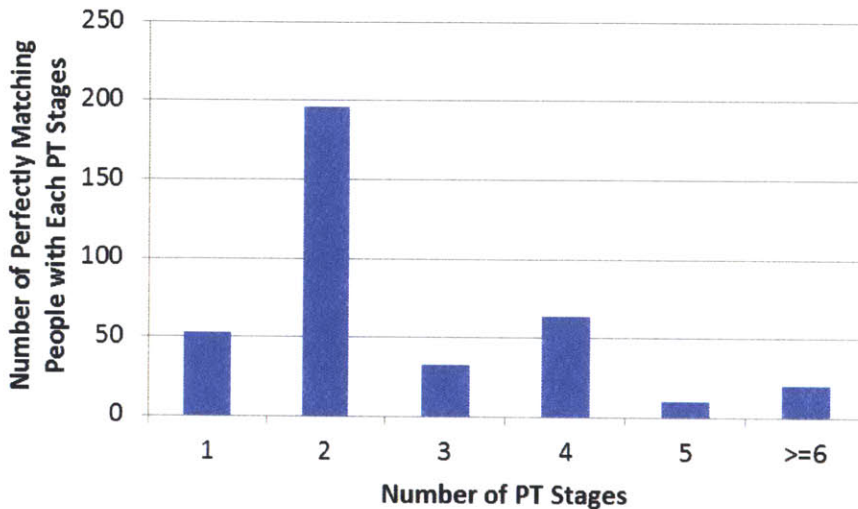


Figure 4-20: Distribution of PT Stages for Exact Journey People

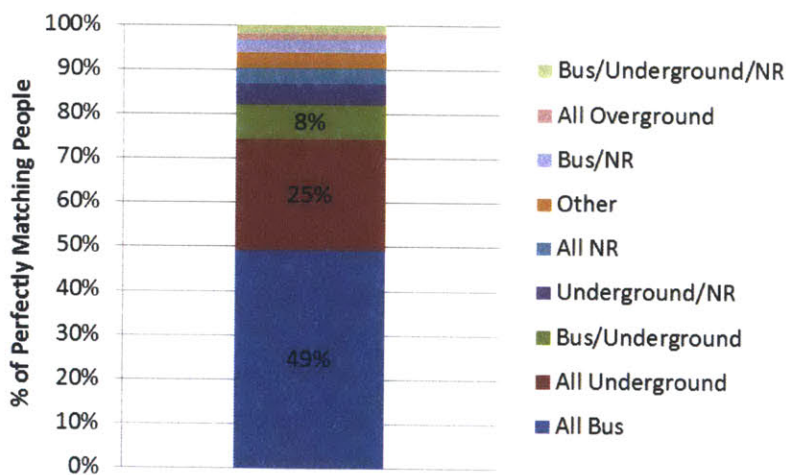


Figure 4-21: Percent of People with Exact Journey Sequence by Modes Used

Figures 4-20 and 4-21 show that most people who had OR for the exact journey as reported

<sup>13</sup>There was one additional person who had the same LTDS and OR stages, but the sequence of OR stages was different than the reported sequence of LTDS stages

in the LTDS had two stages on bus or Underground on their travel day. Many of these people had journeys on one bus route or between one train O-D pair in the morning, and a journey in the opposite direction in the evening, which corresponds to the previous findings that simple journeys were easier to recall than journeys made up of multiple complicated stages.

These perfectly matching journey people made up only 24% of all people who reported PT use on Oyster on their travel day. This highlights the difficulty with combining the survey responses with Oyster data after the survey has taken place. The OR stages can be used for these 376 people to enhance the reported PT stages with accurate stage start and end times, but this is obviously a very small sample compared to all of the people who have reported PT stages on their travel day. The other Type 1 people's reported PT stages could be enhanced with accurate start and end times where there are corresponding OR stages, if they were within a reasonable time frame and journey sequence, but this may require many assumptions about whether the captured OR stage was the PT journey the respondent was referring to. This suggests that it would be more useful and effective to combine Oyster data earlier in the process, such as during the interview to take advantage of the continuous collection of travel behavior data for people who use Oyster cards. This conclusion will be discussed in more detail in Chapter 6.

## 4.12 Summary of Type 1 Analyses

With the voluntary inclusion of Oyster card numbers in the LTDS, the accuracy of PT stages was determined by comparing an individual's reported PT stages with their PT stages captured by the Oyster system. This was done most effectively for the sample of people who volunteered their Oyster card number, reported taking public transport (PT) journeys on their travel day, and had valid Oyster record (OR) stages on that day, the category of survey respondents defined as Type 1 people. Type 1 people were representative of all people who reported taking PT stages on their travel day in terms of the geographic distribution in each borough as well as many demographic and reported trip mode share characteristics. There were some demographic differences that should be kept in mind when considering how the results of the panel of Type 1 people can be expanded for all people who reported LTDS PT stages on their travel day. Type 1 people were not different than the other people who reported PT in terms of how they reported their travel behavior, but they actually traveled more on PT using their Oyster card than they reported. This result suggests that the other people who reported PT, but did not volunteer an Oyster card or have valid OR stages on their volunteered Oyster card on their travel day may also have actually traveled more than they reported on their travel day, and therefore the single day survey underestimated the intensity of PT use on days where the survey sample chose to use the PT mode.

Overall, Type 1 people underreported PT stages by 8.9%. Weekday stages were underreported by about 9.2% and weekend stages were underreported by about 7.5%. By mode, Type 1 people underreported bus stages by 15.6% and reported the same number of train stages as found in OR stages.

Overall, the distributions of reported PT stages and OR stages on the travel day for Type 1 people with weekday travel days were very similar. However, the distribution of the average weekday OR stages over all days where the person had weekday OR stages had a higher median and a narrower range than the reported or OR stages on the travel day, and the average weekday OR stages over all possible weekdays of OR collection had a lower median and a narrower range than the travel day. This shows that people did not use PT every weekday throughout the OR analysis period, but on the weekdays they used PT, they used it more than on their travel day. The results were similar for Type 1 people whose travel day was a weekend, but with an even larger difference between the reported number of PT stages on the weekend travel day and the overall PT use on the weekend days. Thus one could reasonably conclude that at least for PT travel, the single day survey overestimates typical PT use overall, especially on weekend days, but it underestimates the intensity of PT use on days when the survey sample chose to use the PT mode.

The LTDS reported frequency of PT use was higher and had a tighter distribution than the actual PT use found in OR stages for Type 1 people, showing that the LTDS is not capturing the actual variability of PT use for the survey sample. It is not clear how the frequency of use question is used in the TfL modeling process, but if it is used to scale up the frequency of PT use for all London residents, it may be overestimating the number of PT journeys people are making.

There were only 376 people Type 1 people (22% of the PT Sample and 33% of Type 1 people) with perfectly matching reported LTDS PT stages and OR stages. Their stages made up 18.4% of the PT Sample's reported LTDS stages or 24.5% of Type 1 people's PT stages. The remaining 67% of Type 1 people either underreported, overreported, or misreported (or some combination of under, over, and/or misreporting) their PT stages. Approximately 12% of Type 1 people have likely misreported the fare media they used on their travel day, which made up 6.3% of Type 1 people's reported LTDS stages. This is not considered a serious problem, as the reporting of the type of fare media used may have specific implications at TfL, but will not affect the overall level of PT use reported in the LTDS.

About 49% of Type 1 people have likely misreported some portion(s) of the train O-D pair or bus route(s) used, which made up approximately 28.2% of Type 1 people's reported LTDS stages. This will have implications for the LTDS survey, if the specific origins and destinations or bus routes reported are being used for planning purposes. About 10% of Type 1 people overreported their PT stages, most likely in terms of the travel day, that is they may have reported a day other than their agreed upon travel day because they thought was more representative of their regular travel behavior or for other reasons, which made up approximately 5.3% of Type 1 people's reported LTDS stages. Approximately 29% of Type 1 people underreported their PT stages, which made up approximately 20.0% of the Type 1 people's reported LTDS stages. The overall over and underreporting of PT stages will have an impact of the overall level of PT use reported in the LTDS. When taking into account the under and overreported PT stages, 489 stages (666 Underreported - 177 Overreported) should be added to the PT Sample reported LTDS PT stages.



The percentage of trips that had corresponding OR stages for each reported PT stage decreases with the increase in number of stages per trip. Therefore, the accuracy of reporting decreases as the trips become more complicated. Use of Gordon's recently developed method to link the separate stages made by an Oyster card user into full, multi-modal journeys could be a better way to capture accurate records of complicated trips.

93% of Type 1 people had OR stages on the reported PT modes for all of their reported journeys. This shows that overall, most people who reported using PT on their Oyster card actually used their Oyster card on the same mode as reported on their travel day, and there were no significant differences between bus and train journeys. Additionally, over 60% of people had OR stages on the same bus routes reported in the survey, but only 46% of people had OR stages on all bus routes and taps reported in the survey and therefore reported all of their bus routes accurately, but may have over-reporting taking a bus multiple times a day. Therefore, if the use of specific bus routes is important, fully linked Oyster journeys are likely to be a more accurate record of the bus routes used for each person than the survey responses. Depending on the type of train used, 35-58% of people did not report the same O-D pairs captured by their OR. O-D pairs for Underground journeys were matched most accurately at 65% of people with the full reported Underground journey found in the OR. If travel between specific O-D pairs is important, fully linked Oyster journeys are likely to be a more accurate record of the O-D pairs for each person than the survey responses.

LTDS reported stage start times were more peaked than the OR stage start times because many of the reported start times were right on the hour, while the OR stage start times were spread more evenly. Train journeys were reported within 10 minutes of the OR start time about 50% of the time and bus journeys were reported within 10 minutes of the OR start time about 38% of the time. Additionally, inferred speeds and durations from LTDS reports varied greatly from the speeds and durations estimated by the Oyster system. Gordon's method to fully link Oyster journeys for each person who reported using an Oyster card on their travel day and took PT journeys could be used to calculate exact journey start and end times, durations, and speeds for Type 1 people to validate and enhance the reported PT stages in the LTDS survey. This method could also be used in certain cases to estimate bus waiting times, a conclusion that will be discussed in more detail in Section 6.

OR stages and Gordon's recently developed method to link the separate stages made by an Oyster card user into full, multi-modal journeys can only be done for people who had OR stages for the exact reported sequence of LTDS PT stages. These perfectly matching journey people made up only 22% of the PT Sample. This highlights the difficulty with combining the survey responses with Oyster data after the survey has taken place. The other Type 1 people's reported PT stages could be enhanced with accurate start and end times where there are corresponding OR stages, if they were within a reasonable time frame and journey sequence, but this may require many assumptions about whether the captured OR stage was the PT journey the respondent was referring to. This suggests that it would be more useful and effective to combine Oyster data earlier in the process, such as during the interview, to take advantage of the continuous collection of travel behavior data for people who use Oyster cards. This conclusion will be discussed in Chapter 6.



## Chapter 5

# Analyses of Other Person Types

With the voluntary inclusion of Oyster card numbers in the LTDS, the accuracy of PT journey stages was determined by comparing an individual's reported LTDS PT journeys with their PT stages captured by the Oyster system. This was done most effectively for the people who volunteered their Oyster card number, reported using PT on their travel day, and had valid<sup>1</sup> OR stages on that travel day - the people categorized as Type 1 people. However, there were other conclusions that were drawn from the analysis of the other types of people who volunteered their Oyster card number, but did not report using their Oyster card for PT stages and/or did not have valid OR stages on their travel day (Types 2 - 9). This chapter focuses on those results.

### 5.1 Type 2 Analyses

Type 2 people (approximately 7% of all people who volunteered an Oyster card) reported 765 LTDS PT stages using their volunteered Oyster card(s), but did not have valid OR stages on their card(s) on their travel day. They did, however, have OR stages on their card(s) on other days within the OR analysis period meaning they used their card(s) sometimes, but may have reported PT stages from a day other than their travel day, or used a different card on their travel day. The LTDS interviews respondents on all days of the week and captures travel day information for weekdays as well as weekends. Approximately 74% of Type 2 people had a weekday travel day, which is lower than average, meaning slightly more Type 2 people had weekend travel days than other people who reported LTDS PT stages on their travel day using their volunteered Oyster card(s) (Types 1 and 3). Demographic data for Type 2 people is presented in Appendix B.

#### 5.1.1 Type 2 Intensity of PT Use

The OR stages for Type 2 people were collected eight weeks prior to the travel day<sup>2</sup> and indefinitely after the interviews, though this thesis only analyzed OR stages through 18

---

<sup>1</sup>A valid Oyster record stage is any type of PT use (not topping up) on the Oyster card and may or may not be the exact stage as reported in the LTDS.

<sup>2</sup>It is possible to retrieve eight weeks of Oyster data retroactively.

June, 2012. These records were analyzed to find the average number of OR stages for each Type 2 person for the entire OR analysis period. Figures 5-1 and 5-2 show side by side comparisons of the reported PT stages and OR stages for all Type 2 people with travel days on weekdays or weekend days, respectively. The boxplot on the far left (Avg WDayOR Cap WDays) is the distribution of the number of weekday OR stages captured by the Oyster system over all weekdays where the person had OR stages. This represents the distribution of people's average OR stages per weekday that they took PT, as captured by the Oyster system. The boxplot second from the left (Avg WDayOR Pos WDays) shows the distribution of the average number of weekday OR stages over all weekdays where OR stage capture was possible (eight weeks prior to the individual's travel day through 18 June, 2012). The third boxplot from the left (Avg WEndOR Cap WEnds) shows the distribution of the number of weekend OR stages captured by the Oyster system over all weekend days where the person had OR stages. This represents the distribution of people's average OR stages per weekend that they took PT, as captured by the Oyster system. The boxplot fourth from the left (Avg WEndOR Pos WEnds) shows the distribution of the average number of weekend OR stages over all weekend days where OR stage capture was possible (eight weeks prior to the individual's travel day through 18 June, 2012). The boxplot second to the right shows the distribution of reported LTDS stages for people on their travel day, and the boxplot on the far right (Oyster on Travel Day) shows the distribution of OR stages for people on their travel day, which was 0 for all Type 2 people.

Figure 5-1 shows that the distribution of average weekday OR stages over all weekdays where weekday travel day people had OR stages had a higher median and a narrower range than the reported LTDS stages on the travel day, and the average weekday OR stages over all possible weekdays of OR collection had a lower median and a narrower range than the travel day. The weekend results were similar with slightly lower medians and narrower distributions than the weekday results. This shows that overall, people with weekday travel days did not use PT every weekday throughout the OR analysis period, but on the weekdays they used PT, they used it more on average than they reported on their travel day.

Figure 5-2 shows that the distribution of average weekend OR stages over all weekend days where weekend travel day people had OR stages has a similar median and range as the reported LTDS stages on the travel day, and the average weekend OR stages over all possible weekend days of OR collection had a lower median and a narrower range than the travel day. The weekday results were similar with slightly higher medians and wider distributions than the weekend results. This shows that overall, people with weekend travel days did not use PT every weekend throughout the OR analysis period, but on the weekend days they did use PT, they used it similarly on average to what they reported on their travel day.

This shows that at least for PT travel, the single day survey overestimated typical PT use overall, especially for people with weekend travel days, and it underestimated the intensity of PT use on days when Type 2 people with weekday travel days chose to use the PT mode.

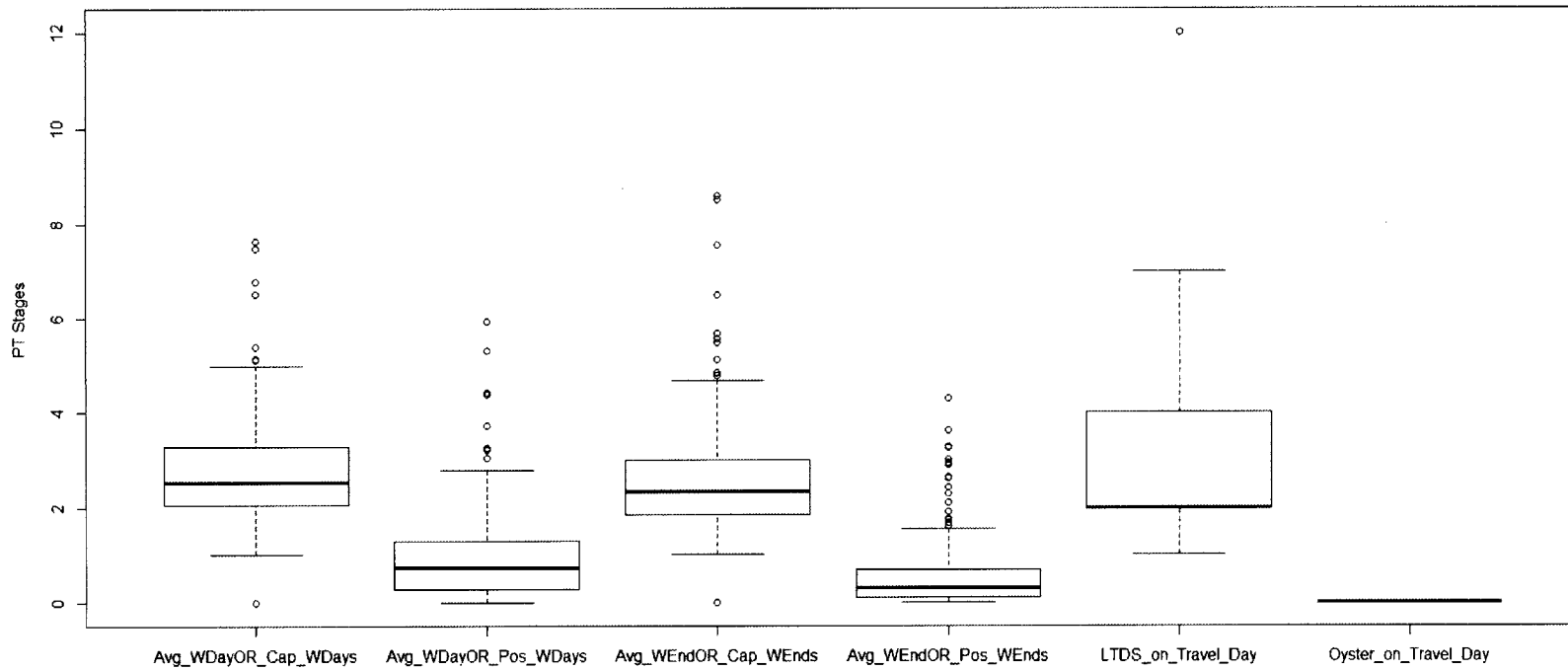


Figure 5-1: LTDS vs. Oyster Stages for Type 2 People with Weekday Travel Days

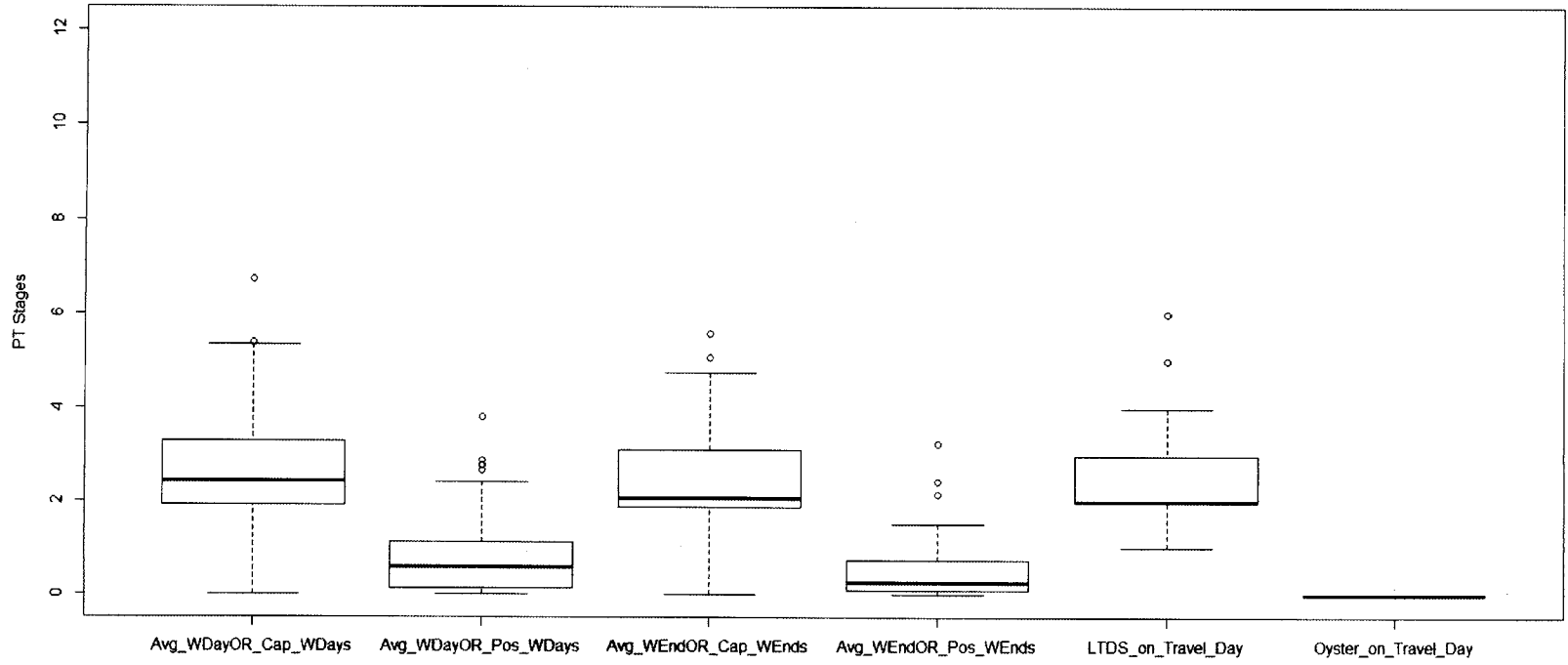


Figure 5-2: LTDS vs. Oyster Stages for Type 2 People with Weekend Travel Days

### 5.1.2 Type 2 Reported LTDS Stages found in OR within Previous Week

Due to the length of the survey, there is no reason for a person to report a PT journey they have never taken, so it is likely that people who reported LTDS stages that had no corresponding OR stages used a different Oyster card on their travel day or reported a day other than their travel day that they believed was more representative of their normal travel behavior. All Type 2 people reported LTDS PT stages on their travel day that were not found in the Oyster system. Analysis of Type 2 people's LTDS PT and OR stages shows that about 45% of Type 2 people (132 people) has some or all (in terms of the specific bus route or train origin-destination (O-D) pair) of their 338 reported LTDS PT stages in their OR stages within the previous week of their travel day. About 20% had only some of their reported LTDS PT stages (194 stages or 25.4% of Type 2 LTDS PT stages) in OR stages within the previous week of their travel day, about 10% had all of their reported LTDS PT stages (61 stages or 8.0% of Type 2 LTDS PT stages) in OR stages within the previous week, but may have had some extra OR stages on that day, and about 15% had the exact LTDS PT stages in the same sequence (83 stages or 10.8% of Type 2 LTDS PT stages) within one week before their travel day.

This suggests that at least 45% of Type 2 people, or 8% of the total PT Sample, reported a different day they may have felt was more representative of their travel behavior and did, in fact, overreport their 338 PT stages (7.6% of the PT Sample reported LTDS stages) on their travel day. The other 55% of Type 2 people (160 people) had none of their 427 reported LTDS PT stages (9.6% of the PT Sample reported LTDS stages) in OR stages within the previous week of their travel day, and therefore most likely misreported the fare media used on the travel day. Therefore interviewers should be sure to record the most used Oyster card for each individual, and be sure to ask about PT stages the person may have taken without having to tap their Oyster card if automatic records are used during the interview process in the future. A summary of this categorization of Type 2 people is shown in Figure 5-3 below.

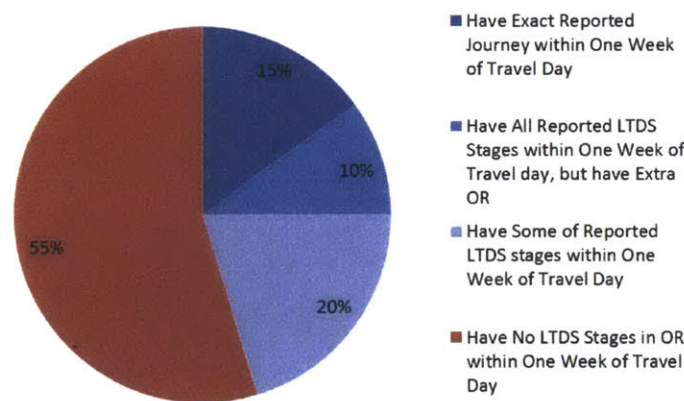


Figure 5-3: Type 2 Oyster Categories

## 5.2 Type 3 Analyses

Type 3 people (117 people) reported 351 PT stages using their volunteered Oyster cards, but did not have valid OR stages on their travel day or any days within the OR analysis period. There is no obvious reason for these people to report travel they have never taken, so it is likely that these people (about 7% of the PT Sample) actually took these 351 PT stages (7.9% of the PT Sample reported LTDS stages), but used a different Oyster card or fare media on this travel day. It is a conservative estimate that all Type 3 people are assumed to only have misreported the fare media because it is likely, given the results of the other person types, that they also overreported, underreported and/or misreported (in terms of bus routes and train O-D pairs) some of their PT stages as well.

Figure 5-4 shows the distribution of number of reported PT stages for Type 3 people compared to all people who reported PT stages using their volunteered Oyster card. The distributions are almost identical, and there is no reason to believe that Type 3 people are any different in terms of travel behavior than other people who reported PT stages, with the exception that Type 3 people most likely misreported their Oyster card number. Therefore interviewers should be sure to record the most used Oyster card for each individual. Demographic data for Type 3 people is presented in Appendix B.



**Figure 5-4:** Reported PT Stages of Type 3 People Compared to All People who Reported PT

## 5.3 Types 6 and 9 Analyses

Type 6 people reported other travel besides using their volunteered Oyster card on PT and Type 9 did not report any travel on their travel day, and neither had valid OR stages on their travel day or any other days within the OR analysis period. Type 6 and 9 people made up approximately 8% of all people who volunteered Oyster card numbers. These people are likely not giving the Oyster card they most frequently (or ever) use and if Oyster analysis is continued in the future, interviewers should be sure to collect the most frequently used



Oyster cards from each respondent that is willing to provide this information. Demographic data for Types 6 and 9 people is presented in Appendix B.

## 5.4 Types 4 and 7 Analyses

Type 4 people reported other travel besides using their volunteered Oyster card on PT, but had valid OR stages on their travel day. Type 7 people did not report any travel on their travel day, but had valid OR stages on their travel day. Type 4 and 7 people combined only made up about 4% of the people who gave Oyster cards (151 people).

### 5.4.1 Types 4 and 7 Unreported Oyster Records on Travel Day

Type 4 and 7 people's records were analyzed together to determine the level of underreporting and possible reasons for this underreporting. Possible reasons include: underreporting a journey due to survey fatigue, reporting using cash or a magnetic ticket for the LTDS stage when they actually used an Oyster card, or someone else in the household using the volunteered Oyster card, without reporting sharing this card within the household.

First, reported stage records and OR stages were analyzed to determine if people were misreporting the type of fare payment used on their travel day. 36 Type 4 people (28% of Type 4 people) reported PT travel but not on Oyster (i.e. they reported using a magnetic/paper ticket or cash, but actually used an Oyster card). Two of those people had the exact OR stage as they reported in the LTDS, 32 people had similar journeys as reported in LTDS, and two people had a journey on a different mode from the one reported in LTDS. The remaining 72% (91) of Type 4 people only reported stages on modes other than PT.

There were eleven instances of OR stages from a Type 4 or 7 person (eight Type 4, three Type 7) matching the reported PT stages of another member of the same household. In eight of these eleven instances, the other household member was a Type 1 or 2 person (one Type 1, seven Type 2) who was missing OR stages for the same reported PT stage on his or her travel day, showing that there are possibly some additional instances of sharing Oyster cards that were not reported during the survey. The Types 1 and 2 household members with missing reported stages matching the Types 4 or 7 OR stages had OR stages on the cards they actually reported on other days (and the same day for the Type 1 person), so the Oyster card from the Type 4 or 7 person cannot be confidently re-assigned to the Types 1 or 2 household members and may just be an instance where the household members were traveling together and the Type 4 or 7 person underreported the journey. Though for conservative estimates, this was assumed to be misreporting the fare media used. The other three instances were OR stages from a Type 4 or 7 person matching the reported stage of a household member who did not volunteer an Oyster card number.

The 36 Type 4 people who reported PT travel but not on Oyster and eleven instances of OR stages from a Type 4 or 7 person matching the reported PT stages of another member

of the same household cannot be considered to have underreported their stages. Instead, it is most likely that they misreported the fare media used on their travel day, which is not considered to be a major concern for TfL planning purposes. The other 104 Type 4 or 7 people, however, have most likely underreported their PT stages due to survey fatigue. These 104 Type 4 and 7 people had a total of 265 OR stages on their travel days, adding to the total underreported stages found in the Type 1 sample, described in section 4.3, to give an additional 5.9% of likely underreported stages, or 20.9% total underreporting of PT stages for people whose discrepancies can be assumed to be due to not making the same journeys that they reported, or not reporting journeys that they made.

Additionally, Types 4 and 7 people's OR were analyzed to determine if their travel day was unique in having PT records, or if they were frequent PT users. Figure 5-5 and Figure 5-6 show the number of days each Type 4 and Type 7 person with an OR stage on bus or train had the OR stage on the same mode within the week before their travel day. These figures show that 82 - 95% of Type 4 and Type 7 people had OR on the same mode at least once the week before their travel day and therefore used PT semi-regularly making it likely that they underreported the journey on the travel day due to survey fatigue.

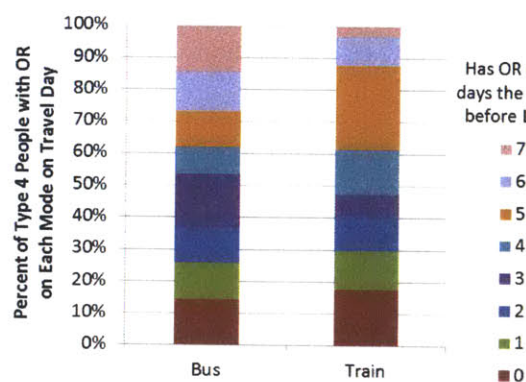


Figure 5-5: Type 4 OR the Week Before Travel Day

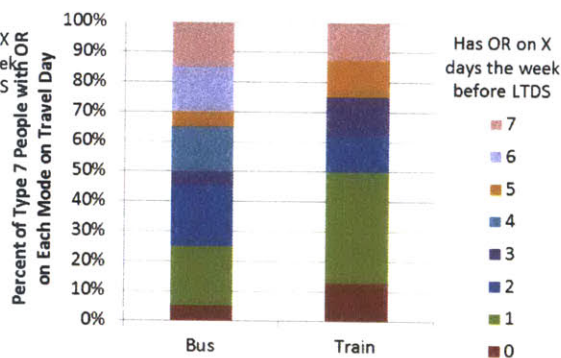


Figure 5-6: Type 7 OR the Week Before Travel Day

### 5.4.2 Types 4 and 7 Intensity of PT Use

The LTDS interviews respondents on all days of the week and captures travel day information for weekdays as well as weekend days. Approximately 72% of Type 4 and 7 people had a weekday travel day, which is similar to all other people who reported Oyster card numbers (Types 1 - 3, 5 - 6 and 8 - 9). Figures 5-7 and 5-8 show side by side comparisons of the reported PT stages and OR stages for all Types 4 and 7 people with travel days on weekdays or weekend days, respectively.

The boxplot on the far left (Avg WDayOR Cap WDays) is the distribution of the number of weekday OR stages captured by the Oyster system over all weekdays where the person had OR stages. This represents the distribution of people's average OR stages per weekday

that they took PT, as captured by the Oyster system. The boxplot second from the left (Avg WDayOR Pos WDays) shows the distribution of the average number of weekday OR stages over all weekdays where OR stage capture was possible (eight weeks prior to the individual's travel day through 18 June, 2012). The third boxplot from the left (Avg WEndOR Cap WEnds) shows the distribution of the number of weekend OR stages captured by the Oyster system over all weekend days where the person had OR stages. This represents the distribution of people's average OR stages per weekend that they took PT, as captured by the Oyster system. The boxplot fourth from the left (Avg WEndOR Pos WEnds) shows the distribution of the average number of weekend OR stages over all weekend days where OR stage capture was possible (eight weeks prior to the individual's travel day through 18 June, 2012). The boxplot second to the right shows the distribution of reported LTDS stages for people on their travel day, which was 0 for all Type 4 and 7 people, and the boxplot on the far right (Oyster on Travel Day) shows the distribution of OR stages for people on their travel day.

Figure 5-7 shows that the distribution of average weekday OR stages over all weekdays where weekday travel day people had OR stages had a higher median and a narrower range than the OR stages on the travel day, and the average weekday OR stages over all possible weekdays of OR collection had a lower median and a narrower range than the travel day. The weekend results were similar with slightly lower medians and narrower distributions than the weekday results. This shows that overall, people with weekday travel days did not use PT every weekday throughout the OR analysis period, but on the weekdays they did use PT, they used it more on average than on their travel day, which was entirely underreported.

Figure 5-8 shows that the distribution of average weekend OR stages over all weekend days where weekend travel day people had OR stages had a slightly higher median and a wider range than the OR stages on the travel day, and the average weekend OR stages over all possible weekend days of OR collection had a lower median and a narrower range than the travel day. The weekday results were similar with slightly higher medians and wider distributions than the weekend results. This shows that overall, people with weekend travel days did not use PT every weekend throughout the OR analysis period, but on the weekend days they did use PT, they used it slightly more on average than on their travel day, which was entirely underreported. Overall the survey is underestimating the typical PT use and intensity of PT use for Types 4 and 7 people. Detailed demographic data for Types 4 and 7 people is presented in Appendix B.

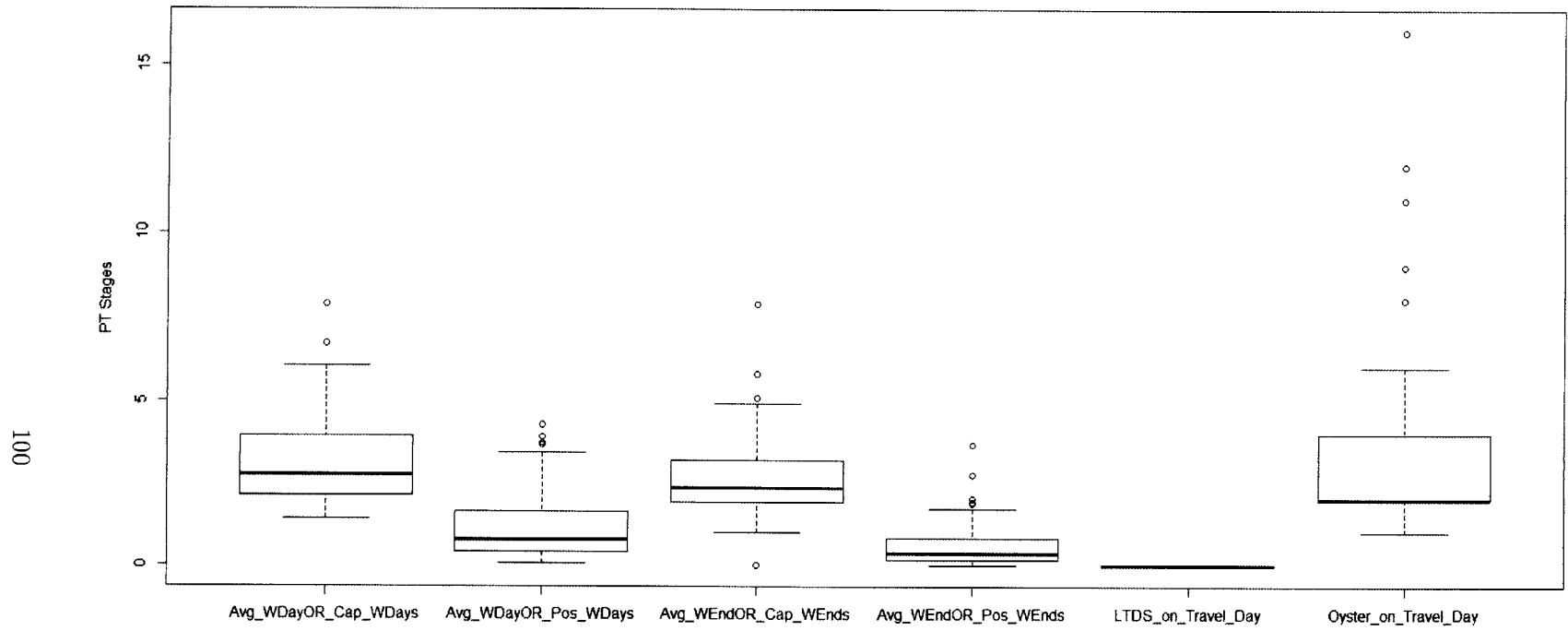


Figure 5-7: LTDS vs. Oyster Stages for Types 4 and 7 People with Weekday Travel Days

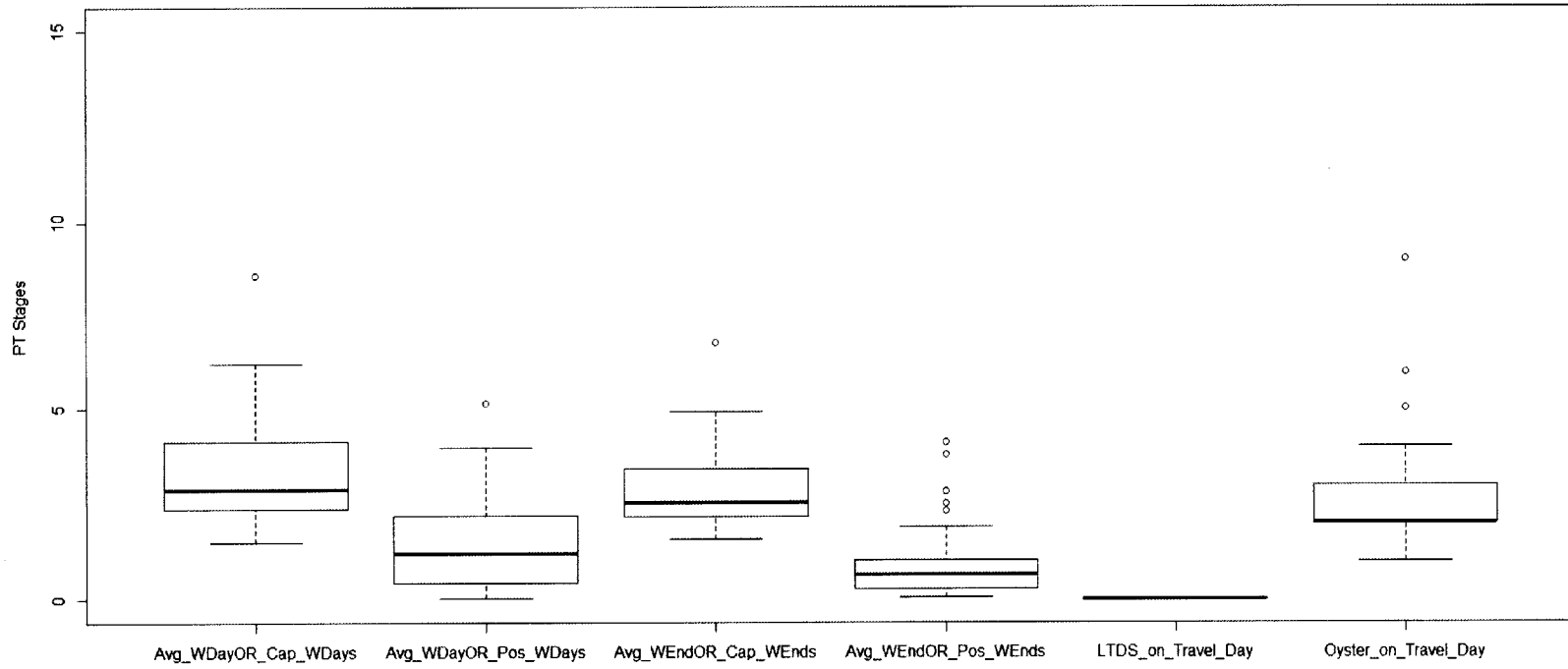
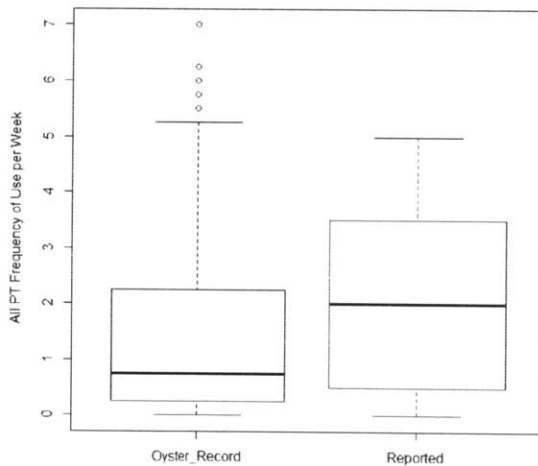


Figure 5-8: LTDS vs. Oyster Stages for Types 4 and 7 People with Weekend Travel Days

## 5.5 Types 5 and 8 Analyses

Type 5 people reported other travel besides using their volunteered Oyster card on PT, and did not have any valid OR stages on their travel days. Type 8 people did not report any travel on their travel day and did not have any valid OR stages on their travel days. However, both of these types had OR stages on other days within the OR analysis period and therefore may be frequent or periodic PT users. These two types of people made up about half of the people who volunteered their Oyster card numbers and it can be reasonably assumed that these types of travelers are represented in similar proportions in the full LTDS sample. The survey asks all people questions about their frequency of use of the PT system, and since these people had no reported PT stages on their travel day, the frequency of use question is likely the only information collected about their PT use. In order to compare this LTDS reported frequency of use with the actual frequency of use as captured by the Oyster system, the LTDS reported frequency of PT use was converted to a use per week based on the categories used in the survey and is described in section 4.2.2.

In addition, the actual PT frequency was calculated by counting the number of days the person had valid OR stages two weeks prior and two weeks after their travel day to normalize the counting method for all people. Analysis of these peoples' OR stages compared with the LTDS reported frequency of PT use is shown in Figure 5-9.



**Figure 5-9:** Types 5 and 8 Reported vs. Actual Frequency of PT Bus Use

The LTDS reported frequency of PT use was in general much higher than the actual PT use found in OR stages for Type 5 and Type 8 people. If the LTDS reported frequency is used to estimate the frequency of PT use for London residents, it is most likely overestimating the number of PT journeys people are making. If these types of travelers ( i.e. did not report PT trips, but actually use PT on a semi-regular basis) make up nearly half of the LTDS sample, this overestimate of PT frequency of use may be substantial in scaling up the full sample. Demographic data for Types 5 and 8 people is presented in Appendix B.

## 5.6 Summary of Other Person Types Analyses

Table 5.1 shows the summary of the analyses of Person Types 2 - 9 in terms of the number of stages that were over, under, or misreported in the LTDS.

**Table 5.1:** Summary of Over, Under or Misreporting of Types 2 - 9

	Sample Description	Type 2	Type 3	Types 4 and 7	Total
Misreporting Fare Media	People (% PT Sample)	160 (9%)	117 (7%)	47 (3%)	324 (19%)
	Stages (% LTDS Stages)	427 (9.6%)	351 (7.9%)	168 (3.8%)	946 (21.2%)
Overreporting (most likely misreporting travel day)		132 (8%)			132 (8%)
		338 (7.6%)			338 (7.6%)
Underreporting				104 (6%)	104 (6%)
				265 (5.9%)	265 (5.9%)

About 55% of Type 2 people had none of their reported LTDS PT stages as OR stages within the previous week of their travel day. There is no obvious reason for these people to report travel they have never taken, so it is likely these people (or about 4% of total people who volunteered an Oyster card) used a different Oyster card on this travel day. Type 3 people (3% of total people who volunteered an Oyster card) combined with the 55% of Type 2 people (an additional 4% of total people who volunteered an Oyster card) who had none of their reported stages in OR stages within the previous week, are likely people who used a different Oyster card on their travel day and most likely did indeed use PT as they reported.

There were also 36 Type 4 people who reported PT travel but not on Oyster and eleven instances of OR stages from a Type 4 or 7 person matching the reported PT stages of another member of the same household have also likely misreported the fare media used on their travel day, which is not considered to be a major concern for TIL planning purposes.

Additionally, Types 6 and 9 people (approximately 8% of all people who gave an Oyster card) had no OR stages on their volunteered cards on any days within the OR analysis period. Therefore, at least 7% and up to 15% of people who volunteered an Oyster card are likely not giving the Oyster card they most frequently (or ever) use and if Oyster analysis is continued in the future, interviewers should be sure to collect the most frequently used Oyster cards from each respondent who is willing to provide this information.

About 45% of Type 2 people had OR stages within the previous week of their travel day for some or all of their reported LTDS PT stages, showing that they have likely reported a different travel day that they may have felt was more representative of their travel behavior. These people are considered to have overreported their PT journeys on their travel day.

Type 4 and Type 7 people had an additional 5.9% of likely underreported stages, when excluding likely instances of people reporting the wrong fare media or sharing of cards within households, which added to the 15.0% of underreported stages for Type 1 people, totals at least 20.9% underreporting of PT Sample reported LTDS stages for people whose discrepancies can be reasonably assumed to be due to not making the same journeys that they reported, or not reporting journeys that they made. When taking into account the under and overreported PT stages, a net total of 73 stages (265 Underreported - 338 Overreported) should be removed from the PT Sample reported LTDS PT stages.

Analysis of OR stages on other days within the OR analysis period showed that Type 2 people did not use PT every day throughout the OR analysis period, but on the days they did use PT, they used it more on average than what they reported on their travel day. Therefore, the Type 2 person's travel day was not captured in the OR stages because the person did not make this trip on the volunteered Oyster card on that day or reported another day they felt was more representative of their average travel, and this reported day was not necessarily representative of all days the person took PT journeys. Overall for Type 2 people, the single day survey overestimates typical PT use overall, and underestimates the intensity of PT use on days when they chose to use the PT mode. Similarly, analysis of OR stages showed that Types 4 and 7 people did not use PT every day throughout the OR analysis period, but on the days they did use PT, they used it more on average than on their travel day, which was most likely completely underreported. Therefore, on average they likely used PT more than the day they underreported those journeys in the survey.

The LTDS reported frequency of PT use was in general much higher than the actual PT use found in OR stages for Type 5 and Type 8 people. If the LTDS reported frequency is used to estimate the frequency of PT use for London residents, it is most likely overestimating the number of PT journeys people are making.



## Chapter 6

# Conclusions and Recommendations

For many years manually collected surveys of travel diary and demographic information were the best way for public transport agencies to gather data and draw conclusions about the travel behavior of their customers. However, there are many drawbacks to using manual surveys as described in section 2.3, and many organizations now have access to automatic data collection (ADC) systems that provide a vast source of information about the travel behavior on public transport (PT) networks that can be used to some degree to validate and enhance the travel diary survey responses and other manually collected travel behavior information. This improved access to ADC systems has introduced the possibility of studying the integration of manually collected survey data and automatic fare card or GPS data over the last few years. This thesis goes further than previous studies by analyzing a much larger and more detailed sample of households than the scale of most GPS studies done in the past, and unlike previous automatic fare card studies, this thesis has household survey data linked with the specific household smart card data, allowing for a more in-depth study of the accuracy and representativeness of manual survey responses than has previously been possible.

This thesis uses Transport for London (TfL) as a case study to determine how and to what extent automatic fare card (Oyster) data can be used to enhance and validate the London Travel Demand Survey (LTDS) single day travel diary responses collected between July 2011 and March 2012 and improve the understanding of Londoners' public transport (PT) travel behavior over time. Analysis of the accuracy and representativeness of the reported PT travel in the LTDS was only possible for the subset of survey respondents who volunteered their Oyster card number, reported using PT on their travel day, and/or had valid<sup>1</sup> Oyster record (OR) stages on that travel day; these people were categorized throughout as the "PT Sample". A subset of people in this sample, called Type 1 people, reported PT stages on their travel day *and* had valid OR stages on their travel day. The other people in the PT Sample either reported LTDS PT stages and had no OR stages, or had OR stages but did

---

<sup>1</sup>A valid Oyster record stage is any type of PT use (not "topping up" fare value on the card or purchasing a period pass) on the Oyster card. The Oyster record stage may or may not be the exact stage as reported in the LTDS.

not report LTDS PT stages.

This chapter summarizes the overall conclusions drawn from the analyses of Type 1 and other person types and provide recommendations for the efficient integration of OR data with survey responses at TfL and other public transport agencies including suggestions for future work.

## 6.1 Conclusions from Combining LTDS and Oyster Data

A total of 3,946 people volunteered to provide up to two of their most frequently used Oyster card numbers between July 2011 and March 2012 (28% of total respondents in that period). Their LTDS survey responses and OR stages were analyzed in varying degrees of detail depending on their reported LTDS PT travel behavior and their PT activity captured by the Oyster system on their travel day and other days within the OR analysis period. A limitation of this research is that there was no guarantee that people actually used their volunteered Oyster card(s) on their travel day or that Oyster card numbers were transcribed correctly during the interview process. This made it difficult to determine if differences in reported LTDS PT stages and OR stages were because the person did not make that journey on that day, or because the person used a different card or fare media than reported. However, almost three quarters of the people who volunteered their Oyster card(s) and reported PT travel using their Oyster card(s) on their travel day had OR stages on their volunteered card(s) on that day - those people are described above as Type 1 people. For these reasons, Type 1 people were assumed to have used their volunteered Oyster card(s) on their travel day and any reported travel discrepancies of these people were assumed to be due to not making the same journeys that they reported, or not reporting journeys that they made<sup>2</sup>. With the other person types (Types 2 - 9), assumptions were made about whether the discrepancies between the LTDS and OR stages were due to the individual misreporting PT travel as opposed to misreporting the card number or fare media used, which was not considered to be of major consequence to TfL's use of the LTDS data.

Only 22% of the PT Sample had perfectly matching reported LTDS PT stages and OR stages, which only accounted for 18.4% of the PT Sample's reported LTDS stages. The remaining 78% of the PT Sample either underreported, overreported, or misreported (or some combination of under, over, and/or misreporting) their PT stages. Approximately 27% of the PT Sample have likely misreported the fare media they used on their travel day, which made up 26% of the PT Sample's reported LTDS stages. This was not considered a serious problem, as the reporting of the type of fare media used may have specific implications at TfL, but does not affect the overall level of PT use reported in the LTDS. About 33% of the PT Sample have likely misreported some portion(s) of the train origin-destination (O-D) pair or bus route(s) used, which made up approximately 21.2% of the PT Sample's reported LTDS stages. This will have implications for the LTDS survey, if the specific origins and destinations or bus routes reported are being used for planning purposes. About 14% of the

---

<sup>2</sup>Some discrepancies may also be due to the person traveling between two un-gated NR or DLR stations using a period pass where they were not required to tap their card, although these trips are believed to be a relatively small proportion of all reported trips.

PT Sample overreported their PT stages (approximately 11.6% of the PT Sample's reported LTDS stages), most likely in terms of the travel day. These 14% most likely reported a day other than their agreed upon travel day that they may have believed was more representative of their regular travel behavior. An additional 26% underreported their PT stages, which made up approximately 20.9% of the PT Sample's reported LTDS stages. Only 51.1% of OR stages had matching LTDS stages and 45.6% of LTDS stages had matching OR stages. When taking into account the under and overreported PT stages, a "net" additional 416 stages, or 9.3%, should be added to the PT Sample reported LTDS PT stages.

When looking at mutually exclusive categories of people who have mis, over, or underreported specific LTDS and OR stages in terms of the reported bus routes and train origin-destination (O-D) pairs, the people who had perfectly matching LTDS and OR stages (or only misreported the fare media used on their travel day) made up only 44% of the PT Sample and 41.3% of the PT Sample's reported LTDS stages. The other 56% of the PT Sample have over, under, or misreported the O-D pair or bus route of their 58.7% of stages, which has significant implications about the accuracy of the LTDS if more than half of the people who have reported PT travel in the survey are incorrectly reporting the number of PT stages and/or the origins, destinations, or bus routes used on their travel day.

## 6.2 Variability of PT Travel

The continuous collection of OR stages for all people who volunteered an Oyster card number allowed for the analysis of OR stages on other days in the OR analysis period to show the weekly variation in personal travel for LTDS respondents compared to the reported frequency of use of each PT mode in the LTDS. The reported frequency of PT use in the LTDS was much higher than the actual use as captured by the Oyster system, and therefore the LTDS is overestimating the PT use overall for London residents, as these constitute the survey target group, excluding visitors and non-residents. The presence of OR stages on the travel day was a much better predictor of being a frequent PT user than reporting PT use on the travel day in the LTDS.

When comparing the Oyster records collected for Types 1 and 2 people over the entire analysis period with the reported PT and OR stages on the travel day, one can reasonably conclude that at least for PT travel, the single day survey overestimates typical PT use overall (if it were to be scaled up to get typical ridership for a year), but it underestimates the intensity of PT use on days when the survey sample chose to use the PT mode and on the actual survey day by the participants.

One of the most significant findings from analyzing the weekly variation in PT travel was that people who reported one or more LTDS PT stages on their travel day had from 38-57% of their remaining person-days (over the period two weeks prior and two weeks after the travel day) with zero PT stages. This suggests that the people who are considered PT users in the survey only used PT about half of the time, and if the one day survey just so happens to assign a travel day on a day where the person used PT, it would overestimate their typical PT use. Therefore, it would be advisable to consider a two or three day survey

as it would better capture the variability in PT use for survey respondents. Alternatively, the use of OR's and Gordon's method (Gordon, 2012) to link the separate stages made by an Oyster card user into full, multi-modal journeys could supplement the one-day survey and provide a much better estimate of the frequency of use and variability of PT travel than the frequency questions asked in the survey or fixed trip rates used in travel demand models.

### 6.3 Selection and Representativeness of the Type 1 Panel

As described above, Type 1 people were assumed to have used their volunteered Oyster card(s) on their travel day and any discrepancies between the reported PT stages and the PT stages captured by the Oyster system were assumed to be due to the person not making the same journeys that he or she reported, or making additional PT journeys that were not reported. The most detailed analysis of the accuracy of the reported LTDS PT stages was done for the panel of Type 1 people created in this thesis because they had the most directly comparable information available about their PT journeys on their travel day.

Type 1 people only made up approximately 29% of all people who volunteered Oyster card numbers. Therefore it was important to ensure that they were a representative sample of all people in the survey who reported PT travel. As described in Chapter 4, Type 1 people were representative of all people who have reported taking PT stages on their travel day in terms of the geographic distribution in each borough as well as many demographic and reported trip mode share characteristics. There were, however, some demographic differences that should be kept in mind when considering how the results of the panel of Type 1 people can be expanded for all people who reported LTDS PT stages on their travel day. Type 1 people were not different than the other people who reported PT stages in terms of the number of PT stages that were reported in the LTDS, but they had more OR stages on their travel day than they reported. This result suggests that the other people who reported PT, but did not volunteer an Oyster card or have valid OR stages on their travel day may also have traveled more than they reported on their travel day, and therefore the single day survey is underestimating the intensity of PT use on days where the survey sample chose to use the PT mode.

Type 1 people's LTDS PT stages were compared to their OR stages on the basis of the overall number of PT stages, origin and destination of PT stages, modes used, and temporal characteristics. The analysis of Type 1 people's stages per trip showed that the percentage of trips that had corresponding OR stages for each reported PT stage decreased with the increase in number of stages per trip. Therefore, the accuracy of reporting decreases as the trips become more complicated. 93% of Type 1 people had OR stages on the reported PT modes for all of their reported journeys. This shows that overall, most people who reported using PT on their Oyster card actually used their Oyster card on the same mode as reported on their travel day, and there were no significant differences between bus and train journeys. Over 60% of people had OR stages on the same bus routes reported in the survey, but only 46% of people had OR stages on all bus routes and taps reported in the survey and therefore reported all of their bus routes accurately, but may have over-reported

taking a bus multiple times a day. Depending on the type of train used, 35-58% of people did not report the same O-D pairs captured by their OR. O-D pairs for Underground journeys were matched most accurately at 65% of people with the full reported Underground journey found in the OR. Therefore, use of OR stages and Gordon's recently developed method to link the separate stages made by an Oyster card user into full, multi-modal journeys could be a better way to capture accurate records of complicated trips, specific bus routes and train O-D pairs.

The LTDS reported stage start times were more peaked than the OR stage start times because many of the reported start times are right on the hour, while the OR stage start times were spread more evenly. Train journeys were reported within 10 minutes of the OR start time about 50% of the time and bus journeys were reported within 10 minutes of the OR start time about 38% of the time. Additionally, inferred speeds and durations from LTDS reports varied greatly from the speeds and durations estimated by the Oyster system. OR stages and Gordon's method to fully link Oyster journeys for each person who reported using an Oyster card on their travel day and took PT journeys could be used to calculate exact journey start and end times, durations, and speeds for Type 1 people to validate and enhance the reported PT stages in the LTDS survey.

The LTDS accurately collects general information about the modes used by each respondent, and is a better source of information about demographic and journey purpose data. However, more detailed information about the exact mode share per person, specific bus routes and train stations used, start times, durations, and speeds could be more accurately recorded with Oyster data, especially if the method developed by Gordon is used to infer trip origins and destinations (including specific bus stops) and link trip stages into full journeys that include exact geographic and temporal data for each Oyster record. This can be done with the most accuracy for people who had OR stages for the exact reported sequence of LTDS PT stages. These perfectly matching journey people made up only 22% of the PT Sample. This highlights the difficulty of combining the survey responses with Oyster data after the survey has taken place. The other Type 1 people's reported PT stages could be enhanced with accurate start and end times where there are corresponding OR stages, if they were within a reasonable time-frame and journey sequence, but this may require many assumptions about whether the captured OR stage was the PT journey the respondent was referring to. This suggests that it would be more useful and effective to combine Oyster data earlier in the process, such as during the interview to take advantage of the continuous collection of travel behavior data for people who use Oyster cards. This conclusion will be discussed in more detail in the next sections.

## 6.4 Recommended Next Steps

Overall, neither source of data (LTDS or Oyster) is perfect. The LTDS likely has many misreported, overreported, and/or underreported trips and inexact start times, durations, and journey sequences, but it does have very detailed information about trip purpose, demographic data, and travel on non-PT modes that cannot be captured with Oyster data alone. Oyster card data provides a continuous collection of the PT travel behavior of millions

of Londoners and visitors every day. These data allow TfL to monitor fare collection and revenue electronically, provide reliable estimates of the travel time flows on its network, and provide accurate records of trips, in terms of locations, start times and durations, and provide information about the travel behavior of visitors that are not captured in the LTDS. However, the Oyster card database only includes basic demographic data, including gender, postcode, and age, for voluntarily registered Oyster customers. Many other cards have no associated demographic data associated with them, making it difficult to discern different travel patterns among specific demographic groups.

#### **6.4.1 Single Day Responses**

The conclusions listed above suggest that combining Oyster data with the survey responses could greatly enhance the validity and understanding of the single travel day recorded in the LTDS. However, as was shown in this thesis, it is difficult to match up these sources after the survey interview has taken place, given that only 51.1% of OR stages had matching LTDS stages, only 45.6% of LTDS stages had matching OR stages, and only 366 people (or 22% of the PT Sample) had the exact reported journey sequence found in their Oyster records. Even when there were matches in terms of the train O-D pairs and bus routes used, there were large differences in many of the start times and durations of these journeys with an average start time difference of matched journeys of 61.2 minutes, forcing one to assume whether or not they are actually the same journeys. This suggests that it would be advantageous to integrate the Oyster records earlier in the LTDS interview process.

Processed OR journeys using the recently developed (Gordon, 2012) method to link OR stages into full multi-modal PT journeys could be used during the interview to aid in the recollection of willing respondent's PT journeys on the previous day, similar to the prompted recall (PR) surveys described in section 2.4.2. Interviewers could ask the respondent for their Oyster card number while they are scheduling the interview (which is generally a few days before the interview day), and if they had their permission, get the Oyster records for the travel day from TfL in advance to shorten the length of the interview process. In the near term, interviewers could use Oyster card readers (similar to the card readers used by ticket inspectors) connected to an electronic tablet to display the Oyster stages collected on the previous day(s). The paper survey could continue in its current form, with the tablet display of the previous day's records as a starting point with the respondent's filling in missing auto or walking stages in addition to trip purpose and other demographic data. To facilitate this process, the interviewer could explain that using the card reader to supplement the survey form could greatly shorten the overall length of the survey process. Over time this tablet could replace the paper survey forms and greatly enhance the LTDS survey by improving accuracy and reducing reporting errors, and data entry time and cost, allowing the interviewer to spend more time on other types of questions that are more interesting and important, like the attitudes of Londoners about different modes of transportation or new project schemes.

#### **6.4.2 Travel Behavior Over Time**

Oyster card data can be used to supplement the questions regarding frequency of use of PT modes asked in the survey. The reported frequency of PT use in the LTDS was found to

be much higher than the actual use as captured by the Oyster system. OR stages could be analyzed for each person who volunteered their Oyster card number for any period of time, but previous research summarized in Block-Schachter (2009) showed that after two weeks of travel behavior data, there was little additional variability gained by adding days to the sample in the Cambridge, MA area. Two or more weeks of OR stages could be analyzed for each LTDS respondent who volunteered their Oyster card number (in addition to the travel day), and the period of time studied could be tested to find an appropriate length of analysis depending on the resources available and variability observed on the network in London. Data analyzed should show the variability in modes used, difference between number of OR stages, travel time, activity duration, etc. depending on the mode, day of the week or other variables.

Additionally, a panel of Type 1 people that includes detailed individual and household demographic data collected in the LTDS could be created to build O-D matrices of full journeys from the Oyster database on a monthly or yearly basis (or other time period) to improve the understanding of changes in user travel behavior in great detail over a span of time not available with current survey methods, with the opportunity to disaggregate the data by demographic characteristics not available with the Oyster data alone.

The Oyster database can only be used to validate and enhance LTDS survey responses with respect to PT modes. However, the LTDS is used to study all modes of travel in London. Other studies described in 2.4 have been done using GPS enabled devices to validate the travel behavior of all modes of transportation analyzed in household travel surveys. Pilot studies could be undertaken to use GPS devices to validate the diary information for other modes of transportation, using the type of PR surveys described in section 2.4.2. For example, The Future Mobility Study (FMS) described in Cottrill et al. (2012) is a smartphone-based PR travel survey that aims to support data collection initiatives for transport modeling purposes. They developed a mobile-phone application that traced users over a certain time period and allowed the users to input additional information about the travel modes used, journey purposes, and travel companions. The survey required an extensive period of testing in order to develop a practical system that was understood by participants, frugal with mobile phone battery use, and useful to practitioners. The pilot study has not resulted in a significant sample yet, but has provided valuable insight into user needs regarding the interface, as well as training data for the background intelligence for stop and mode detection. It demonstrates the capability of smartphone-based travel surveys and the effort needed for successful development. In addition to GPS receivers, smart phones can also locate the user by alternative means of WiFi network signatures and the mobile network, which also works inside buildings (Zilske and Nagel, 2012) and throughout the subways in many places. This could provide insights into the paths chosen and interchange locations on networks like the London Underground where interchange information is only collected at stations where people tap into a reader to show that they did not cross into a certain zone for fare calculation purposes. This research area is still developing, but presents an exciting new area of travel demand data collection opportunities that should be considered by transportation planning agencies, like TfL, in the future.





# Appendix A

## LTDS Trip Purpose Definitions

These trip purposes, defined by the surveying agency, Kantar Operations, are summarized from their training materials and aggregated for the purposes of this thesis.

- Home - This can only be used if the respondent states they are going home (i.e. to the sample address). The only exception is if a “Visitor” to the household states they are going home which will not be the sample address.
- Work - Usual place of work/workplace - This can only be coded if it is the respondent’s usual place of work. If they are traveling somewhere else in the course of their work (e.g. to go to a meeting, visit a client, or make a delivery/pick something up), one should use the following two codes as appropriate.
- Delivering/loading/picking something up (related to work) - Use this code if the respondent is conducting an errand in the course of their work. For example, use this code if the respondent is delivering/picking up a package at a post office or to/from a client (this is essentially “ad-hoc” and not an integral part of their job).
- Other work - This is for all other trips made in the course of work not covered by the other work codes. For example, use this code if the respondent trip is to see a client or a trip to a conference that is work related.
- Recreation - made up of the following:
  - Entertainment/recreation - Use this code if purpose of trip is ‘leisure or pleasure’. This would include going to bingo or the cinema/theater. Also included in this category are holidays/weekends away, going for a walk, walking the dog, etc.
  - Leisure - Use this code for all other purposes, going for a walk, walking the dog, etc.
- Sport - Use this code for either as a player or as a spectator
- Social - Made up of the following:
  - Visiting friends - Use this code if the respondent is visiting friends or relatives at their home
  - Other Social - Use this code for social visits including the pub and restaurants

- Shopping - Made up of the following:
  - Shopping (Food/grocery) - Use this code if the respondent is going to the super-market or any food shopping
  - Shopping (all other types of shopping) - Use this code if respondent is going to buy all other types of shopping including window shopping
- Personal - Use Services/Personal Business - Use this code if respondent is going to use services of any kind. Essentially this code covers trips made for “personal” reasons and includes, for example, hairdressers, dry-cleaners, lawyers, banks, building societies, etc. This code also includes buying goods at places other than shops, such as trunk sales, stations, garages, etc.
- Medical - Health or Medical Visit - Use this code if the respondent goes to the doctor or dentist
- School - Education - This can only apply if the respondent goes somewhere to be educated (including evening classes). If respondent is a teacher/lecturer etc., they should be coded as going to work as appropriate.
- D-off/P-up - Dropping off/Picking up Trips - Made up of the following:
  - Drop off/Pick up/Accompany someone to/from work - For example, a wife who drops her husband off at work/a station from which they get to work. If she then goes back home/proceeds to her workplace or somewhere else (e.g. to drop children off at school), this is a different/new trip
  - Drop off/Pick up/Accompany someone to/from school/college/university - Use this code if a respondent is taking a child to school/college/university - either all or part of the way
  - Drop off/Pick up/Accompany someone to/from health or medical visit - Use this code if the respondent is accompanying/escorting someone to the hospital or doctor
  - Drop off/Pick up/Accompany someone to/from another place - Use this code if respondent is 'accompanying/escorting' someone elsewhere (e.g. friend/relative to station/airport)
- Worship - Worship or religious observance - Use this code if respondent is going to worship in a church, mosque, temple, synagogue, etc. Do not use if respondent is going to worship in a private home or public building - this should be recorded as “personal business”

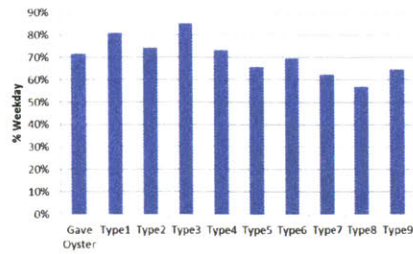
## Appendix B

# Demographic Characteristics

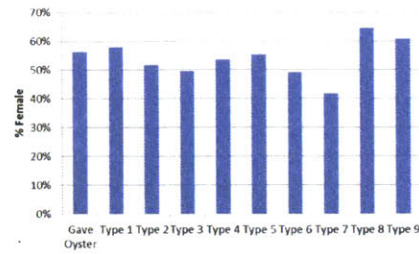
- **Type 1 - PT Users with Similar LTDS and OR Travel** - people reported LTDS PT stages using their volunteered Oyster card(s) and had valid OR stages on their card(s) on their travel day.
- **Type 2 - PT Users with Different LTDS and OR Travel** - people reported LTDS PT stages using their volunteered Oyster card(s), but did not have valid OR stages on their card(s) on their travel day. They had OR stages on their card(s) on other days within the OR analysis period.
- **Type 3 - PT Users with LTDS but No OR Travel** - people reported LTDS PT stages using their volunteered Oyster card(s), but did not have valid OR stages on their card(s) on their travel day or any other days within the OR analysis period.
- **Type 4 - PT Unreported on LTDS with PT OR on Travel Day** - people reported other travel besides using their volunteered Oyster card(s) on PT, but had valid OR stages on their card(s) on their travel day.
- **Type 5 - PT Unreported on LTDS with Different OR Travel** - people reported other travel besides using their volunteered Oyster card(s) on PT, and did not have any valid OR on their card(s) on their travel day. They had OR stages on their card(s) on other days within the OR stages analysis period.
- **Type 6 - PT Unreported on LTDS with No OR Travel** - people reported other travel besides using their volunteered Oyster card(s) on PT, and did not have valid OR stages on their card(s) on their travel day or any other days within the OR analysis period.
- **Type 7 - No Travel on LTDS with PT OR on Travel Day** - people did not report any travel on their travel day, but had valid OR stages on their volunteered card(s) on their travel day.
- **Type 8 - No Travel on LTDS with Different PT OR Travel** - people did not report any travel on their travel day and did not have any valid OR stages on their volunteered card(s) on their travel day. They had OR stages on their card(s) on other days within the analysis period.
- **Type 9 - No Travel in LTDS or OR** - people did not report any travel on their travel day and did not have any valid OR stages on their volunteered card(s) on their travel day or any other days within the OR analysis period.

**Table B.1:** Comparison of Demographic Characteristics

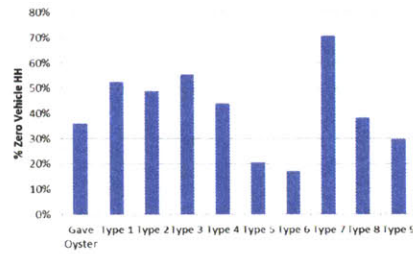
Demographic Characteristic		Gave Oyster	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9
Travel Day	% Weekday	71%	81%	74%	85%	73%	66%	70%	63%	57%	65%
Gender	% Female	56%	58%	52%	50%	54%	55%	49%	42%	65%	61%
Ethnicity	% White	72%	68%	63%	68%	70%	78%	78%	71%	70%	76%
	% Asian or Asian British	13%	13%	17%	14%	12%	12%	16%	8%	16%	9%
	% Black or Black British	10%	12%	13%	14%	11%	7%	4%	13%	11%	7%
	% Other Ethnic Group	3%	3%	3%	3%	3%	2%	2%	8%	2%	7%
	% Multiple Ethnic Groups	2%	3%	3%	2%	4%	2%	0%	0%	2%	1%
HH Income	Lower Income (< £25k)	46%	46%	60%	38%	50%	40%	45%	83%	55%	51%
	Middle Income (£25 - £75k)	39%	37%	29%	42%	36%	45%	41%	13%	34%	39%
	Higher Income (> £75k)	14%	16%	11%	21%	14%	15%	14%	4%	11%	9%
Household Size	%1 Person HH	21%	24%	26%	22%	21%	17%	17%	13%	21%	20%
	%2 Person HH	35%	35%	31%	29%	36%	36%	34%	33%	34%	45%
	%3 Person HH	18%	17%	17%	23%	15%	20%	19%	21%	15%	12%
	%4 Person HH	17%	15%	15%	12%	15%	18%	21%	13%	19%	14%
	%5 Person HH	6%	6%	8%	7%	4%	3%	4%	0%	6%	4%
	% 6 or more Person HH	3%	3%	3%	7%	4%	3%	4%	0%	6%	4%
Zero Vehicle HH	% Zero Vehicle HH	36%	52%	49%	56%	44%	20%	17%	71%	38%	30%
Disability	% No Disability	87%	88%	88%	94%	90%	89%	84%	75%	77%	61%
Age	% 5-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	% 11-15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	% 16-24	13%	17%	15%	11%	15%	9%	7%	8%	15%	8%
	% 25-44	39%	45%	43%	56%	39%	35%	45%	38%	28%	27%
	% 45-59	21%	17%	14%	13%	17%	26%	20%	25%	23%	16%
	% 60+	28%	22%	28%	21%	28%	30%	29%	29%	34%	49%
Total		3,946	1,148	292	117	127	1,474	224	24	466	74



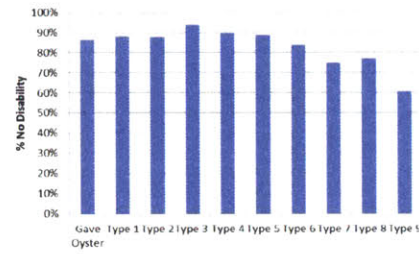
(a) % Weekday



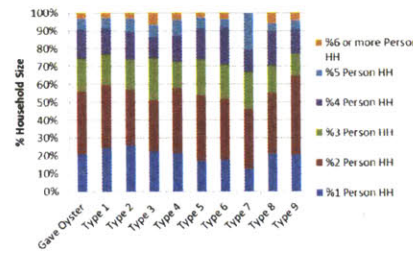
(b) % Female



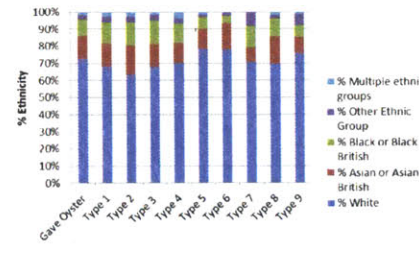
(c) % Zero Vehicle Household



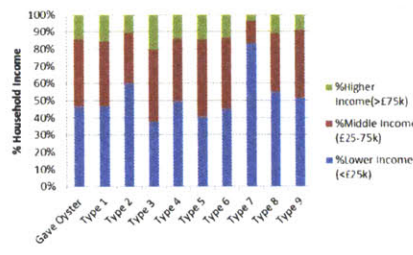
(d) % No Disabilities



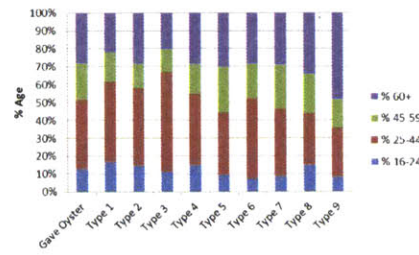
(e) Household Size



(f) Ethnicity



(g) Household Income



(h) Age

**Figure B-1:** Demographic Characteristics of Person Types Compared to All People who Volunteered an Oyster Card



# Bibliography

- Bagchi, M. and White, P. The potential of public transport smart card data. *Transport Policy*, 12(5):464 – 474 (2005). ;ce:title;Road User Charging: Theory and Practices;ce:title; ;xocs:full-name;W. Saleh;/xocs:full-name;.  
**URL:** <http://www.sciencedirect.com/science/article/pii/S0967070X05000855>
- Block-Schachter, D. *The myth of the single mode man: how the mobility pass better meets actual travel demand*. Ph.D. thesis, Massachusetts Institute of Technology (2009).
- Bricka, S. and Bhat, C. Using GPS Data to Inform Travel Survey Methods. In *Innovations in Travel Demand Modeling Conference, Transportation Research Board, Austin, TX* (2006a).
- Bricka, S. and Bhat, C.R. Comparative analysis of global positioning system-based and travel survey-based data. *Transportation Research Record: Journal of the Transportation Research Board*, 1972(1):9–20 (2006b).
- Chapleau, R., Trépanier, M., and Chu, K.K. The ultimate survey for transit planning: Complete information with smart card data and GIS. In *8th International Conference on International Steering Committee for Travel Survey Conferences, Lac d’Annecy, France* (2008).
- Cottrill, C., Dias, I.F., Lim, H.B., Ben-Akiva, M., and Zegras, P.C. The Future Mobility Survey: Experiences in Developing a Smartphone-Based Travel Survey in Singapore (2012).
- Gordon, J. *Intermodal Passenger Flows on London’s Public Transport Network: Automated Inference of Full Passenger Journeys Using Fare-Transaction and Vehicle-Location Data*. Master’s thesis, Massachusetts Institute of Technology (2012).
- Muhs, K.J. *Automatically Collected Data to Infer Travel Behavior: A Case Study of the East London Line Extension*. Master’s thesis, Massachusetts Institute of Technology (2012).
- Pelletier, M.P., Trépanier, M., and Morency, C. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4):557–568 (2011).
- PTV, Nu Stats, L. Regional Travel Survey: Final Report. Technical report, Atlanta Regional Commission (2011).
- Stopher, P., FitzGerald, C., Bretin, T., and Zhang, J. Variability in day-to-day

- travel—analysis of a 28-day GPS survey. In *Paper to be presented to the 29th annual meeting of the Australasian transport research forum, Brisbane* (2006a).
- Stopher, P., FitzGerald, C., Greaves, S., and Biddle, T. What can we learn from GPS measurement of travel. In *29th ATRF meeting, Brisbane* (2006b).
- Stopher, P., FitzGerald, C., and Xu, M. Assessing the accuracy of the Sydney Household Travel Survey with GPS. *Transportation*, 34(6):723–741 (2007).
- Stopher, P. and Shen, L. In-Depth Comparison of Global Positioning System and Diary Records. *Transportation Research Record: Journal of the Transportation Research Board*, 2246(1):32–37 (2011).
- Stopher, P.R., Prasad, C., and Zhang, J. *Can GPS Replace Conventional Travel Surveys?: Some Findings*. Institute of Transport and Logistics Studies (2010).
- Stopher, P.R. and Zhang, Y. Repetitiveness of Daily Travel. *Transportation Research Record: Journal of the Transportation Research Board*, 2230(1):75–84 (2011).
- Transport for London. Travel in London, Key Trends and Developments, Report number 1. Technical report, Transport for London (2009).
- Transport for London. Transport for London Company Information. <http://www.tfl.gov.uk/corporate/about-tfl/4510.aspx> (2013). Accessed: 06 March, 2013.
- Trepanier, M., Morency, C., and Blanchette, C. Enhancing household travel surveys using smart card data. In *Transportation Research Board 88th Annual Meeting*, 09-1229 (2009).
- Xu, Y. *Using Volunteer Tracking Information for Activity-Based Travel Demand Modeling and Finding Dynamic Interaction-Based Joint-Activity Opportunities*. Master's thesis, University of Tennessee, Knoxville (2011).
- Zilske, M. and Nagel, K. Towards volunteered digital travel demand data. In *Proceedings of the 7th International Conference on Geographic Information Science, Extended Abstracts. Also VSP WP*, pages 12–19 (2012).